African Union

THE AFRICA-EU PARTNERSHIP

LE PARTENARIAT AFRIQUE-UE

Evaluation as an Integral Part of Quality Assurance

Petra Pistor

IQA-4-Africa

From Pan-African Policy to Practice

This initiative is implemented on behalf of the European and African Union Commissions by:

GLOBAL OBSERVATORY Knowledge, Innovation and Development

OBREAL Observatorio de las Relaciones Unión Europea-América Latina

DAAD Deutscher Akademischer Austauschdienst German Academic Exchange Service

enqa. European Association for Quality Assurance in Higher Education

HAQAA2 Harmonisation of African Higher Education Quality Assurance and Accreditation

This HAQAA2 training material is provided by: FH MÜNSTER University of Applied Sciences

Evaluation as an Integral Part of Quality Assurance

Petra Pistor

Preliminaries

This is an introductory text to the following screencasts in the cinema of the topic *Evaluation as an Integral Part of Quality Assurance*:

- What is evaluation?
- Evaluation Standards and Guidelines
- How to Make Things Measurable
- Questionnaire Design
- Evaluation Instruments – Case Study of Redeemer's University, Ede, Nigeria
- How to Use Evaluation Results – Case Study of FH Münster, Germany

The text is not a complete summary of the information given in the screencasts. It can introduce you to the respective topic and help you to remember the key facts later. You can make use of the full potential of the learning material and have the most fruitful learning experience, if you watch the respective screencasts in the cinema and work on the reflective questions, which you find at the end of this document.

Further reading to deepen your knowledge can be found in the bookshelf of the topic *Evaluation as an Integral Part of Quality Assurance*. Material for the transfer from theory into practice, like worksheets, templates etc. can be found in the toolbox of the topic *Evaluation as an Integral Part of Quality Assurance*.

All material is part of a compendium that was developed for the HAQAA2 Training Course IQA-4-Africa – from Pan African Policy to Practice.

Keywords

evaluation, empirical social research, operationalisation, data collection, quantitative and qualitative studies, explorative, descriptive and explanatory studies, internal and external evaluation, reporting, data usage

Expected Learning Outcomes:

On successful completion of the material *Evaluation as Integral Part of Quality Assurance*, you should be able to:

- explain the key concepts of evaluation and the relevance of social empirical research for evaluation processes and quality assurance,
- precisely define the research objects you want to evaluate by the help of an operationalisation process,
- apply basic rules for designing questionnaires while developing evaluation instruments,
- take into consideration quality criteria for evaluation, when designing own evaluation projects.

1. What is Evaluation?

Geared towards development, following different functions

In everyday language we use words like *assessment*, *monitoring*, *testing*, *measuring*, *appraisal, benchmarking* and perhaps also others as synonyms for evaluation. Very fundamentally, we can say that the English word *evaluation* goes back to the Latin word *valere*, which stands for *being strong* or *being of value*. Thus, evaluation must have something to do with our intuitive understanding of quality as was introduced in the second module of this training programme.

Also, when we look at evaluation from a more scientific point of view, the definition of the term *evaluation* – as is the definition of quality – is context dependent (cf. Module 2) and there are different schools of thought, which define the term differently. In this text and the respective screencasts, we define evaluation as the **systematic description and assessment of particular phenomena on the basis of explicit or implicit criteria**. Going beyond mere assessment, evaluations help us to collect comprehensive information about a research object and therefore can support a decision-making process for the deduction of improvement measures for this research entity, for example a study programme in our university.

According to the German scholar Lars Balzer (2005), professional evaluation is characterised by the following aspects:

- It is focused on a clearly defined object. This research object is also called *evaluand*.
- It is conducted by experts, who are also referred to as *evaluators*.
- It is based on precisely defined and transparent evaluation criteria.
- Information and data are collected by the help of sound methods of empirical social research.

Balzer also states, that the result of evaluation was “a feedback of exploitable results in the form of descriptions, justifiable interpretations and recommendations to as many stakeholders as possible in order to improve the evaluand and support future actions.” (Balzer, 2005, p. 16, transl. PP). Thus, a focus is put on the developmental intention of evaluation.

Originating from the area of development cooperation, where it was implemented to assess programme effectiveness and value for money of aid activities, evaluation nowadays can be found in many professional areas. It also has become an important part of the activities in Higher Education Institutions worldwide. That is why evaluation is reflected in policy documents of the Higher Education Area, like for example the African Standards and Guidelines for Quality Assurance (ASQ-QA) or the European equivalent (ESG). In these documents reference is made to evaluation in various fields of a HEI’s activities, like teaching, research, financial management etc.

Evaluations can be based on different paradigms and therefore fulfill different functions. We differentiate the following evaluation paradigms (cf. Kromrey 2001):

- **research paradigm**: The evaluation is carried out in order to expand the knowledge base about a research object and to test hypotheses about this object (also: theory based evaluation, evaluation research).
- **control paradigm**: The evaluation is conducted to execute control over a process or intervention and to measure its effectiveness, efficiency and acceptance and to ensure accountability (also: impact evaluation).
- **development paradigm**: The evaluation serves the purpose to improve the research object and provides information that forms the basis for a change process (also: accompanying research).

Focusses, perspectives and underlying principles

Depending on what the focus of analysis is, evaluations will have different underlying concepts. **Formative evaluations** support the process of the intervention to be evaluated, e.g. by collecting information at different points of time during the process. Formative evaluations can be described as "evaluative activities undertaken to furnish information that will guide program improvement" (Rossi, Lipsey, and Freeman 2004, p. 63). Whereas **summative evaluations** focus on a result of a process. They are "evaluative activities undertaken to render a summary judgement on certain critical aspects of the program's performance, for instance, to determine if specific goals and objectives were met." (ibid., p. 65).
Evaluators can approach an evaluand from different perspectives. Either before the implementation of an intervention (ex-ante), during the intervention (interim) or after the intervention (ex-post).
Table 1 (Stockmann, 2004, transl. PP) summarises the different perspectives, focuses and characteristics of evaluations in connection with the phase of the programme or intervention to be evaluated:

phase of an intervention	**point of time / perspective of analysis**	**focus of analysis**	**evaluation concept / character**
planning phase	ex-ante / anticipatory	analysis of framework conditions and target group for a planned action knowledge gain about the basis and starting situation of a planned action	pre-formative / formative: process oriented, constitutive, constructive
implementation phase	interim / accompanying	re-analysis of framework conditions and target group for a planned action analysis of first effects of the planned action knowledge gain about what has to be adapted to enhance the success of the action	formative / summative (both possible)
impact phase	ex-post / retrospective	analysis of effects of the planned action cost-benefit analysis knowledge gain about the success of the action	summative: concluding, summarising, focussed on results/effects

Table 1: Characteristics of evaluations according to Stockmann 2005 (translation PP).

Apart from that, we can differentiate **internal evaluations** from **external evaluations**. By internal evaluations, we refer to evaluations that are conducted inside an organisation by members of that organisation itself. For example, the evaluation of a study programme by an internal QA office of a Higher Education Institution would be an internal evaluation. There is also a special form of internal evaluation, namely the self-assessment, where persons responsible for a programme or intervention evaluate their own project. For example, the evaluation of a study programme by the coordinator and the lecturers of the study programme along the lines of the African Quality Rating Mechanism (AQRM), would be an internal self-assessment.

We speak of external evaluations, whenever an evaluation is carried out by a body outside of our organisation, for example the accreditation of a study programme by a national accreditation board.

3. Evaluation Standards

JCSEE Standards, AGE, DeGEval Standards and what it all comes down to

Evaluation results oftentimes serve as a basis for taking decisions. Especially in Higher Education, where we mostly and ideally find evaluations conducted following the development paradigm (see above), improvement measures are designed on the fundament of evaluations. These improvement measures may include decisions like to change the order of modules of a particular study programme or to order more books on a specific topic in the library. But also more serious and impactful decisions might be based on evaluation results, like the allocation of funds, the extension of a contract of a lecturer, the restructuring of a department. Thus, since impactful decisions might be based on evaluations, it is important that these evaluations are conducted on the basis of ethical and methodological standards.
Since the 1970s, when evaluation still was a rather young phenomenon and evaluation research was more and more professionalised, committees came into being which deal with good practice and standards in evaluation work, like the Joint Committee on Standards for Educational Evaluation (JCSEE, founded 1975) or the German Association of Evaluation (Deutsche Gesellschaft für Evaluation, DeGEval, founded 1997). (cf. Pistor/Stammen, p. 20). In 1999, the African Evaluation Association (AfrEA) came into being as an umbrella organisation of several national evaluation associations and networks. One of its aims to this day is the promotion of research on evaluation and training placed in African contexts and evaluation approaches.
All of the aforementioned associations have developed standards that combine procedural and ethical principles and guidelines to give evaluators orientation for their work. Although the standards of JCSEE, DeGEval, AfrEa (and supposedly many others) look different at first glance, because they use different headings to formulate the standards, a deeper look into the content of the standards reveals that they all follow the same overarching principles: relevance, feasibility, fairness, accuracy, accountability (see also JCSEE 2010 , DeGEval 2016, AfrEA 2020). However, unlike the others, the African Evaluation Guidelines (AEG) emphasise the national and cultural determinants for evaluations and demand that they should be taken into account in evaluation procedures.
Derived from the core principles of the above-mentioned standards, we can identify the following success factors for evaluations:

- **relevance**: The evaluation is more likely to be accepted (and to make any sense at all), if the results are relevant to a certain question and to stakeholders (no “nice to know”).
- **communication**: If stakeholders know about the goal of the evaluation, they will be more likely to support the endeavour. If they are informed about progress and results of an evaluation, it can increase their supportiveness for upcoming evaluation projects.
- **process organisation**: Thoroughly planning and managing your evaluation project will decrease the probability of unforeseen difficulties and facilitate a smooth progress of the evaluation.
- **trustworthiness**: If you can demonstrate that as a QAO or evaluator you are neutral and without prejudice, this will not only increase your credibility but also that of your evaluation projects.
- **standard conformity**: All is summarised in the aforementioned ethical and methodological standards. If you adhere to them, less can go wrong with your evaluations.

- **resource allocation**: Evaluators need working hours to fulfil their tasks. Only if the appropriate resources are provided, the evaluation can receive the working time that it deserves.

Any evaluation project is more likely to run smoothly and successfully, if you as a quality assurance officer keep these factors in mind.

4. Empirical Social Research

Approaches, strategies and designs

Since we partly base serious decisions on evaluations (see above), we need to pay special attention to the quality of our evaluation results. The results have to meet scientific standards to be comprehensive and useful for the purpose of our evaluations, namely the improvement of our activities (development paradigm). To guarantee such scientific quality, professional evaluators make use of the standards and methods of empirical social research. Empirical social research is a way of gaining knowledge using empirical evidence by means of direct and indirect observation or experience. It is an important part of evaluation practice. And we are convinced that it can support your work as a QAO to know some of the basic rules and catchwords.
Empirical social research fills entire study programmes, which is why we will only refer to a few defining characteristics in the following text and in the respective screencasts of the module. However, more comprehensive literature on standards and methods of empirical social research can be found in the bookshelf of this module.
There are different kinds of empirical studies. We can use the following areas to characterise them in more detail: the approach of the study, the research strategy and the research design (but note: these labels are sometimes used synonymously).

Research approach

The research approach can be classified into explorative studies, descriptive studies and explanatory studies:

- **explorative studies** give us a first insight into new areas of research. They do not want to describe phenomena in detail or explain causalities. They are not used to test hypotheses about given phenomena. Since they explore new ground, they often serve as the starting point for further in-depth research. Explorative studies are employed, whenever little is known about our research object and it is difficult to formulate hypotheses.
- **descriptive studies** describe phenomena and provide detailed information about our research object. They can be used to verify of falsify our hypotheses, but they do not want to investigate the causalities of given phenomena.
- **explanatory studies** explain causalities and provide information about backgrounds, causes and connections between certain phenomena.

Research strategy

With regard to the research strategy, we differentiate cross-sectional studies and longitudinal studies:

- **cross-sectional studies**: In a cross-sectional study data is collected at one particular point of time. The data we receive shows a “snapshot” of the characteristics of our target population at that particular time. A survey among undergraduates and

graduates asking them about their perception of their study conditions one time at the end of the semester is an example for a cross-sectional study.

- **longitudinal studies**: Longitudinal studies allow us to observe a development over time concerning our research object/our target population. Here, data is collected at more than one point of time. That means, repetitive observations of the same research object are conducted over a longer period of time, as for example in graduate tracer studies, where graduates are surveyed immediately after graduation, one year after graduation and five years after graduation.

Research design

The research design determines the overall character of data collection and the framework of methods and tools applied to collect the data. As for the overall character, we can differentiate quantitative and qualitative studies:

- **quantitative studies**: … measure phenomena by the help of statistical, mathematical or computational techniques. We rather deal with numbers than with words. This makes the research process structured and rigid. Usually, we use quantitative research, if we want to reach larger groups of respondents, because a quantitative research design helps us to calculate lager amounts of data.
- **qualitative studies**: … collect and interpret non-numerical data, for example in interviews. The research process is not that structured and allows for more flexibility. Usually, with qualitative research designs we can reach less respondents, but we usually can investigate our research object in more depth.

The research design also determines the method of data collection. Here, we differentiate between structured, semi-structured and unstructured designs. We also differentiate oral and written modes of data collection. Table 2 (Pistor/Stammen, 2005, p. 73) gives an overview of the different modes of data collection and their characteristics. For more detailed information refer to the screencast *What is Evaluation?*.

Survey	Mode	Example	Characteristics
structured	oral	guided individual (or group) interview, telephone survey	• rigid structure, rigid contents, rigid questions and wording • To collect quantifiable data and aspects • „measure“
	written	mail survey, online survey	
semi-structured	oral	guided discussion, group interview	• different levels of flexibility and specifity • to collect qualitative and/or qualitative data and aspects • „interpret“ and/or „measure“
	written	expert survey	
unstructured	oral	expert survey, group discussion	• flexible structure, flexible contents, flexible questions and wording • To collect qualitative data and aspects • „interpret“
	written	informal survey of experts	

Table 2: Overview of data collection methods.

5. How to make things measurable

Three questions at the beginning of each research project

Before any evaluation project starts, evaluators should be very clear about the purpose of the evaluation and how the evaluation shall be implemented. Therefore, evaluators should answer the following questions at the very beginning of their evaluation:

- Why do we want to conduct the evaluation? (= purpose)
- What do we want to find out by the evaluation? (= research question)
- How do we want to conduct the evaluation? (= research method)

Answering the questions why, what and how can help us to frame our evaluation project at the starting point.

Any empirical study - thus also evaluations - follows more or less the same process. We can subdivide the process into the following phases (cf. Pistor/Stammen 2015):

1. **orientation and definition:** We clarify the problem the research should have an answer to, we review literature on concepts and previous research results and formulate questions on the basis of our problem clarification and literature review.
2. **preparation:** We choose an appropriate research design, decide on appropriate methods of data collection, define the target population and our sample (= part of the totality of our target group).
3. **data collection:** We develop an instrument for data collection, e.g. a questionnaire or an interview guideline, test the instrument, finally collect our data for example by an online survey or a guided interview.
4. **data analysis:** We clean the data, i.e. we detect and correct (or remove) inaccurate records from our database. Both in quantitative and qualitative research, raw-data has to be coded. This means for qualitative studies, the data has to be organised in categories first. By doing this, also qualitative data can later be translated into numbers. After coding, the data is ready for analysis.
5. **reporting:** Before a report can be written, we – although neutral – conduct an initial interpretation of the results of the evaluation research project. Then we compile and disseminate the report.
6. **use of data:** For the use of evaluation results in Higher Education Institutions we usually bring together relevant stakeholders, who interpret and reflect on data and develop improvement measures (cf. also the reflection phase of the PDCA cycle in module 2).

Operationalisation

For any evaluation research project, it is important to operationalise our mental concepts, i.e. to transfer them from abstract concepts to variables in our evaluation research project. Concepts are our ideas, notions and beliefs. They are abstract and only mental creations. Concepts are the ideas and images that come to our minds, when we hear the word that denotes the concept.

Operationalisation helps us to

a) make something abstract measurable
b) make our evaluation valid, because we create a common understanding between our respondents and ourselves

With regard to a): How would you define *study success*? Personal feeling of success? Good grades? Well-paid job after graduation? Number of job promotions? And how would you

measure it? While *study success* is not directly observable and measurable, the last three mentioned operationalisations (grades, income, promotions) are.

Concerning b): Concepts may vary "from brain to brain". If we would only ask our students in a survey, whether the lecture was *good* and they would respond that it was good, we might draw the wrong conclusion, because the mental concept of a good lecture of our students might be: simple, entertaining, little homework. But our mental concept of a good lecture may be: structured, variety of teaching methods, assignment aligned to intended learning outcome (cf. also module 4). Thus, if we do not operationalise our mental concepts, we might receive answers to questions, we have not even asked. In the process of operationalisation we usually see the following levels: concept → dimensions → variables (that can have different values).

- **concept**: A concept is a mental creation (see above).
- **dimension**: A dimension is a specifiable aspect of an abstract concept. For example, the concept *personal satisfaction* could be defined by the dimensions *satisfaction with family*, *job satisfaction*, *satisfaction with physical health*.
- **variable and values**: Variables are characteristics that vary from research object to research object. Variables have values/attributes that describe the characteristics. For example: The variable *hair colour* can have the values *blond*, *red*, *brown*, *black* (and literally all other colours). We can differentiate independent and dependent variables. Independent variables are variables that can have an effect on dependent variables. Dependent variables are those variables which we observe and measure to determine the effect of the independent variable.

Quality criteria of measurements

You surely have already heard about validity, reliability and objectivity. These are quality criteria of the measurements and – since they might become important in a discussion about evaluation results at your institution, as a QAO you should know them.

- **validity** is given, if our instrument, e.g. our questionnaire, measures what it was supposed to measure. For example, a course evaluation should measure the didactic quality of a course, not the popularity of the teacher.
- **reliability**: Your instrument is reliable, if it produces the same results, when the measurement is repeated under the same circumstances. If you put a 5-pound-weight on a scale repeatedly and the scale will show 5 pounds every time, it is reliable. Reliability only refers to the reliability of the instrument in terms of reproducibility of measurements. An instrument can be reliable, although it is not valid.
- **objectivity**: Objectivity refers to the independency of an instrument. An evaluation is objective, if it is not influenced by the beliefs, values, attitudes and the assumptions of the evaluator.

6. How to Use Evaluation Results

Reporting, interpreting, discussing

According to me it does not make any sense to collect data at great expense and then leave them to rot in so-called "data graveyards". Evaluation takes time and money. It can be very powerful to support the development of individuals, study programmes and organisations as a whole. Therefore, it is essential to use evaluation results. Prerequisites for the use of data are

a) the data is of good quality (see standards for evaluation) and
b) the data is provided in a way appropriate for the data users and use.

As QAO or evaluator you can contribute to the usability and use of data by the following activities (cf. BetterEvaluation 2018):

- Identification of primary intended users.
- Identification of the primary intended use.
- Identification of reporting requirements.
- Guiding the process of data use.

In fact, the users and the intended use of evaluation results should be clear before you start your evaluation (see above), because they will affect the design of your evaluation. In addition to that, also reporting requirements will have an influence on your data collection phase.

It is important to identify the stakeholders inside and outside your organisation, who will actually use the evaluation results for decision-making. According to the charity BetterEvaluation (2009-2020), these stakeholders are "the specific people, in a specific position, in a specific organization who will use the evaluation findings and who have the capacity to effect change (for example, change policies and procedures, improve management strategies). Who they are will depend on your evaluation." If possible, you should involve them in the preparation of your evaluation, because that makes it more likely that the evaluation meets their needs and that the results will be used in the end (ibid.). The overall purpose of an evaluation will influence the evaluation schedule, resources, stakeholders involved and the evaluation design, implementation, context and impact (cf. ibid.).

- Knowledge about primary users and primary use of your evaluation will help you to define the reporting requirements. Answering the following concrete questions can help you to decide how the evaluation results can be compiled in a report and how they can be disseminated: Who needs to be updated?
- What decisions do they need to make?
- When do they need to make these decisions?
- What information do they need to make these decisions?
- What format do they need this information in? (cf. ibid.)

In the toolbox of module 3: *Evaluation as an Integral Part of Quality Assurance*, you will find some checklists and tools that can help you to identify users, uses and reporting requirements of your evaluation.

As a Quality Assurance Officer in a Higher Education Institution, also guiding the process of data use may fall into your responsibility, i.e. preparing, organising and documenting all processes around the interpretation of evaluation findings and the deduction of improvement measures. In this capacity you will have to make sure that the evaluation results are provided in a form appropriate to the event. For example, you might want to prepare a powerpoint presentation rather than a very long report, because it will be easier to present immediately before the discussion in – let's say – a faculty board meeting begins, thus everybody's memory will still be fresh. In addition, it will be your responsibility to ensure that improvement measures are discussed and agreed on in a correct and sustainable form, e.g. in the framework of an undersigned target agreement between a faculty and the rectorate. In the cinema of this module, you will find examples on how evaluation results can be used in a Higher Education Institution.

Reflective Questions

What do you think?

- What kind of evaluations do you already conduct in your Higher Education Institution?
- Do you think they already respect the mentioned standards of evaluation? What could be done even better?
- What challenges for quality managers might occur with the application of evaluation standards?

References

And recommendations for further reading

AfrEA (2000): The African Evaluation Guidelines. 2020 Version. Draft. Online https://afrea.org/ (last access on 02.08.2020).

Balzer, L. (2005): Wie werden Evaluationsprojekte erfolgreich? - Ein integrierender theoretischer Ansatz und eine empirische Studie zum Evaluationsprozess.

BetterEvaluation (2009 - 2020): Online www.betterevaluation.org (last access on 02.08.2020).

DeGEval – Deutsche Gesellschaft für Evaluation (2016): Standards für Evaluation. Erste Revision. Kurzfassung.

Komrey, H. (2005): Evaluation. Ein Überblick. In: Heidrun Schöch (Ed.): Was ist Qualität. Die Entzauberung eines Mythos. Pp. 31-85).

Pistor, P. & Stammen, K.-H. (2015): Tools and Procedures for Quality Assurance in Higher Education. Training on Internal Quality Assurance Series. Module 2.

Rossi, H. & Freeman, E. (1999): Evaluation. A Systematic Approach. 6th Edition.

Stockmann, R. (2004): Was ist eine gute Evaluation? Einführung zu Funktionen und Methoden von Evaluationsverfahren.

Torres, R., Preskill, H., & Piontek, M.E. (2005). Evaluation strategies for communicating and reporting, enhancing learning in organizations. 2nd Edition.

Yarbrough, D.B., Shula, L.M., Hopson, R.K., & Caruthers, F.A. (2010): (= JCSEE). *The Program Evaluation Standards: A guide for evaluators and evaluation users.* 3rd Edition.