Pistor, Petra; Stammen, Karl-Heinz:

Tools and Procedures for Quality Assurance in Higher Education Institutions - Training on Internal Quality Assurance Series | Module 2

In: Training on Internal Quality Assurance Series

DOI: http://dx.doi.org/10.17185/duepublico/43223

URN: urn:nbn:de:hbz:464-20170215-100037-2

Link: http://duepublico.uni-duisburg-essen.de/servlets/DocumentServlet?id=43223

Petra Pistor and Karl-Heinz Stammen

Tools and Procedures for Quality Assurance in Higher Education Institutions

Training on Internal Quality Assurance Series | Module 2

Solveig Randhahn and Frank Niedermeier (Eds.)

With financial support from the

**Federal Ministry
for Economic Cooperation
and Development**

Imprint

Please cite the use of our course book series in presentations, trainings, papers etc. according to scientific standards. You can cite this book as:

Pistor, P. & Stammen, K.-H. (2017). Tools and Procedures for Quality Assurance in Higher Education Institutions. Module 2. In Randhahn, S. & Niedermeier, F. (Eds.) *Training on Internal Quality Assurance Series*. Duisburg/Essen: DuEPublico. Retrieved from: http://dx.doi.org/10.17185/duepublico/43223

Acknowledgment

Our modules and books have been prepared and written in a joint effort of the University of Duisburg-Essen and the University of Potsdam under the DIES (Dialogue on Innovative Higher Education Strategies) Programme conducted by the German Academic Exchange Service (DAAD) and the German Rectors' Conference (HRK) with funds from the Federal Ministry for Economic Cooperation and Development (BMZ). We take this opportunity to thank the DIES Programme and all the partners from Africa, Europe and Southeast Asia who were involved in the development process and express our deepest gratitude for the received support, without which the modules and course books would not have been possible to realise.

We want to further express our sincere gratitude for the most valuable support and contributions from the involved partners

- Autorité Nationale d'Assurance Qualité de l'Enseignement Supérieur (ANAQ-Sup), Sénégal,
- ASEAN Quality Assurance Network (AQAN),
- ASEAN University Network (AUN),
- Association of African Universities (AAU),
- Conseil Africain et Malgache pour l'Enseignement Supérieur (CAMES),
- European Association for Quality Assurance in Higher Education (ENQA),
- National Accreditation Board (NAB), Ghana,
- National Council for Tertiary Education (NCTE), Ghana,
- National Universities Commission (NUC), Nigeria,
- Southeast Asian Ministers of Education Organization Regional Centre for Higher Education and Development (SEAMEO RIHED),
- United Nations Educational, Scientific and Culutral Organization (UNESCO),
- University of Professional Studies Accra (UPSA), Ghana

and especially from Prof. Dr. Shahrir Abdullah, Richard Adjei, Prof. Dr. Goski Bortiorkor Alabi, Prof. Dr. Bassey Antia, Prof. Dr. Arnulfo Azcarraga, Gudrun Chazotte, Assoc. Prof. Dr. Tan Kay Chuan, Kwame Dattey, Prof. Dr. Ong Duu Sheng, Prof. Zita Mohd. Fahmi, Mae Fastner, Assoc. Prof. Dr. Nantana Gajaseni, Robina Geupel, Josep Grifoll, Juliane Hauschulz, Dr. Pascal Hoba, Dato' Syed Hussein, Benjamin Jung, Prof. Abdel Karim Koumare, Dr. Vipat Kuruchittham, Prof. Dr. Chiedu Mafiana, Prof. Dr. Duwiejua Mahama, Barbara Michalk, Prof. Dr. Le Quang Minh, Nguyen My Ngoc, Johnson Ong Chee Bin, Concepcion V. Pijano, Prof. Dr. Philipp Pohlenz, Sonja Pohlmann, Dr. Suleiman Ramon-Yusuf, Dr. Sylvia Ruschin, Dr. Chantavit Sujatanond, Dr. Oliver Vettori and Marc Wilde.

The authors

Petra Pistor

Centre for Higher Education Development and
Quality Enhancement (CHEDQE)

University of Duisburg-Essen, Germany

petra.pistor@uni-due.de
https://www.uni-due.de/zfh/

Petra Pistor currently acts as Deputy Managing Director of
the Centre for Higher Education Development and Quality
Enhancement (CHEDQE) at the University of Duisburg-Essen
(UDE). Her responsibility here lies in two of the centre's five
teams: data management and evaluation. Petra Pistor has
worked on a variety of evaluation projects in higher education.
Her working activities are focused on different evaluation
projects and trainings in the field of higher education quality
management both in Germany as well as internationally,
especially in East and also West Africa. As a freelancer she
works as consultant and trainer in various contexts for higher
education institutions.

Karl-Heinz Stammen

Centre for Higher Education Development and
Quality Enhancement (CHEDQE)

University of Duisburg-Essen, Germany

karl-heinz.stammen@uni-due.de
https://www.uni-due.de/zfh/

Karl-Heinz Stammen has a Graduate Diploma in Social Science and completed teacher training. Since 2005, he has worked as a research associate at the Centre for Higher Education Development and Quality Enhancement (CHEDQE) of the University of Duisburg-Essen, where he heads the data management team.

Amongst others, Karl-Heinz Stammen coordinates the student panel, course evaluation and graduate tracer studies of the university. His working activities are focused on the design and realisation of different higher education and student research projects as well as trainings in the field of quality assurance and enhancement in higher education especially in East Africa.

List of Abbreviations

DeGEval   Gesellschaft für Evaluation (German Evaluation Society)

EHEA   European Higher Education Area

ESG   Standards and Guidelines for Quality Assurance in the EHEA

HEI   Higher Education Institution

JCSEE   Joint Committee on Standards for Educational Evaluation

KIT   Karlsruher Institut für Technologie

PDCA cycle   Plan-Do-Check-Act cycle

QA   Quality Assurance

QMS   Quality Management System

SSC   Science Support Center

UDE   University of Duisburg-Essen

UP   University of Potsdam

Table of Contents

Preface

Introduction to the Module

Prerequisites for the Module
- Learners are familiar with different notions of quality and shortcomings in defining quality,
- they know the PDCA-cycle and different models for defining quality and quality assurance,
- they know the special circumstances of non-profit and particularly educational institutions which have an influence on quality concepts and quality assurance.

Intentions of the Module

In the study material to Module 2 first of all the concept of "evaluation" is defined and discussed taking into account the special circumstances at higher education institutions. As far as we define evaluation, it is an integral part of quality management, namely the continuous collection of information relevant for monitoring and enhancement – so to speak the "check"-part of the PDCA-cycle (plan, do, check, act). In this course book, readers will be made aware of the possible shortcomings of evaluation in higher education and will be given practical hints for enhancing the quality of evaluative processes. All this will be reflected taking into account national and international standards for evaluation and quality assurance.

This course book also deals with basic knowledge of empirical social research methodology as well as technical demands and procedures of particular instruments and processes in the field of quality management. This knowledge is very important since methods and tools of empirical social research are the toolbox for obtaining data and information in the bounds of evaluations. Hence, the precise conception and systematic conduct of qualitative and quantitative data collection as well as data analysis and interpretation are core elements of this study material.

Quality managers are confronted with various tasks in the field of evaluation: designing an evaluative process that fits to the organisational and cultural framework of a higher education institution, keeping in mind the combination of evaluations with other QA-tools (e.g. data management) and the integration into a QA-system, designing and applying methods and tools in order to obtain relevant information, etc. quality managers act as coordinators, facilitators, mediators, social scientists (although they seldom hold a degree in social sciences) and sometimes even as psychologists (because quality management and evaluation are sensitive processes). Thus, the role of quality managers is not only multifaceted, but also demanding. In the bounds of the study material on hand we will therefore reflect on the different facets of this role, whenever appropriate.

On successful completion of the module, you should be able to...

- apply the knowledge gained about prerequisites, paradigms, standards and typical shortcomings in evaluations in higher education when organising an evaluative process at your own institutions,
- take into account the basic principles of qualitative and quantitative empirical social research methods when designing your own instruments for data collection, e.g. surveys and interviews, and interpreting your own data,
- know what to take into account, when interpreting data,
- apply basic principles for data reporting,
- apply these instruments in regard to the particular characteristics and framework of your own institutions.

Chapter 1

Introduction to Theory and Concepts of Evaluation

On successful completion of this chapter, you should be able to…

■ define the term "evaluation" and its relationship to assessment, analysis and checking,

■ describe evaluative processes along the lines of their underlying paradigms, intentions and perspectives of analysis,

■ describe standards for evaluation and take them into account for your everyday work as quality managers.

1 Introduction to Theory and Concepts of Evaluation

1.1 Introduction

The term "evaluation" is – also in the context of higher education – resounded in the whole world. But what does it stand for in different contexts and what do variable meanings in different contexts imply for the realisation of evaluations? At the beginning, this course book will give you a manageable definition of the term "evaluation" and introduce different paradigms, concepts and standards that determine the character of evaluative processes. After that, special features and potential areas of conflict of evaluations in higher education institutions will be discussed and practice-oriented recommendations will be deduced for the implementation of self-evaluations and external evaluations (also referred to as peer reviews) as instruments for quality assurance and quality enhancement.

According to our understanding, evaluations go beyond sheer assessment. In addition to data collection and interpretation – to what we would refer to as assessment – evaluations comprise the deduction of improvement measures as key element and thus a process of decision-making.

Therefore, the availability of relevant information is crucial for evaluations, because they support faculties, quality managers and higher education leadership with their decision-making processes. And this is why it is of utmost importance that this information is comprehensive and true (valid). For the collection of relevant and valid information, evaluations use methods of social empirical research, which make sure that the information collected mirror the reality correctly and accurately and therefore are suitable to form a basis for discussion and the deduction of improvement measures.

Quality assurance units at higher education institutions should therefore ideally have staff with expertise both in the field of evaluation theory, higher education management and in methods of empirical social research, because it is equally important to know the superordinate aim of an evaluation and its connection to strategic plans (evaluation and higher education management) as it is to know opportunities and limits of data collection instruments that should be used for that purpose (empirical social research). In reality, however, oftentimes all three fields of qualification must be represented by one staff member employed as quality manager. With the course book on hand we invite you therefore to get familiar with the basics in evaluation theory and social empirical research methodology.

1.2 Evaluation – Concept Definition

In everyday language we refer to the process of assessing something by the term "evaluation" (from Latin "valere" = strong, being of value). But with this mere description, nothing is conveyed about the actors involved, the entity to be analysed, the degree of scientific foundation of the process, etc.

A variety of different definitions and concepts can be subsumed under the heading "evaluation". A basic classification of different approaches to the concept will be given in the following. As a fundamental agreement for this course book, we will adhere to the following definition:

Definition of evaluation and assessment

Evaluation is the systematic description and assessment of particular phenomena on the basis of explicit or implicit criteria. It supports the process of decision-making about development measures, which in turn can be approached by evaluative processes (see the PDCA-cycle in Module 1).

Evaluations use methods of empirical social research, as will be executed in the following chapters. For the study material on hand, evaluations will be understood as the process encompassing data collection, data interpretation and the deduction of development measures. The terms "evaluation" and "assessment", however, are oftentimes used synonymously, but in this course book we will differentiate them following Charles Secolsky and Brian Denison:

> *Assessment = "... collection, analysis and interpretation of information related to a particular issue or outcome of interest."*
>
> *Evaluation = "... determining the worth, value or effectiveness of something – often some kind of program."*
> ***(Secolsky & Denison 2012, 18)***

Thus, assessment is an integral part of evaluation.

1.3 Evaluation Paradigms

Evaluations can have different interconnected functions. In other words, they can be executed following different paradigms. As a rule, evaluations never follow one of these paradigms individually, but rather adhere to a bundle of them with more or less emphasis. Generally, four different evaluation paradigms are distinguished in the recent discussion (Stockmann 2004):

a) Knowledge-gain paradigm – evaluation for the provision of knowledge
In the first place, evaluations serve the purpose of generating information about particular phenomena, entities or processes, e.g. as a study in the framework of a research project. Evaluations are not an end in themselves, but the information they provide shall be of use for the questioner, e.g. the responsible person or the target group of a process. For example, it could be of interest, whether a programme (e.g. a study programme) works smoothly, what the demands of a given target group are or if any measures applied matched the demands of a target group. Regarded through the lenses of the knowledge-gain paradigm, evaluations are of use to produce insight into the nature of something. Only in a second step, they are perceived to be the basis for analysis and the deduction of measures.

b) Control paradigm – evaluation for execution of control

Evaluations can also provide information about the success of a measure, a programme etc. They can reveal information about if and why or why not an activity led to the intended outcome. Thus, evaluation can easily be linked to control. In saying that, we refer to control as the process of checking the degree of having reached a particular aim, not the checking of achievements of individuals or groups. In our understanding, the latter is not very likely to contribute fruitfully to the development of quality.

c) Legitimation paradigm – evaluation for documenting achievements

Any measure or programme usually follows an aim and should be oriented towards the demands and expectations of respective stakeholders (e.g. target group of the measure or initiator of a programme). Evaluations make it possible to generate information about the achievement of objectives and the adequacy of measures. Evaluations also allow the documentation of resources needed and the process as a whole. This documentation can comprise information about the productivity and success of a measure and can be given to all relevant stakeholders as proof.

d) Development paradigm – evaluation as a starting point for change

Evaluations provide knowledge (see a). But the compilation of information and the generation of knowledge about a topic is not an objective in its own. In fact, the information collected and the knowledge gained can serve as a basis for analysis and the evolvement of deep insight (causalities and backgrounds) into a subject matter and the deduction of improvement measures. Information and data gathered during an evaluation can be the fundament of the dialogue between different stakeholders and the person in charge of a measure. This dialogue provides the opportunity to analyse the information and data-base respecting different perspectives and to jointly develop adequate measures for improvement.

"Hidden Agendas" – Evaluations can also be (mis)used as a tactical move. They can be used to justify political decisions already made – and one or other evaluation has to face this objection. These kind of pseudo-evaluations do not follow the objectives described above, but are conducted, because evaluations have become a fashionable tool for business development and higher education management. Thus, they are very suitable to serve as a sheer marketing tool for an institution or certain actors. One can argue that such evaluations following the "tactics paradigm" are a legitimate tool for policy making. Taking into account the resources usually spent for evaluative processes, one can easily invalidate this argument. Evaluations that are only conducted out of political reasons, do not justify the use of these resources. (Rossi, Lipsey, and Freeman 2004, 37)

1.4   Evaluation Concepts

While in Chapter 1.3 paradigms of evaluations have been described, the different procedural concepts for evaluations will be discussed in the following.

Whenever it comes to the assessment of measures or bundles of measures (programmes), evaluations can follow different objectives. This has an influence on their underlying concept and the moment of their conduction.

For example, evaluations can support the planning process for a measure by providing information about the target group or the framework conditions of a measure. When evaluations are used for this purpose, they are conducted before the beginning of a planned measure and are called ex-ante (lat. "from before") evaluations. That is: Evaluations that take place before the programme or action to be evaluated takes place.

If evaluations shall accompany and describe the process of a measure or a programme, they are referred to as interim (lat. "in between") evaluations. That is: Evaluations that take place during the course of an action or programme.

Finally, evaluations that follow the aim to generate information about the effectiveness of a measure after it is finished, are called ex-post (lat. "from after") evaluations. That is: Evaluations that take place after an action or programme has taken place.

The terms ex-ante, interim and ex-post denominate the **point of time** or the **perspective of analysis** (anticipatory, accompanying, retrospective) of an evaluation.

Formative and summative evaluations

The terms formative and summative on the contrary denominate the underlying concept, the character of an evaluation: Formative evaluations support the process of the measure to be evaluated – e.g. by doing assessments at different points of time during the process - that is: **Formative evaluations** can be described as "evaluative activities undertaken to furnish information that will guide program improvement" (Rossi, Lipsey, and Freeman 2004, 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."* **(Rossi, Lipsey, and Freeman 2004, 65).**

The following matrix depicts the interconnections of the two ways of description:

Phase of a Measure or Programme	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   *Concept and perspective of analysis of evaluations (adapted from Stockmann 2004, 6)*

1.5   Evaluation Standards

On-going professionalisation in the field of evaluation research, starting in the 1970s in the US and Europe, has brought about institutions and committees that deal with good practice and standards in evaluation work. For example, the Joint Committee on Standards for Educational Evaluation (JCSEE)[1] has developed and continuously revised a bundle of standards for evaluations also in the education sector that have been adapted by other associations as well (see, for example, the Standards for Evaluation by the German Association of Evaluation: DeGEval 2001).

These evaluation standards combine procedural and ethical principles and shall support the quality assurance of evaluations as well as give evaluators orientation for their work. The JSCEE formulates the following standards and guidelines for quality assured evaluations that can be subdivided into five conceptual packages (Yarborough et al. 2011):

---

1 Created in 1975, the Joint Committee is a coalition of major professional associations (e.g. American Educational Research Association, National Council on Measurement in Education etc.) concerned with the quality of evaluation. The Joint Committee is housed at the Center for Evaluation and Assessment, University of Iowa.

Utility Standards *The utility standards are intended to increase the extent to which programme stakeholders find evaluation processes and products valuable in meeting their needs.*		
U1	Evaluator credibility	Evaluations should be conducted by qualified people who establish and maintain credibility in the evaluation context.
U2	Attention to stake-holders	Evaluations should devote attention to the full range of individuals and groups invested in the programme and affected by its evaluation.
U3	Negotiated purposes	Evaluation purposes should be identified and continually negotiated based on the needs of stakeholders.
U4	Explicit values	Evaluations should clarify and specify the individual and cultural values underpinning purposes, processes, and judgements.
U5	Relevant information	Evaluations should clarify and specify the individual and cultural values underpinning purposes, processes, and judgements.
U6	Meaningful processes and products	Evaluations should construct activities, descriptions, and judgements in ways that encourage participants to rediscover, reinterpret, or revise their understandings and behaviours.
U7	Timely and appropriate communicating and reporting	Evaluations should attend to the continuing information needs of their multiple audiences.
U8	Concern for consequences and influence	Evaluations should promote responsible and adaptive use while guarding against unintended negative consequences and misuse.

Table 2   *JCSEE's Utility Standards.*

Feasibility Standards *The feasibility standards are intended to increase evaluation effectiveness and efficiency.*		
F1	Project manage-ment	Evaluations should use effective project management strategies.
F2	Practical procedures	Evaluation procedures should be practical and responsive to the way the programme operates.
F3	Contextual viability	Evaluations should recognize, monitor, and balance the cultural and political interests and needs of individuals and groups.
F4	Resource use	Evaluations should use resources effectively and efficiently.

Table 3   *JCSEE's Feasibility Standards*

Propriety Standards		
The propriety standards support what is proper, fair, legal, right and just in evaluations		
P1	Responsive and inclusive orientation	Evaluations should be responsive to stakeholders and their communities.
P2	Formal agreements	Evaluation agreements should be negotiated to make obligations explicit and take into account the needs, expectations, and cultural contexts of clients and other stakeholders.
P3	Human rights and respect	Evaluations should be designed and conducted to protect human and legal rights and maintain the dignity of participants and other stakeholders.
P4	Clarity and fairness	Evaluations should be understandable and fair in addressing stakeholder needs and purposes.
P5	Transparency and disclosure	Evaluations should provide complete descriptions of findings, limitations, and conclusions to all stakeholders, unless doing so would violate legal and propriety obligations.
P6	Conflicts of interests	Evaluations should openly and honestly identify and address real or perceived conflicts of interests that may compromise the evaluation.
P7	Fiscal responsibility	Evaluations should account for all expended resources and comply with sound fiscal procedures and processes.

Table 4   *JCSEE's Propriety Standards.*

Accuracy Standards		
The accuracy standards are intended to increase the dependability and truthfulness of evaluation representations, propositions, and findings, especially those that support interpretations and judgements about quality.		
A1	Justified conclusions and decisions	Evaluation conclusions and decisions should be explicitly justified in the cultures and contexts where they have consequences.
A2	Valid information	Evaluation information should serve the intended purposes and support valid interpretations.
A3	Reliable information	Evaluation procedures should yield sufficiently dependable and consistent information for the intended uses.
A4	Explicit programme and context descriptions	Evaluations should document programmes and their contexts with appropriate detail and scope for the evaluation purposes.
A5	Information management	Evaluations should employ systematic information collection, review, verification, and storage methods.

Accuracy Standards *The accuracy standards are intended to increase the dependability and truthfulness of evaluation representations, propositions, and findings, especially those that support interpretations and judgements about quality.*		
A6	Sound designs and analyses	Evaluations should employ technically adequate designs and analyses that are appropriate for the evaluation purposes.
A7	Explicit evaluation reasoning	Evaluation reasoning leading from information and analyses to findings, interpretations, conclusions, and judgements should be clearly and completely documented.
A8	Communication and reporting	Evaluation communications should have adequate scope and guard against misconceptions, biases, distortions, and errors.

Table 5   *JCSEE's Accuracy Standards*

Evaluation Accountability Standards *The evaluation accountability standards encourage adequate documentation of evaluations and a meta-evaluative perspective focused on improvement and accountability for evaluation processes and products.*		
E1	Evaluation documentation	Evaluations should fully document their negotiated purposes and implemented designs, procedures, data, and outcomes.
E2	Internal meta-evaluation	Evaluators should use these and other applicable standards to examine the accountability of the evaluation design, procedures employed, information collected, and outcomes.
E3	External meta-evaluation	Programme evaluation sponsors, clients, evaluators, and other stakeholders should encourage the conduct of external meta-evaluations using these and other applicable standards.

Table 6   *JCSEE's Evaluation Accountability Standards*

? Questions & Assignments

1. According to which paradigm have evaluative processes at your higher education institution been conducted? Do you see any assets and drawbacks connected to the different paradigms?
2. Do you think, the JCSEE-Standards for evaluation are sensible and helpful? Why or why not?
3. What challenges for quality managers might occur with the application of evaluation standards?

Chapter 2

Evaluation at
Higher Education Institutions

⌖ On successful completion of this chapter, you should be able to...

- explain the characteristics of internal, external, self- and peer-evaluation,
- plan an evaluation in a sensitive manner to the stakeholders involved, aligned to the typical phases of an evaluative process, taking into account typical potential obstacles,
- compare assets and drawbacks of different case studies of evaluations in the field of teaching and learning,
- describe the different factors for successful evaluations.

2   Evaluation at Higher Education Institutions

2.1   Different Forms of Quality

As discussed in Module 1, the definition of the concept of quality poses a challenge to non-profit organisations, service organisations and above all to educational organisations. According to Nickel (2010), one can find different qualities in higher education institutions:

- **Quality of objectives** in research, study and teaching as well as in the field of community outreach;
- **Quality of outputs** (results) of study programmes and research activities;
- **Quality of outcomes** (effects) of study programmes, research activities and community services;
- **Quality of internal processes** (to reach the aforementioned objectives, outputs and outcomes).

Additionally, educational organisations differ from other forms of organisations by some inherent specialties that make the definition and assessment of quality very difficult (Röbken 2012; Stockmann 2006):

- **Vagueness of objectives**: While other kinds of organisations follow clear-cut and transparent aims (hospitals heal people, factories produce goods and maximise their profit, etc.), higher education institutions are geared to only vaguely defined aims. Is the objective of a higher education institution the personal development of people, the transfer of recent research results to the community, the preparation of people for the job market, or are there even other aims, a higher education institution can follow?

    Differences of quality concepts between industry/economy and educational institutions

- **Variety of stakeholders**: In the field of higher education a variety of target groups and key actors pose their demands on the quality of higher education products and processes. Students, their parents, the state, teachers and researchers have different expectations of higher education institutions. Thus, they will assess the offer of a HEI along the lines of different implicit criteria.
- **Client involvement**: The success of educational programmes is highly influenced by the addressees of the course and thus their quality is only partly controllable by the educational organisation itself.
- **Problematic technologies**: In contrast to product manufacturers, educational organisations can hardly or not at all standardise the "production steps" of education. How, for example, can a teaching and learning process that is heavily dependent on the personalities of the interacting teachers and learner groups, be subdivided into distinct process steps and be standardised?
- **Subject relation**: At higher education institutions experts in a variety of different subjects work together (expert organisations). The definition of quality can differ from subject to subject. For example, a lesson in the natural sciences that only involves the reading of books would be assessed as of being lower quality, whereas a lesson in the humanities would be assessed as of good quality in the same case.

The aforementioned factors make it extremely difficult to define when one of the various activities of a higher education institution can be assessed as good. The evaluation of an item, a programme or product always involves the presence of criteria and standards against which assessments can be made. Different from indus-

try and the economy, the definition of standards in higher education is not an easy task, as you can see from the following explanations.

A criterion in this context is a feature of an item, a measure or a programme that allows one to make statements about its quality. Waterproofness, for example, is a criterion for outdoor jackets. If the wearer stays dry after being flooded by 100 litres of water, the jacket can be assessed as good (in the respect of watertightness). Another criterion for the quality of outdoor jackets can be windproofness, yet another thermal insulation etc.

A standard prescribes the level to which a criterion should be attained in order to be assessed as good. For example, the criterion waterproofness of an outdoor jacket would be achieved if the inside stays dry after 30 minutes in heavy rain. "Thirty minutes in heavy rain" would be the standard for the criterion waterproofness.

Criteria and standards

Only very rarely, are standards defined in the field of higher education (although sometimes the term "standard" is used misleadingly). A quality standard according to the definition above in the context of higher education would read: "The quality of a study programme is good, if not more than 30 students are taught by one person." In reality, quality in the higher education sector more often is assessed by the use of open criteria. Oftentimes you find formulations like "the ratio of students and teaching personnel is adequate" or "enough literature resources are provided for the students". The definitions or rather: The translations of the terms "adequate" and "sufficient" into observable or measureable entities lie in the eye of the beholder. In other words, to be able to exercise valid and relevant assessments of quality, the actors involved must agree on standards that they think are a sign of the good quality of an item, measure or programme. There are no universal standards for that. Usually, this definition of criteria and standards is done by getting inspired by how others (other higher education institutions) use them and adapting them to their own framework conditions after discussion with all relevant stakeholders.

Quality definitions and requirements of different stakeholders can also be contradictory! While for the institution a high graduate ratio and fast-to-finish study programmes may be a quality feature, students may regard study programmes with some freedom and options to choose from (not to mention a degree which can be completed in the regular time) as higher in quality.

Last but not least, the definition of criteria and standards is also influenced by the perspective of analysis and the character of an assessment or an evaluation. If it is an ex-post evaluation that should assess the degree of how an objective is being achieved, a target-performance comparison on the basis of key performance indicators is possible (evaluation following the control paradigm). If a higher education institution, for example, has the hypothesis that the number of participants of a given seminar is low, because its timeframe is bad for students who have a job, one can measure the number of participants again after the seminar has been rescheduled. In doing so, one can test, if the measure applied (the rescheduling of the course) was successful (=more participants in the course). The criterion (one criterion amongst others) for good quality in this case is the number of course participants.

Evaluations at higher education institutions usually shall be effective and support change processes. They should guarantee that management decisions are made on the basis of sensible information. Data and information that are gathered in the bounds of an evaluation shall be valid and reliable (see Chapter 5.4), if they shall be used as a control instrument. The validity and reliability of data can be increased by following the rules of empirical social research.

2.2   Internal and External Evaluation

Evaluations can be subdivided into internal and external evaluations. Internal evaluations are assessment processes that are conducted by actors, who are part of the organisation to be evaluated. If the evaluators simultaneously are actors in the field to be evaluated, e.g. members of the faculty who assess the processes and products of their own faculty, one talks about internal self-evaluations. If the assessment is done by, for example, a QA unit integral to the higher education institution, the evaluation is an internal one, but no self-evaluation.

On the other hand, external evaluations are those ones that are conducted by experts from outside the organisation.

While internal evaluations have the advantage that they take less effort and provide more comprehensive internal knowledge to the evaluators about the subject to be assessed, they simultaneously bear the risk of a close connection between evaluator and evaluand (subject to be evaluated). This can mean that under certain circumstances staff members evaluate the achievements of their colleagues, which is a sensitive process, because not everyone likes to be judged by their colleagues and a tendency euphemise certain aspects might evolve. Moreover, a personal involvement in processes that are the object of an evaluation can hinder the process of free brainstorming about potential measures of improvement.

If the evaluation is conducted by an institute within the organisation, e.g. a QA office, one should also bear in mind that this can lead to conflicts within the organisation that could be quite sustainable. With external evaluations this potential conflict between evaluators and persons connected to the item or programme evaluated is externalised and is not sustained within the organisational framework. External evaluations have the advantage that usually the evaluators are independent and thus can bring in new perspectives for the evaluation unhindered. On the other hand, external evaluators will need far more time to come to a comprehensive final assessment about the evaluand, because they first will have to become familiar with the backgrounds and frameworks of the evaluand (Stockmann 2006, 263). External actors will never have the amount of insight a person has, who is working in an organisation every day. This is why it is best to always combine the external perspective with information from internal actors of an organisation (Stockmann/Meyer 2010, 81). The following matrix summarises assets and drawbacks of internal and external evaluation:

	Internal Self-Evaluation	Internal Evaluation by QA Office or Similar	External Evaluation by Peers
Assets	■ Fast, less effort	■ Some effort, some inde-pendency	■ Independency
	■ Insight into details	■ High method authority, some insight into details	■ High method authority
	■ Cost-efficient	■ Guided follow-up	■ High credibility
			■ Resistance possible
Drawbacks	■ Low method-authority	■ Internal conflict potential	■ Resistance possible
	■ Lack of neutrality and distance to evaluand		■ More costly
	■ Organisational blindness		■ No guided follow-up
	■ Low reform potential		

Table 7 *Potential assets and drawbacks of internal and external evaluation*

Table 7 gives an overview over potential assets and drawbacks. Assets and drawbacks of one evaluation format can be valid for the other as well and vice versa, depending on the design of the evaluation and the organisational structure, within which an evaluation takes place.

In the context of higher education institutions on all levels and in all fields – in teaching and learning, research, departments or universities as a whole – when conducting evaluations, a combination of internal self-assessment and external assessment has developed. Particularly with institutions of higher education that are characterised by subject relation and often referred to as "expert organisations" (see above), the approach of so-called informed peer-review has proven to be adequate. In an informed peer-review a comprehensive self-evaluation process takes place before the peers are invited to assess an evaluand. The results of the self-assessment are compiled in a self-assessment report (SAR) which serves as an information basis for the peer reviewers. The peers thus base their assessment on information from the report and also from discussions with relevant stakeholders during a site-visit.

Not all participants of an evaluation are familiar with underlying concepts and practical processes of evaluations. Sometimes they need training in order to fulfil their role as interviewee or as peer-reviewer. For most evaluations, peer reviewers are trained how to conduct interviews, how to write a report etc.

2.3   Participatory Evaluation

Applied social research evaluation, in contrast to research, aims at providing useful information for the purpose of development. This is particularly true for evaluations following the development paradigm (Stockmann 2006, 262).

The deduced improvement measures of an evaluation sometimes affect different actors of an organisation, which makes it even more important that these measures make sense to them. The relevance of evaluations for the actors involved is an important factor for the success of the process, as is also formulated in the utility standards in Chapter 1.3.

Relevance for actors can be achieved by including the actors' questions and advice in the evaluation process and let them contribute actively to the planning process (see U2 Attention to Stakeholders). The contribution of affected actors additionally has a reflex on the quality of the process itself, because the internal actors have more knowledge about details and backgrounds, which can help to gain an even deeper insight into the evaluand.

Last but not least, the participation in the planning process of an evaluation provides a lot of insight into processes of an organisation and knowledge acquisition about evaluation theory and thus can also be seen as an additional qualification of the persons involved.

While it is good advice to actively involve the affected staff members of an organisation or organisational unit to be evaluated in the planning phase of an evaluation and in the reflexion of results and planning of improvement measures, the data collection should in any case lie in the hands of a qualified evaluator with a good qualification in empirical social research, in order to ensure the scientific quality of data collection (see U1 Evaluator Credibility). In the phase of data collection, persons affected by the evaluation should only be involved as informants.

When planning an evaluation that is supported by evaluation experts, it is a good idea to constitute a working group that should ideally consist of members of all affected stakeholder groups, so that evaluation objectives and questions can be discussed and ideally agreed on (U3 Negotiated Purposes).

We understand transparency throughout the whole process as an element of a participatory approach to evaluation, that is: In our opinion a timely and comprehensive communication of the process development and its results among the stakeholders who are not part of the aforementioned working group is essential (see U7 Timely and Appropriate Communicating and Reporting).

Even if in our opinion a participatory approach to evaluation is advisable, a thorough consideration should be given to the decision about whether and to what extent which actors should be involved in the evaluation. The following aspects should be considered when selecting stakeholders to be involved:

- Participation does not exclude clear-cut responsibilities – Quality management and also evaluation are always also a "matter for the boss". Within the framework of a working group, finding a consensus is always

worth the effort of seeking it. Sometimes it is even a good idea to involve those stakeholders who are notoriously critical of an evaluation. But if you cannot reach consensus within an evaluation working group, the decision of the initiator or sponsor of the evaluation, which is usually someone responsible (deputy vice-chancellor for academic affairs etc.), might be needed.

- "Too many cooks spoil the broth." – When conducting a participatory evaluation, you do not really need to involve ALL potential stakeholders in the same way. For example, when planning an evaluation with a working group it is important that this group is able to work, that is: according to our experience, it is not comprised of more than six members. When selecting members of the working group, one should also bear in mind that persons are chosen, who can represent their unit, group etc., that is, who are accepted by their fellow group members and have a comprehensive knowledge about the evaluand.

- Hierarchies and roles – For the constitution and the actual tasks of a working group, you should keep in mind that mechanisms of the normal job reality are still valid, when you put together staff members in a team working for a different purpose. Although in an evaluation working group, you might want all members of the group to be on the same level, staff members of a higher hierarchy level usually stay "the boss" in the evaluation working group. This might result in other members of the team not conveying their true opinion about things, because they might not want to criticise the work of superior staff etc. Here, it is crucial to address evolving conflicts within the working group immediately and openly in order to foster a cooperation based on trust. This aspect is of particular relevance if student members are involved.

- Participation of students and alumni – The involvement of these actor groups is, in our experience, often a particular challenge, because students and graduates frequently cannot spend much time at their (former) higher education institution due to study or job commitments. Sometimes they simply are not interested in an evaluation process. With graduates another problem is that they might me harder to reach, since they are no longer present at the institution. However, reacting to these challenges, it has proved to be a good compromise to keep low the amount of attendance at meetings for students or alumni.

- Stakeholders are not automatically experts – As formulated above, it is not a good idea to assign the task of developing instruments for data collection and of planning the whole evaluation to any member of the working group, since a high amount of expertise in the field of evaluation theory and empirical social research is necessary to accomplish this task in a manner that guarantees useful results.

- Proper participation or no participation – It is absolutely legitimate to conduct evaluation without participation of the affected actors, as long as they are informed about all relevant aspects of the evaluation. Sometimes a fake participation of the stakeholders of an evaluation is realised; that is: They are invited to participate in an evaluation, but in reality crucial decisions on key aspects or questions of the evaluation are made above their heads (see "hidden agendas" in Chapter 1.3). This practice can weaken trust in superiors and the structures of an organisation in the long term, which is why we strongly recommend to only involve actors, if their contributions are welcome.

As a quality manager it is your task to make recommendations on the parties and stakeholders that should be involved in an evaluation and to make sure that they are able to contribute fruitfully to the discussion in a working group and to the evaluation process as a whole.

2.4   Phases of an Evaluation

According to our understanding, any evaluation can be subdivided into four phases, that are similar to the ideal-typical phases of a research process as described in Chapter 4.1: 1. Planning phase, 2. Data collection phase, 3. Reflexion phase, 4. Follow-up phase (for other subdivisions of evaluation phases refer to Stockmann 2006, 290, or Röbken 2012, 73). It is generally a rule that single phases of the process overlap or are exercised simultaneously. For example, during the data collection phase, there sometimes will be the need to interpret already existing data and to restructure or adapt the other phases. In the following, the particular evaluation phases will be explained taking into account the special circumstances at a higher education institution.

I. Planning Phase: In this initial phase of an evaluation, the task of a quality manager and an evaluation working group is to plan the process of the evaluation taking into account the special circumstances of a higher education institution and in a way that it adheres to the feasibility standards explained in Chapter 1.3.

This means that all work packages and process steps should be planned systematically and realistically in terms of time by a coordinator (the quality manager), who is responsible for the progress of the process (see Table 3, F1 Project management). The evaluation should be planned in a sensible way, that is: It should be tailored to the objectives of the evaluation and the evaluand (see Table 3, F2 Practical procedures). Moreover, the planning should be done in a way that supports acceptance of all stakeholders by respecting their respective interests and perspectives for the planning process as much as possible (see Table 3, F3 Contextual viability).

Evaluations do not usually belong to the everyday work of scientists, teachers, administrative personnel, external reviewers and students, and thus represent an extra workload for these actor groups. This is why it is of utmost importance to facilitate the evaluation by as much monetary and personnel resources as is necessary for the objectives of the evaluation (see Table 3, F4 Resource use), but with as little as possible, in order to guarantee an effective and efficient process.

During the planning phase, the following interconnected aspects should be discussed and decisions must be made in order to ensure a smooth evaluation process:

- The evaluand: When deciding about the evaluand, the question "What is to be evaluated?" should be asked in a critical and thorough manner. One might be tempted to pack a lot of different questions on different levels into one evaluation, because it is planned anyway. To operate on the premises of data economy and the reduction of additional workload, it is strongly recommended to limit the evaluation to a clear-cut evaluand. This is particularly true, because a clear-cut evaluand objective of the evaluation make a deeper level of investigation possible.
- Objective of the evaluation: In order to be sure about the real objective of an evaluation and thus to be able to make it known to all stakeholders, quality managers should ensure that the following questions are answered before the beginning of the evaluation: What is the aim of the evaluation? Should it be conducted as a target-performance comparison using a summative approach? Should information be collected that in the first place should be used for informing the public about ongoing processes/projects? Or should the evaluation generate knowledge that allows the deduction of improvement measures? As you can see, it all

comes down to answering the question about the evaluation paradigm (see Chapter 1.1).

- Instruments to be applied: For a sensible and valid assessment of an evaluand, it is best to collect as comprehensive information as is possible, in order to be able to base judgements on it. As a general rule it can be recommended to get a detailed insight into the given evaluand by respecting different perspectives (triangulation). Here, you as quality manager should first collect information about which data and information already exist at your higher education institution to avoid unnecessary work. To choose appropriate additional instruments you should ask the following questions: Which (additional) data and information are needed? Who or what can provide these data and information? How should the data and information be acquired? Should a qualitative or quantitative research strategy come into practice (see Chapter 3.3 for assets and drawbacks of each research strategy)? In addition, the selection of the instruments to be used is also dependent on the framework conditions of the evaluation. For example, it can be sensible to ask, whether the technical prerequisites are given for an online-survey or – more generally – whether the timely and monetary resources are sufficient to carry out an evaluation and whether the staff in charge of the evaluation is appropriately qualified. When selecting instruments to be applied in an evaluation – as is the case when planning an evaluation in general – you can best adhere to the rule: As much as necessary, as little as possible!

- Evaluators and other stakeholders: Dependent on the evaluand, the structural framework conditions of a higher education institution (Does the higher education institution employ quality managers? Does a quality assurance unit exist?), and the instruments to be applied etc., a decision has to be made regarding who will, in which manner, contribute to the evaluation: Shall the evaluation be conducted internally, or shall external experts be invited (see Chapter 2.2)? How participatory shall the evaluation be designed (see Chapter 2.3)?

- Scheduling: At the end of a planning phase for an evaluation a so called Project Action Plan (PAP) should be in place. The PAP should contain the phases of an evaluation as well as the timeframe for data collection and responsible actors for the scheduled activities and other stakeholders to be involved.

> *"Data triangulation, also referred to as data sources triangulation, depicts the use of multiple data sources in the same study for validation purposes. According to Denzin (1978), there are three types of data triangulation; namely, time, space and person. These types of data triangulation come as the result of the idea that the robustness of data can vary based on the time data were collected, people involved in the data collection process and the setting from which the data were collected (Begley 1996)"* (cited from Hussein 2009, 3). Triangulation can help to increase data or information validity, because the weaknesses (blind spots) of one instrument applied can be eliminated by the strengths of another instrument.

II. Data Collection Phase: The data collection phase forms the core of an evaluation. During this phase, the data and information necessary for answering the questions of the evaluation are collected and processed in a way that makes them understandable for stakeholders who are not experts in empirical social research.

The data collection phase can be subdivided into the following four steps:

a) operationalisation and development of instruments,
b) data collection,
c) data processing,
d) reporting

(see Chapter 4.1 for more detailed information about the phases).

This phase, thus, is the process step of an evaluation, where methods of empirical social research come into place. The design and the timeframe of these aforementioned steps are highly dependent on the design of the evaluation in general.

III. Reflexion Phase: Are the collected data processed in a way that allows also data-laymen to understand them? A systematic reflexion of the evaluand on the basis of these data should be started (see also reporting and dissemination phase, Chapter 4.1). At this point, it is, in our opinion, not sufficient to present the processed data and results to those responsible and the decision makers, but a reflexion process is recommended that is moderated by you as quality manager. The reflexion of data in the form of a guided discussion contributes a lot to the quality of data-based assessments. Generally, it is a good choice to conduct the data reflexion in the bounds of one or more workshops. In this workshop or these workshops first all information on hand should be presented by the quality manager in a comprehensible form to relevant stakeholders, questions should be answered and the data should be reflected and discussed. Afterwards final statements and – this is particularly true for evaluations according to the development paradigm – improvement measures should be formulated that should be included in an evaluation report. For the reflection of the data and information, the relevant stakeholders should receive also the raw data well before the reflection workshops begin.

IV. Follow-up Phase: Evaluations should produce effects, that is: Evaluations should produce results which are directly connected to the objective of the process. With summative evaluations that are often used for legitimation for particular interest groups, the evaluation objective can be reached by publishing an evaluation report. Evaluations according to the development paradigm, which are most common in the context of higher education institutions, should foster changes and be effective within the organisational structures. Here, it is usually not sufficient to comment on the results of the evaluation and to present the deduced improvement measures. According to us, it is important to give the improvement measures a binding character, which can for example be achieved via target agreements between faculty (or another organisational sub-unit) and the vice-chancellery, or – in the case of course evaluation – via a talk between the lecturer and the dean. A binding character of the follow-up process increases the effectiveness of an evaluation and makes the progress of development measureable. More detailed information about how to design the follow-up phase of an evaluation and how to close the quality loop in general can be found in Module 5.

2.5 Institutional Evaluation at the University of Duisburg-Essen

At the University of Duisburg-Essen (UDE) in Germany, a procedure of evaluating all organisational units has been in place since 2006. This evaluation procedure takes into account the processes and achievements of all sub-units in a cycle of six years. Hereby, the evaluation is directly connected to the university-internal target agreements between the sub-units and the university's leadership. The following units undergo the evaluation:

- faculties,
- research units,
- service units,
- central administration.

Note: Because of the fact that the whole portfolio and the organisational structure of the particular units, that is: the whole institution, is considered for an evaluation, the term "institutional evaluation" has been established at UDE and other German institutions that have similar procedures in place. Internationally, "institutional evaluation" usually refers to the evaluation of higher education institutions as a whole. By "institutional evaluation" in this course book, however, we refer to the evaluation of organisational sub units.

In the framework of institutional evaluations at the University of Duisburg-Essen, all areas of achievement, (teaching and learning, research, service and management) of the respective organisational unit to be evaluated are taken into account for the evaluation, if applicable. These areas are assessed by stakeholders of the organisational unit itself and the results of the self-assessment are put into a self-assessment report that is given to external experts, who resume their assessment in an external report in turn. Although institutional evaluations at the UDE are conducted on the level of organisational sub-units, there is also space to address particular questions on the level of study programmes within the evaluation.

The whole procedure of institutional evaluation at UDE, which takes place every six years, takes 9 months to be carried out and prepares every second cycle of target agreements which take place every three years. The procedure is fixed in the evaluation regulations and thus is obligatory for all organisational units. Whereas the evaluation procedure of faculties is more standardised, that one for other organisational units is more focussed on special questions and more flexible.

The process is oriented to international standards for evaluation and quality management, as for example the European Standards and Guidelines for Quality Assurance in the European Higher Education Area (ESG) (ENQA 2009) or the standards of the German Association for Evaluation (DeGEval 2001).

UDE's institutional evaluations are coordinated by its quality assurance unit CHEDQE (Centre for Higher Education Development and Quality Enhancement).
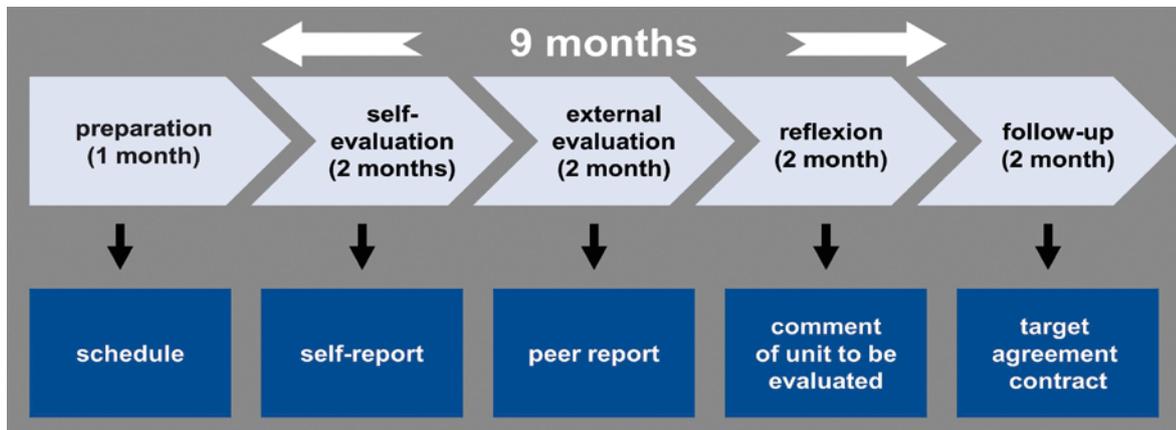
Figure 1 *Institutional Evaluation at University of Duisburg-Essen*

The standardised procedure of faculty evaluation is subdivided into five phases and will be described in more detail in the following:

I. Preparation Phase (see Planning Phase): As explained in Chapter 2.4, in the planning phase the following aspects must be clarified: evaluand, objectives, instruments, stakeholders, schedule. Evaluations at UDE are committed to the development paradigm and thus it is important to produce consensus and transparency with all relevant stakeholders in the planning phase. This happens in three steps:

1. Preparatory Meeting	
Participants of the meeting:	■ Pro-vice-chancellor for resource planning and development ■ Quality manager from CHEDQE ■ Representatives in charge of department for resource planning and development ■ Representatives in charge of Science Support Center (SSC)
Content/objectives:	■ Reflection of status quo of the faculty to be evaluated on the basis of key performance indicators provided by the UDEs controlling unit ■ Collection of specific questions for the evaluation, especially for the external reviewers

2. Kick-off Meeting	
Participants of the meeting:	■ Same as in preparatory meeting ■ Representatives of the deanery
Content/objectives:	■ Presentation of the process by CHEDQE ■ Discussion of the special questions of the evaluation ■ Detailed planning of process and schedule ■ Establishment of a consensus of all stakeholders
3. Informing Staff Members	
Participants of the meeting:	■ All relevant stakeholders of unit (scientific and non-scientific staff, students)
Content/objectives:	■ Make the process transparent, minimise anxieties and reservations

Table 8   *Three steps to produce consensus and transparency in UDE evaluations*

II. Self-Evaluation (see Implementation Phase): The faculty to be evaluated receives guidelines for the self-assessment report and is asked to describe and assess the following aspects:

■ organisational structure,
■ profile in study and teaching,
■ research profile,
■ quality management,
■ diversity management.

For all of these aspects, a description of the status quo, a long-term strategic planning (next six years), a short-term operative planning (tasks for the next three years) should be written down. The deanery is responsible for writing the report.

Usually, before writing the report, the faculty to be evaluated conducts a self-assessment exercise. No prescriptions in the form of checklists etc. are made on how to design this process, but the quality managers of CHEDQE offer support for planning and moderating the exercise, e.g. moderating a SWOT analysis.

Institutional evaluations at the UDE are embedded in a holistic quality management system (QMS). That means that results of course evaluations or other surveys among students or staff as well as data from a data warehouse containing common key performance indicators, should be used as a basis for reflection in the self-assessment exercise.

III. External Evaluation (see Implementation Phase): For the external assessment at UDE up to six experts are invited. For the selection of these, the vice-chancellor decides, whom to invite on the basis of a list provided by the faculty itself. The UDE has guidelines and criteria in place for the selection of the external experts (e.g. external experts must not be in close research cooperation with the faculty to be evaluated etc.) in order to ensure an unbiased and neutral perspective of the experts.

The external experts receive the self-assessment report and specific questions by the vice-chancellor and faculty and are asked to write a short comment before the on-site visit on the basis of these documents. The comment shall include main strengths and weaknesses the experts identify from the documents provided, but also open questions and the demand for additional documents. This makes the preparation of the on-site visit easier and more fruitful.

During the two-day on-site visit, the external experts will talk to representatives of the vice-chancellery, the deanery and different stakeholders from the faculty (scientific and non-scientific staff and students). All discussions are guided and moderated by staff of CHEDQE. The on-site visit ends with a final workshop of the external experts, during which they prepare their pre-final assessments and recommendations, that will be put into the external report, and a presentation in front of members of the vice-chancellery and the faculty and a discussion with these stakeholders to avoid misunderstandings etc.

IV. External Evaluation (see Implementation Phase): As stated above, according to us, an evaluation does not end with the collection of data and information (via surveys, self-evaluation or external evaluation), but serves as a basis for the deduction of improvement measures in the first place. Before measures can be deduced, however, a reflexion of the gathered information must take place.

The design of the reflexion phase lies in the hand of the faculty itself. The faculty, however, is obliged to summarise the reflexion process and – more importantly – the deduced measures in the framework of a commentary to the external report.

V. Follow-up Phase: At the University of Duisburg-Essen, the follow-up phase marks the transition from evaluation to university-internal target agreements and the conduction of respective measures agreed upon.

Within the bounds of a final discussion between the pro-vice-chancellor for resource planning and strategic development, representatives of the evaluated faculty's deanery, the Science Support Center and CHEDQE the evaluation process is officially finished first. The discussion in the first place serves the purpose of jointly reflecting the results of the self-evaluation, the external report and the faculty's commentary to the external report.

Finally, in the framework of this talk, measures that would be recommended by the vice-chancellor to discuss again officially in the frame of the target agreement process are agreed upon.

You will get to know details about the process of target agreements at UDE in Module 5.

2.6 Evaluation of Teaching and Learning Focussed

Evaluations at higher education institutions take place in a variety of areas (research, teaching and learning, services, management etc.) and many organisational levels (faculties, institutes etc.). The number of potential evaluands at higher education institutions is practically unlimited.

Teaching is, besides research, one of the core processes at higher education institutions, and in Europe and the USA evaluations of this core process have been established since the 1970s. Today, the regular evaluation of study and teaching is reflected in the higher education laws of some countries or bodies, brought about to ensure the quality of higher education in general, which of course includes the area of study and teaching.

In fact, in the area of study and teaching, some of the most advanced and best tested methods and instruments for quality assurance are in place with a variety of foci and levels of analysis. Some of them will be summarised in the following:

a) Course Evaluation: The evaluation of courses takes place at practically all higher education institutions worldwide. Course evaluation can be exercised using a variety of methods – from job-shadowing by colleagues to group discussions with students. The by far most common method for evaluating courses, however, is the evaluation of courses by standardised questionnaires that can be paper-based or online. With this method, students are asked to answer questions about their overall satisfaction with a particular course, didactical design of the course and the performance of the teacher.

b) Graduate Tracer Studies: Surveys of graduates have become increasingly important in recent years for the evaluation of study and teaching at higher education institutions. Graduates can give a retrospective assessment of how far the contents and didactical design of their studies have fostered their knowledge and competences, particularly those aspects that qualified them for the job market.

When surveying graduates, special challenges have to be dealt with, because usually, the target group is not as easily reachable as the students, who are still physically present at the higher education institution. In most cases, graduates are surveyed using standardised online surveys. In countries where there is no well-established legal address management that allows research for addresses after graduates have moved etc., other methods of surveys like telephone interviews are used for graduate tracer studies. The same applies for countries where the technical infra-structure for online-surveys is weak.

c) Study Programme Evaluation and Accreditation: On the level of study programmes, evaluations analyse for example the structure or the orientation of the programme to the job market. Besides the internal and also external evaluation of study programmes, in most countries the accreditation of study programmes is legally binding. Accreditation ensures that a study programme adheres to legal frameworks and higher education policies. In the framework of an accreditation process a study programme is checked against the legal criteria and – if it complies with the prescriptions – is awarded an accreditation certificate by a legal body that accredits study programmes. In Chapter 4 and 5 of Module 3 you will learn more about internal and external study programme evaluation, accreditation and the linkages thereof.

d) Evaluation of Framework Conditions: Quality of study and teaching is not only about the content or structural design of courses or study programmes, but is massively determined by a variety of framework conditions that affect teaching and learning processes. For example, room conditions, technical infrastructure, equipment of the library and internet have an influence on the quality of teaching and learning. For gathering information about these factors, a variety of methods and instruments can be chosen, like the survey of students and teachers etc.

e) Key Performance Indicators for Study and Teaching: Key Performance Indicators in the area of study and teaching are a good quantitative starting point for evaluations. For example, data on student-teacher-ratio or drop-out-ratio (number of students of one cohort, that did not finish their studies), as well as graduates' final grades, offer a first overview (for example comparison over the course of time or benchmarking with other subjects or institutions) of problems or improvement potentials. This can be an ideal starting point for an in-depth analysis of these problems etc. Chapter 4 of Module 4 will provide more detailed insight into the concept and usage of key performance indicators.

Of course, there are innumerable other methods and tools that come into practice for the evaluation of different aspects of study and teaching, like for example drop-out surveys, tracer studies, document analysis of study programme descriptions, employer satisfaction surveys, secondary analysis of rankings etc. In the limited bounds of the course book on hand, however, we cannot deal with all of them.

In fact, we want to concentrate on the description and critical discussion of course evaluation practice, because a) course evaluation is often the first instrument for quality assurance of teaching and learning that is implemented at higher education institutions and b) the critical discussion of the procedures below is easily transferrable to other and more complex processes of quality assurance.

2.7 Comparative Case Studies of Course Evaluation

For many years there has been a form of course evaluation by student surveys implemented at higher education institutions worldwide. The procedures in place differ in terms of implementation (e.g. the manner of distributing the survey (online or paper-based), the instruments (contents of the questionnaires) and the use of results. In the following, three case studies will be discussed along the lines of similarities and differences in these fields.

Course evaluation has been implemented for many years at both the Karlsruher Institut für Technologie (KIT), National University of Singapore (NUS) and University of Potsdam (UP), but different approaches and practices are in place:

Course Evaluation at Karlsruher Institut für Technologie (KIT)

1. Background Information

The KIT was founded via a merger of the Research Center Karlsruhe and the University of Karlsruhe in 2009. In winter semester 2013/2014, 24,530 students were enrolled at the KIT.

Quality enhancement activities at KIT lie in the responsibility of a staff position directly associated with the institute's leadership. All processes of quality assurance and enhancement are laid down in a quality handbook and a university quality enhancement act. In May 2014, the QA-system of the KIT was accredited by the German Accreditation Council. Retrieved on September 1, 2014, from www.kit.edu/index.php

2. Evaluation Purpose

Course evaluations at KIT not only pursue the objective to provide teachers with feedback from their students, but also should serve for accountability towards the state and public. Course evaluations should form the empirical basis for improvement of study programmes. Retrieved on September 1, 2014, from http://www.pst.kit.edu/lv-eva.php

3. Process of the Evaluation

Each semester, 35-50% of the courses at KIT are evaluated by students. The courses to be evaluated are selected by the dean of studies of each faculty. In the second half of the semester, teachers of selected courses receive paper-based questionnaires from the central person in charge (see above) to be distributed among their students. The questionnaires are filled in by the students during one session of the course, collected by one of the students and sent to the central staff in charge for the purpose of data processing. Retrieved on September 1, 2014, from http://www.pst.kit.edu/211.php and http://www.pst.kit.edu/downloads/6_Evaluationsordnung_2013-05-31.pdf

4. Content of the Questionnaires

All questionnaires of the KIT contain a set of obligatory questions concerning the quality of course content and infrastructure, teaching performance, student counselling, organisation of the course and engagement of course participants as well a question about the overall students' satisfaction. All faculties can add faculty specific questions to this obligatory questionnaire part. Retrieved on September 1, 2014, from http://www.pst.kit.edu/downloads/6_Evaluationsordnung_2013-05-31.pdf

5. Data Processing and Use of Data

The arithmetic means of answers to all questions in the questionnaire are combined in a report. For five core questions, a traffic light logic has additionally been established that marks all arithmetic means on the basis of 5-point scales scoring between 1 and 2,5 as good (green), results scoring 2,5 to 3,5 as average (yellow) and those scoring above 3,5 as poor (red). In addition to that, an index is calculated on the basis of the core questions, weighting the questions differently. The teaching quality index is used to describe the overall quality of teaching within a particular course. Thus, the courses can also be labelled as good, average and poor.

Core Question	Weighting in the Teaching Quality Index
Please state your overall satisfaction with the course (scale from totally satisfied to not satisfied at all)	50%
The workload needed for this course is… (scale from adequate to not adequate)	12.5%
How do you judge the structure of the course? (scale from very good to very poor)	12.5%
In how far is the teacher motivating in this course? (scale from very much to not at all)	12.5%
In how far does the teacher respond to questions and inputs of students? (scale from very much to very little)	12.5%

Table 9   *Teaching Quality Index of KIT (http://www.pst.kit.edu/230.php)*

The results of course evaluations are distributed as follows:

a. Teacher – receives results of all evaluated courses and is obliged by KIT's evaluation regulations to discuss them with the students of the particular course.
b. Faculty board, Commission of Studies and Dean of Studies – receive results anonymised by aggregation with hints to good, average and poor courses and are obliged to deduce measures for improvement (e.g. talk to teacher, recommendation for teacher to participate in higher education didactics courses etc.).
c. Deputy-Vice-Chancellors for Studies and the Commission for Studies and Teaching of the senat – receive the aggregated results of all faculties with the hint to good, average and poor courses and discusses the deduced measures for improvement (see above) with the Deans of Studies.

Course Evaluation at the National University of Singapore (NUS)

1. Background Information

The National University of Singapore (NUS) is a medium-size full university of approximately 40,000 students and 2,500 teaching staff. At every semester, all courses undergo teacher, pedagogy, and content evaluation. An increased emphasis on teaching quality has led to a number of strategies for its evaluation and promotion, including student feedback, peer review reports, teacher appraisal, teaching excellence awards, etc.

The state of quality teaching at NUS is captured in a number of publications by the Centre for Development of Teaching and Learning (CDTL, www.cdtl.nus.edu.sg) including a handbook (Pan, 2008), regular newsletters, an annual conference, case studies, and other teacher training materials. The teacher at NUS sees clearly that teaching quality has a direct effect at performance appraisal.

2. Evaluation Purpose

Traditional teaching evaluation focuses on: i) the quality of the knowledge that teachers transmit to students, and ii) the transmission quality of this knowledge. This view is teacher-centered. A student-centered view asks "What is the quality of learning that results from the teaching?"

Teaching evaluation at NUS takes place at two levels. The first is the level of surface teaching (e.g. depth and breadth of topics coverage). End-of-semester course evaluation is for assessing surface teaching.

The second is the level of deep teaching where the results may not be immediately observable and would be realized only a few years after graduation. The outcome-based evaluation of a whole program of study may be said to assess deep teaching. Outcome-based assessment such as of the Washington Accord is a good example of the assessment of deep teaching. Other educational program assessment models are moving also in the direction of being outcome-based and assessing deep learning.

3. Process of the evaluation

Course evaluation begins at the 10th week of classes where students are required to complete an online feedback questionnaire. End-of-semester examination registration is contingent upon completion of the feedback.

The feedback consists of three parts. Part I is a general feedback about the course. Part II asks about various aspects about the course. Part III is about feedback of the teachers (positive and negative aspects of the teacher in open-ended questions).

The feedback has to satisfy two criteria. The first is a minimum number of lectures for evaluation.

Teachers who cover less than five sessions (lectures) in a course are not assessed. This is because it is unlikely that the feedback can be meaningful given the low contact hrs. The second criterion is that the class has to have at least 10 students. I.e., classes with less 10 students are not assessed. Due to resource issues, NUS discourages such small classes anyway, and no class can have less than eight registered students.

At the end of the semester, only after all examination grading have be completed and submitted to the University, are the results of the feedback exercise are communicated to the teachers. This is to not allow the teachers grading to be influenced by the students' assessment.

4. Content of the questionnaires

Annex 1 is a sample of a complete end-of-course evaluation including the questions asked, ratings made on the teacher, summary ratings from respondents, etc (from www.cdtl.nus.edu.sg). Each faculty member is rated on six questions, including how the teacher has enhanced the students' thinking ability, how the teacher has increased their interested in the subject, provide useful and timely feedback, etc.

5. Date processing and use of data

The six questions used a Likert scale of 0.0 – 5.0. Average and standard deviation scores for the faculty member, the department, and the faculty as a whole are provided. In this regard faculty members are ranked against each other. It can be said that the ranking is harsh in that comparison is made to three decimal places.

Students' positive and negative open-ended comments on the teacher are shown also. There is also a nomination for the best teacher award. Good teaching is rewarded this way. Ineffective teachers are counselled at the end-of-year performance review.

The information on all courses taught in the past years are kept by the University and may be viewed online. This information is used for promotion and tenure purposes. The University emphasises continually the need to achieve teaching scores in the evaluation.

Tan Kay Chuan

Course Evaluation at the University of Potsdam (UP)

1. Background Information

The University of Potsdam was founded in 1991 and had around 20.000 students in the winter semester 2013/2014.

Besides other tasks, as for example the promotion of e-learning activities, staff career development, higher education didactics, the Centre for Quality Development in Teaching and Learning is in charge of the quality enhancement activities at the UP on institutional level.

All processes of quality assurance and enhancement are laid down in a quality handbook and a university quality enhancement act.

In the year 2013, the QA-system of the UP was accredited by the German Accreditation Council.

2. Evaluation Purpose

In the first place, the course evaluation serves the purpose of fostering dialogue between students and teachers and to establish a culture of vital feedback among students and their teachers in the long run. On the basis of course evaluation results, both groups shall develop measures for improvement discursively. However, anonymised and aggregated course evaluation results are sent annually to the deanery and by request, it is possible for deans to receive the results of individual teachers in order to implement measures if necessary.

3. Process of the Evaluation

At UP, course evaluations are mainly conducted with questionnaires via paper and pencil or through the own online evaluation platform. The platform gives teachers access to a list containing all of their courses. From this list, they chose the courses to be evaluated (each faculty has its own regulations for criteria for choosing the courses that have to undergo an evaluation, if any) and are provided with paper questionnaires or access data for an online questionnaire. Retrieved on January 5, 2015, from https://pep.uni-potsdam.de/articles/kurse-bewerten/Kurse-bewerten.html

4. Content of the Questionnaires

Teachers have the possibility to choose between either a) their respective faculty specific questionnaires or b) a new learning outcome based individual survey. Every faculty has its own questionnaires, a generic one and partially also for specific course formats (e.g. for practical/laboratory courses). The questionnaires are developed in close cooperation between the Centre for Quality Development and the faculties, resulting in tailored instruments for different faculties. However, all questionnaires con-

tain questions about teaching performance, students' engagement, infrastructure and framework conditions as well as overall students' satisfaction.

The new learning outcome based survey instrument has been introduced in 2014 and allows each lecturer to select the learning outcomes of their course to be evaluated from a predefined list and let them add own individual questions. This will generate an individual survey that addresses the learning outcomes of the course. With it, the idea is to have more valuable individualised feedback for the courses. Still, global questions such as about the students' satisfaction, or motivation are being asked to allow some form of comparison between all courses.

5. Data Processing and Use of Data

No quality indices are calculated for courses at Potsdam University. Arithmetic means of the answers of each question are put together in a report. The results of course evaluations at UP are processed and communicated as follows:

a. Teacher – receives results of all his evaluated courses and is obliged by UP's evaluation regulations to discuss them with the students of the particular course.
b. Dean/Dean of Studies – receives results anonymised by aggregation in order to deduce general measures for improvement for all courses. Deans can also ask the Centre for Quality Development for results of individual teachers. These teachers are informed if a request is made.
c. University's Leadership – does not receive any results of course evaluation.

? Questions & Assignments

1. Which major differences do you recognise between the two procedures for course evaluation of KIT and University of Potsdam? What do you think, on which evaluation paradigm could the procedures above be based?
2. What are the advantages and disadvantages for each of the procedures described above?

2.8 Course Evaluation – What to Ask and How to Ask it

As discussed in Chapter 2.1, the question of what quality is, is more difficult for educational institutions than it is for the economy and industry. This is, because processes and products of higher education institutions are vastly determined by expectations and behaviour of different actors. This in turn means that these actors should be respected when quality is to be assessed.

Before we can assess processes of teaching and learning, we must agree on what quality should be in this context. In other words, from what can we see that the quality of our processes and products is good?

This question cannot be answered easily. If we asked ten professors, what constitutes high quality of teaching and learning for them, we would most likely get ten different answers. The notion of quality can differ a lot from person to person. Even if we only looked at a tiny detail of the huge area of study and teaching, as for example a particular course, it would be very hard to find a common understanding of what a high quality course is, because here too, different actors (namely students and teachers) with different expectations are involved and the course takes place within particular framework conditions.

Thus, the quality of a course is a combination of the quality of different aspects of a course. To ask questions about all of these aspects in a sufficiently in-depth way, is nearly impossible to do with one single questionnaire, as is usually used for course evaluation. For example, the time the course takes place, the rhetoric of the teacher, as well as the contributions of the students themselves all influence the quality of a course positively or negatively.

Empirical social research can help us to find a shared understanding of a concept (e.g. quality of a course) and thus to make it empirically approachable, that is: measureable.

The challenge now is to transfer the concept of course quality into a questionnaire. This process of translating is referred to as operationalisation in social empirical research and will be explained in more detail in Chapter 5.5 along the lines of the Munich Multifactoral Model for Course Quality of Heiner Rindermann (1998). By this model, Rindermann presents a concept of course quality that tries to systematise the aspects or dimensions that can have an influence on course quality: Rindermann differentiates the dimensions of teacher, students and framework conditions as parts of the concept course quality. To get to know more about these dimensions (e.g. the behaviour of the teacher or the previous knowledge of students), we need so called indicators that give us information about the behaviour of the teacher or about the previous knowledge of the students.
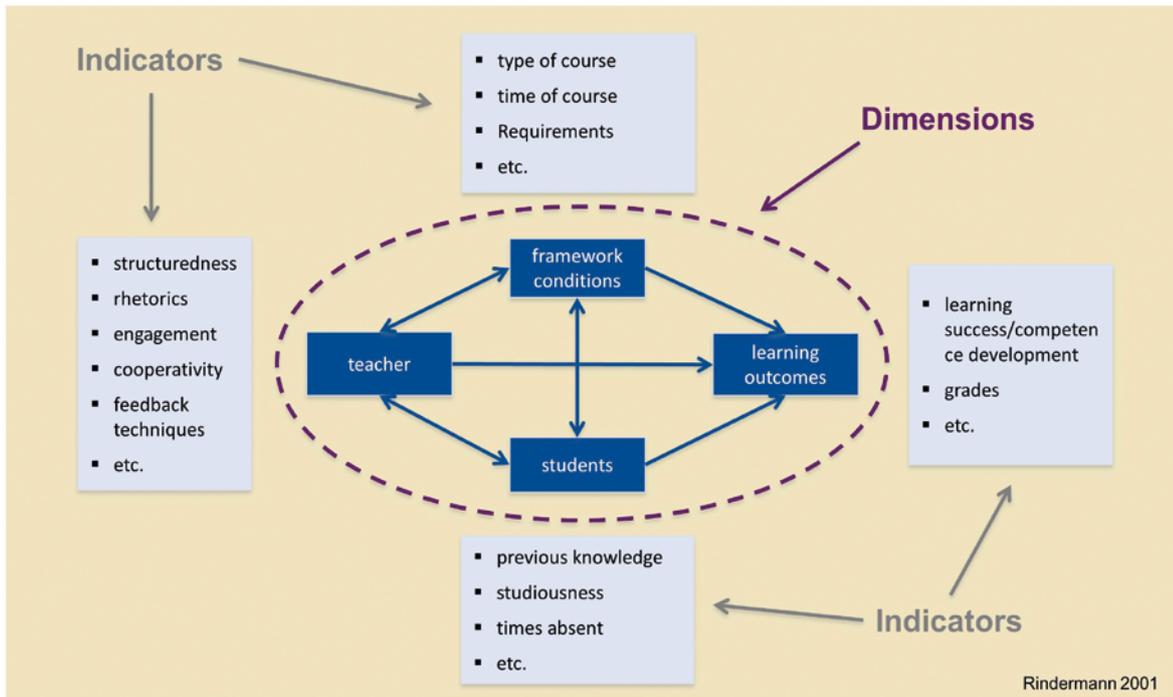
Figure 2 *Adapted Munich Multifactoral Model for Course Quality (adapted from Rindermann 2001)*

As we can see, the quality of a course is a very complex construction. Because of the fact that not every influential aspect can be considered in a questionnaire, results of a students' course evaluation survey do not represent the whole truth, but rather should be interpreted cautiously, also taking into account additional information from discussions with students etc.

But how can course evaluation by students' surveys support quality enhancement in teaching and learning? Which functions can it have and which expectations can't it meet? In the following, we will critically discuss some selected aspects of common course evaluations:

- Usage of results – Sometimes the results of course evaluation are linked to monetary or other gratifications or sanctions. This practice seems to be based on a notion of students as clients, who passively consume the product, namely the teaching performance. We cannot follow this definition of teaching as a good provided to passive clients. On the contrary, we think that conducting a course is an interaction to which both, teacher and students, contribute and that its success is also determined by the framework conditions in which a course takes place. Most of these framework conditions, again, can hardly be influenced by the teacher, e.g. room conditions etc. To use results of course evaluations as a means for controlling and rewarding or sanctioning teachers, who are only one of many factors that influence course quality, is accordingly not a good idea. Students' course evaluation is thus not appropriate for evaluations that follow the control paradigm. The results of course evaluations can rather be used to foster the dialogue between teachers and students about the course and to deduce appropriate measures of improvement. Students' course evaluation is therefore a good instrument for conducting evaluations following the development paradigm.
- Scope – At many higher education institutions, the results of course evaluation are aggregated and the summarised results are used to assess the quality of teaching of a study programme. This practice must be

regarded cautiously and critically, because the instruments used for the evaluation of a course are – naturally – tailored to course level. Measures of improvement on study programme level therefore cannot be sensibly deduced.

- Embedding the process – Good teaching becomes not only manifest in the form of didactically well-planned courses, but is also achievable by a well-structured curriculum and framework conditions that foster teaching and learning. That is why the assessment of particular courses by students can only be seen as one part of a more comprehensive view on the huge area of teaching and learning. This more comprehensive view is achievable by combining course evaluation with other instruments (study programme evaluation etc.), as is discussed in more detail in Chapter 5 of Module 3.

- Follow-up – The provision of results of course evaluation for the teacher and also the dialogue with students about the results will not result in a better teaching performance of the teacher, but capacity building in higher education didactics and counselling can support the teachers to change their teaching habits.

As a quality manager, it is your task to ensure good practice with the conduction of course evaluation and to include the aforestated recommendations in your everyday work.

2.9   Factors for Success of Evaluations at Higher Education Institutions

In the preceding chapters, we have discussed case studies of course evaluation and evaluation of faculties, but there are many more kinds of evaluations in higher education institutions. In the frame of this course book it is, however, impossible to discuss more than these three.

Nevertheless, in the following we will shortly discuss some aspects that you as quality manager should keep in mind when planning and conducting evaluations following the development paradigm, in order to increase the quality of the results and the success of the whole process. Please note, that here we can only give some general advice and inspiration. Some of the recommendations given below may not be applicable to all forms of evaluations.

- Relevance – Evaluations in higher education and in general should only be conducted if they are relevant for the stakeholders. That means, evaluations should produce results that help the involved stakeholders to improve their processes, structures or products. Otherwise the financial and personal expenditure, that is needed for planning and conducting an evaluation, is not justifiable. Relevance of an evaluation can be ensured by:
  - conducting the evaluation along the lines of concrete problems or actual questions within the institution; that is most of the time and orientation to strategic planning;
  - including expectations and perspectives of the different involved actor groups;
  - transferring the results into development measures following defined processes, that is: a connection to higher education management.
- Communication and participation – Evaluations always have an influence on different actor groups within a higher education institution and it is often observed that anxiety and resentment occur, because people may expect incorrect or bad assessments. Anxiety and resentment, in turn, can hinder a smooth and sen-

sible evaluation, because the involved actors may be very critical towards the evaluation in general, may be tempted to provide wrong information or may even refrain from participating in the evaluation. This is why, it is of importance to communicate aim, progress and results of the evaluation and involve relevant stakeholders in the process. This can be achieved by:

- establishing a working group for the evaluation, comprising representatives of all relevant actor groups that accompanies the process throughout the whole evaluation;
- continuous information of all stakeholders affected by the evaluation via sensitisation meetings, a project website, a newsletter or workshops that provide information about objectives, methods and progress as well as results of the evaluation.

- Process organisation – generally, evaluations should be accompanied by an expert in the field, even if they are internal self-evaluations. You, as the quality manager, are such an expert at your institution and it is your obligation to:
  - plan the evaluation, respecting common standards for evaluations and for social empirical research;
  - balance scientific standards and a practicable approach, that might be needed for the specific circumstances at a higher education institution;
  - use the techniques of project management for planning and conducting the evaluation;
  - plan the timeframe realistically including enough buffer for unexpected things;
  - ensure the reachability of a person responsible, who is coordinating the process;
  - conduct the process as intensely as necessary, but with as less effort as possible.

- Resource allocation – Evaluations mean extra work. A lot of evaluations remain below expectations, because monetary resources for coordinators, software etc. is lacking. As individual cost items for evaluations you should think of:
  - salary/honorarium for an evaluation expert, qualified in methods and standards of empirical social research;
  - workload/working hours of staff participating in an evaluation working group, if applicable;
  - honorariums, travel and accommodation costs for external experts;
  - costs for evaluation software (e.g. for the conduction of surveys);
  - costs for printing documents, such as own reports etc.;
  - costs for the organisation of workshops etc. (catering, moderation material etc.).

With regard to human resources, you should ensure that the staff involved are, at least partly, released from their every-day duties. Generally, the evaluation should only use as much (human) resources as are definitely needed for reaching the objectives of the evaluation. This is particularly true for human resources, because a workload that is too high can influence the whole evaluation negatively.

- Trustworthiness – Evaluations always have to do with assessment and thus – at least indirectly – with judging the work of staff within an organisation. Evaluations following the development paradigm should foster change in an organisation or of a process. Both change and assessment of one's own work are factors that can foster anxieties and resentment and thus influence an evaluation negatively. We believe it is of utmost importance to keep the following aspects in mind, when planning and conducting an evaluation:
  - You as quality manager, coordinating an evaluation, have to be neutral, have integrity and must understand the evaluation without prejudice and fixed expectations. You must not take sides in the evaluation. This also counts for the stakeholders involved in an evaluation working group.

- The objectives of an evaluation should be communicated clearly, even if it might produce resentment and arguments from the involved stakeholders. "Hidden agendas" (see Chapter 1.3) are to be avoided. It is your task as quality manager to openly address anxieties and resentment and any possible resistance by stakeholders. This can be done by clarifying the backgrounds of decisions and the discussions that have taken place before the evaluation.
- Participation of stakeholders should only be allowed, if it is really wanted (see Chapter 2.3). This means that opinions and recommendations of the actors involved should only be asked for, if the evaluation process is flexible enough to include those in the process. A pseudo-participation can make an evaluation implausible and also a quality manager seem unreliable.

Chapter 3

Instruments of Empirical Social Research as Tools for Evaluation

⊙ **On successful completion of this chapter, you should be able to...**

- ■ recognise the relevance of empirical social research for evaluation,
- ■ distinguish between techniques and methods of empirical social research,
- ■ describe and differentiate between main purposes and approaches of empirical social research,
- ■ name the three fundamental questions of empirical social research,
- ■ distinguish between quantitative and qualitative empirical social research strategies and decide which approach is appropriate in a particular given situation,
- ■ describe and differentiate between quantitative and qualitative data,
- ■ name the purpose of social research methods.

3 Instruments of Empirical Social Research as Tools for Evaluation

3.1 The Relevance of Empirical Social Research for Evaluation

As described in Chapter 2, in higher education we use evaluation to gather and revise processes and their effects on research, teaching and administration. Thereby, the revision is based on information that informs about such processes and their effects. That means, access and gathering such information is elementary for any planning, doing, checking and acting in evaluation processes (PDCA-cycle, see Module 1). The quality of such information has to meet scientific standards to be comprehensive and useful. To guaranty such scientific quality, we use methods and techniques of empirical social research that help to collect and analyse data.

We define data as information in terms of numbers, abstracted by using methods and techniques from social sciences (see Module 4, Chapter 1.1).

Difference between information and data

Such methods and techniques (see Chapter 4.3) are manifold and can be used for different evaluation processes (see Chapter 2).

Quality managers do not have to be experts for all available methods, but they should know the broad variety of socio-scientific methods and techniques for collecting and analysing data. This knowledge gives them a fundamental basis to be able to deal with methods and techniques of empirical social research in evaluation processes themselves. Furthermore, quality managers should be able to interpret data (gained e.g. from surveys). Finally, a good methodological understanding is important to be able to interpret the quality of data that has not been collected by the quality manager him-/herself, but that derive from external sources. Neumann (2000) describes social research, to be considered by quality managers, as follows:

Why empirical social research matters

> *"Social research is a collection of methods people use systematically to produce knowledge. It is an exciting process of discovery, but it requires persistence, personal integrity, tolerance for ambiguity, interaction with others and pride in doing quality work."*
> *(Neuman 2000, 2)*

We can sum up the following conclusion:

Methods and techniques of empirical social research help to
- gather and collect,
- analyse, and finally
- interpret data and information to be able to
- deduce adequate activities to follow-up.

Difference between techniques and methods

Up to now, we have talked about socio-scientific methods and techniques, but we have not differentiated these two terminologies yet: We use the term techniques for instruments and tools necessary for adequate data collection. Methods describe the procedures in which such techniques or tools are to be used. For example, a questionnaire is a tool (technique), which is used to do a survey (method) (Kumar 2005, 126).

Based on the PDCA-cycle, methods and techniques of empirical social research are used for revision purposes ("to check"), but also to deduce and plan activities ("to act"; "to plan"). Dealing with methods and techniques usually includes a bias between procedures that is recommended in books and reality. For example, financial or time resources are measured too narrowly as recommended scientifically. Rossi, Lipsey, and Freeman (2004) summarise this bias as follows:

> *"(…) evaluations should meet high standards of scientific research and at the same time be dedicated to serving the information needs of program decision-makers. The problem is that in practice these two goals often are not especially compatible. Conducting social research at a high level scientific standard generally requires resources that exceed what is available for evaluation projects."*
> ***(Rossi, Lipsey, and Freeman 2004, 23 et seq.)***

Dealing with such challenges (e.g. in time or financial means) that come up in evaluation processes, we always have to try not to neglect or misuse the rules of empirical social research. Otherwise, the results can be arbitrary, and with it not reliable and wrong. Furthermore, results that are not scientifically understandable won't bear a critical revision and with it, won't be accepted. Considering the rules of empirical social research in a pragmatic way helps quality managers to realise a successful evaluation process. Rossi, Lipsey, and Freeman (2004) state:

> *"In practice, therefore, the evaluator must struggle to find a workable balance between the emphasis placed on procedures that ensure the validity of the findings and those that make the findings timely, meaningful, and useful to the consumers. (…) In many cases, evaluations will justifiably be undertaken that are "good enough" for answering relevant policy and program questions even though program conditions or available resources prevent them from being the best possible designs from a scientific standpoint."*
> ***(Rossi, Lipsey, and Freeman 2004, 25)***

> ### ? Questions & Assignments
>
> 1. What is the difference between information and data?
> 2. What is the difference between techniques and methods of empirical social research?
> 3. Describe what kind of methods and tecniques of empirical social research should be used in evaluation processes and why.
> 4. What challenges can arise when using methods and techniques of empirical social research at higher education institutions?

> ### 👓 Further Reading
>
> - Babbie, E. R. (2004). *The practice of social research* (10th edition). Southbank: Wadsworth, 342-365.
> - Neuman, W. L. (2000). *Social research methods: Qualitative and quantitative approaches* (4th edition). Boston: Allyn and Bacon, 1-19.
> - Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th edition). Thousand Oaks: SAGE, 21-30.

3.2 Purposes and Approaches of Empirical Social Research

As already described, methods of empirical social research are used to collect and analyse data systematically. This is not as easy as one might think at first glance, especially when considering the main purpose of empirical social research, which is to gather and analyse aspects or social phenomena of reality, which can be rather complex and cannot be observed in a direct way (e.g. interaction between lecturers and students; teaching and learning processes and their outcomes etc.).

Purposes of empirical social research

At higher education institutions, we use socio-scientific methods in evaluation processes, for example, to define the value and benefit of a specific study programme. Other examples that refer to evaluation processes based on methods and techniques of empirical social research include the following:

- To describe or explain phenomena or incidences (e.g. reasons for drop-outs),
- to make predictions on future incidences (e.g. the demand of additional training for bachelor graduates in master programmes),
- to deduce and evaluate interventions or activities to achieve certain defined targets (e.g. activities to minimise students drop-out by 5%).

When using methods and techniques of empirical social research in evaluation processes at higher education institutions, there usually exists a concrete demand. Depending on such demand, this may include surveys that are based on approaches in the fields of exploration, description and explanation (Babbie 2004, 87 et seqq.).

Approaches of empirical social research

Explorative data analyses are used to get an overview of research fields that are still unknown. That means, such surveys are not primarily used to determine reasons for certain incidences. For example, when a higher education institution wants to implement a new IT-system, this can have structural but also content-related influence on workplaces: Workplaces can be reduced, they can be up- or down-graded, or new workplaces can be developed. An explorative data analyses based e.g. on a survey of the staff of a higher education institution can collect data on expectations, challenges or the fears of those staff members that arise from the change to a new IT-system.

Based on this explorative data analyses, we can proceed with a **descriptive data analysis**. The purpose of such analyses is, e.g. to describe opinions of a certain stakeholder group as precisely as possible, e.g. based on statistical data. For example, one result of the explorative analyses to introduce a new IT-system at a higher education institution might be that the staff has the impression that students have difficulties designing their course schedules. If a module on online course schedule-designing in the planned new IT-system were possible, you could do a student survey to find out which functions students think would be useful to design their course schedules.

If you want to find out about the linkages between reason and outcome, you should do an **explanatory data analysis**. For example, at a higher education institution there is the hypothesis that students who have to work to be able to finance their studies, need more time to graduate and receive worse grades than other students. Such hypotheses can be assessed by an explanatory analysis. Hypotheses in general describe rather specifically what you expect to happen in your study. This example also makes clear why we talk about empirical social research: Hypotheses based on theoretical considerations are revised with regard to the respective social reality (Kumar 2005, 74 et seqq).

Research without hypotheses?

At the same time, not all data analyses are based on hypotheses. Sometimes, in an exploratory designed analysis there is no formal hypothesis, and perhaps the purpose of the analysis is to develop some specific hypothesis or prediction that can be tested in future research (Trochim 2005, 8). That means evaluation processes at higher education institutions that include methods and techniques of empirical social research are not always based on theoretical approaches. In this case, they refer to evaluation criteria that have the function of hypotheses. Such evaluation criteria can be defined by the authorities that order an evaluation (e.g. vice-chancellery), by the target group (e.g. members of a faculty), participating stakeholders (e.g. senate), a quality manager or by all parties together. In this case, the purpose of empirical social research is to collect and analyse data systematically based on the defined evaluation criteria (Stockmann & Meyer 2010, 77 et seqq).

Based on this, we can conclude that evaluation processes at higher education institutions cannot always be defined in the narrow meaning of socio-scientific research projects. Instead, they include socio-scientific methods and techniques to define, e.g. the value and the benefit of a study programme. In this case, evalua-

tion processes are still evidence-based because they refer to empirically collected and scientific findings that help in the decision-making process (McMillan & Schumacher 2010, 430).

To combine hypotheses or evaluation criteria, data collection and analyses systematically, you should consider some fundamental questions at the beginning of each process that includes methods and techniques of empirical social research (Atteslander 2010, 4). In evaluation processes it may occur that it is the task of the quality manager to pose such questions. In case the quality manager is not responsible, s/he should ensure that somebody else answers the following three fundamental questions that are closely linked and depend on each other:

<div style="float:right">Three fundamental questions of empirical social research</div>

1. *For what reasons is data to be collected?* This question refers precisely to the purpose of using the data: Why and for what reason should we do a certain evaluation? What are the hypotheses for such evaluation?
2. *What data is to be collected?* This question refers precisely to the research question: What exactly do we want to find out based on the evaluation?
3. *How is the data to be collected?* This question refers precisely to the methods of collecting and analysing data that shall be used for the evaluation.

Depending on how we answer these questions, we can select the adequate methods and techniques of empirical social research to answer our question (see Chapter 4.3).

(?) **Questions & Assignments**

1. Please give 2-3 examples of using methods and techniques of empirical social research at your higher education institution.
2. Explain the differences between explorative, descriptive and explanatory data analyses.
3. Please give examples of evaluation criteria that could be defined for your higher education institution.
4. Please name the three fundamental questions of empirical social research and explain why they are important with regard to evaluation criteria, data collection and the purpose of evaluation.

3.3 Empirical Social Research Strategies

In addition to the approaches described in Chapter 3.2, we can also use different research strategies. These can be divided into quantitative and qualitative strategies. They differ essentially with regard to the methods of data collection because the result of quantitative strategies is numerical data, meanwhile of qualitative strategies it is non-numerical data (Babbie 2004, 26 et seqq.).

Quantitative and standardised strategies and data

Quantitative research strategies are structured procedures, in which the design of the research process and the expected data material is already defined in advance (Kumar 2005, 12) and the respective incidences are described based on numbers. For example, if you can describe student satisfaction with their study programme with numbers from 1 to 5, these are numerable and countable quantitative data. Since the students can define their satisfaction in the same way (by numbers from 1 to 5), we call it a standardised data collection that also delivers standardised data.

A big advantage of quantitative strategies is that due to this standardisation the collected data can be analysed relatively easily and in huge quantities with software for statistical analyses, because of the numerical data basis.

Qualitative and non-standardised strategies and data

Contrary to quantitative strategies, qualitative research strategies put more emphasis on words or text than on numbers. They focus on relationships and describe the social world as seen through the eyes of the subjects to discover how it is constructed. Qualitative research strategies usually collect data in a non-standardised way. For example, a non-standardised form of data collection is when students' satisfaction (see Chapter 4.3) is described in an interview verbally (e.g. based on the question to describe the satisfaction with the study programme). The resulting text is defined as qualitative data because the students' statements exist as a non-numerical text. A big advantage of qualitative strategies is that the collected data has a high significance because of the missing standardisation. That means, the opinions and views of the interviewed persons can be collected without restrictions.

Differences between quantitative and qualitative data

The following table gives an overview on the differences between quantitative and qualitative data (Babbie 2004, 26 et seqq.; Trochim 2005, 121 et seqq.).

Data	
Quantitative	**Qualitative**
Deals with numbers, more explicit	Deals with descriptions, can be richer in meaning
Data can be measured	Data can be observed (not measured)
Easier to aggregate, compare and summarise	Can be coded quantitatively
Opens up possibility of statistical analysis	

Table 10 *Differences between quantitative and qualitative data*

Coding: Transformation from qualitative to quantitative data

Up to now, we have stated that qualitative data do not exist in numerical form (other than quantitative data), but as words or text. This is not completely accurate, because qualitative data can be translated into quantitative data, when words or text are defined with a numerical value. In doing so, text-based answers or comments on questions such as "Please add any additional comments", can be categorised. Each of such categories receives a numerical value (also named as codes), which can be analysed statistically. Based on such coding, qualitative data is transformed into quantitative data (Trochim 2005, 121 et seqq.).

Example: Coding of Qualitative Data

In a student survey, ten students shall indicate their sex and origin. Since there are no pre-defined answers, the students write down their answers in text-form. That means we receive qualitative, non-numerical data. To be able to calculate how many male and female students have answered and where they are from, we need quantitative, numerical data.

That is why we code the answer "female" with 1, and "male" with 2. Based on this, we can see that five male and five female students have participated in the survey.

Also the answers on the origin are given codes. We can see that seven students are from Bangkok, two from Chiang Mai and one from Singapore.

Respondent	Sex (qualitative)	Sex (quantitative)	Origin (qualitative)	Origin (quantitative)
Student01	Male	2	Bangkok	2
Student02	Female	1	Bangkok	2
Student03	Male	2	Chiang Mai	1
Student04	Female	1	Bangkok	2
Student05	Male	2	Chiang Mai	1
Student06	Male	2	Singapore	3
Student07	Female	1	Bangkok	2
Student08	Female	1	Bangkok	2
Student09	Male	2	Bangkok	2
Student10	Female	1	Bangkok	2

The example shows that the choice of research strategy does not necessarily determine exclusively the format of the data to be analysed, but in practice it often does. Deciding to use a quantitative strategy means that you will mostly receive numerical data (Babbie 2004, 26). Quality managers should be aware that evaluation processes usually consist of qualitative and quantitative elements. Due to this, a good understanding of quantitative and qualitative research strategies is rather helpful and recommendable.

The following table gives an overview on different aspects of quantitative and qualitative research strategies that should be considered when choosing an appropriate strategy:

Aspect	Research strategy	
	Quantitative	**Qualitative**
Approach	Structured, rigid, predetermined	Unstructured, flexible, open
Main purpose	Measure objective facts Quantify extent of variation in a phenomenon	Construct social reality, cultural meaning Describe variation in a phenomenon
Sample size	Emphasis on greater sample sizes	Fewer cases
Dominant research topic	Explains prevalence, incidence, extent, nature of issues, opinions and attitude; discovers regularities	Explores experiences, meanings, perceptions and feelings
Analysis of data	Numeric data to frequency distributions, cross-tabulations or other statistical procedures	Responses, narratives or observation data to identification of themes and describes these

Table 11   *Differences between quantitative and qualitative research strategies (Kumar 2005, 17 et seqq.) (own table)*

Quantitative or qualitative: Which strategy is suitable?

Remembering the fundamental questions mentioned in Chapter 3.2, we can see that answering these questions also helps to find an appropriate research strategy. Asking "why" helps to find an answer with regard to the purpose: On whom or what shall the analysis focus? On individuals (e.g. students' learning success) or social interaction or processes (e.g. outcome of a tutorial programme)? That means, the analysis and the purpose determine the quantitative or qualitative values to be examined (Trochim 2005, 11). Only afterwards can we ask "what" to analyse: What is the concrete research question? Who can help to answer this research question? In the following, we can answer the third question on "how" to do the data collection and analysis. Having answered these questions, we can decide if qualitative or quantitative strategies are more suitable to answer the questions of the evaluation process.
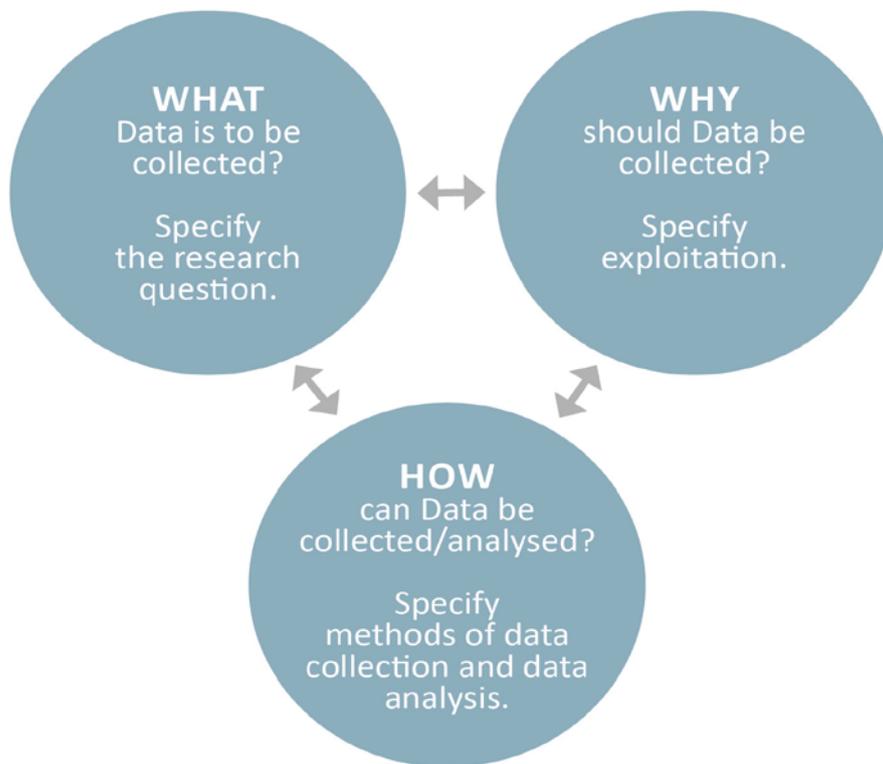
Figure 3   *Questions to find an appropriate research strategy*

⑦ Questions & Assignments

1. What is the difference between quantitative and qualitative research strategies? Please name key advantages of each strategy.
2. You were given an example of coding qualitative data. Please describe another example for data collection, as is carried out in an evaluation process at your own higher education institution.
3. Please describe a possible approach to finding out about an appropriate research strategy.

3.4   Empirical Social Research Methods

Having discussed quantitative and qualitative strategies of empirical social research and learned about their differences, we will now find out about what exactly methods of empirical social research are. In Chapter 3.1 we have already learned that methods are procedures, in which techniques and tools of data collection are used. In social sciences, there exist manifold recognised methods that can be used for data collection. These can be methods such as observations, surveys, experiments or content analysis (more detailed description see Chapter 4.3).

Using these methods and their techniques helps to plan and structure data collection in such a way that it produces efficient data that are valid and understandable. As in any other research, social research also has to accomplish the following characteristics: "it must, as far as possible, be controlled, rigorous, systematic, valid and verifiable, empirical and critical." (Kumar 2005, 7). Due to the already described bias between scientific recognition and use-oriented application of evaluation processes, a responsible use of techniques and methods of empirical social research helps to reduce mistakes of data collection and for the resulting outcomes to be useful and acceptable.

The following table shows which method can be used for what:

What shall be examined?	Method
Opinions of human beings	Survey
Human behaviour (in natural situations)	Survey, Observation
Human behaviour (in experimental situations)	Experiment
Products of human beings (e.g. text)	Content analysis

Table 12   *Outcome-based research methods*

Based on this, we will focus more in detail on the survey-method, since for quality managers, it probably is one of the most important methods to collect data.

> ### ? Questions & Assignments
>
> 1. Please name different social research methods.
> 2. What are purposes of social research methods in evaluation processes?

Chapter 4

Empirical Social Research Process: From Research Question to Data Collection

On successful completion of this chapter, you should be able to…

- follow the main phases of an empirical study when conducting your own evaluative projects,
- describe and explain the differences between empirical study designs,
- distinguish several data acquisition methods,
- discuss advantages and disadvantages of survey research methods,
- judge the degree of representativeness of a survey,
- choose appropriate data collection methods according to the given framework conditions.

4 Empirical Social Research Process: From Research Question to Data Collection

4.1 Ideal Phases of an Empirical Study

Before getting to know methods and techniques of empirical social research more in detail, we will focus on the ideal phases of socio-scientific analysis. These phases are similar in structure to the phases of an evaluation and they are equivalent in content to an elaborated data collection phase (see Chapter 2.4).

A socio-scientific analysis can be structured into five ideal phases. Ideal means that each phase is relevant and should be considered for the analysis, meanwhile within a phase, aspects of less importance can be neglected for a pragmatic procedure (see Chapter 3.1). In each phase we have to make decisions and take further steps that depend on and influence each other. That's why these phases cannot always be exactly differentiated, but there are fluent cross-overs between the phases.

Five ideal phases of an empirical study

In the **orientation and definition phase (phase I)**, we define the concrete subject-matter to be analysed. In evaluation processes, the subject-matter is often determined by the principal authority, e.g. the vice-chancellory of a higher education institution. Sometimes, such requests can be rather vague, so that one or more research questions have to be formulated more specifically. Therefore, we can use a literature review. If this is not enough, we can also describe the type of statements to be formulated after the analysis. An example for such a statement could be: X% of students that have to work besides studying need more time for finishing their studies than the average. Describing such statements helps to formulate the necessary hypotheses or evaluation criteria (see Chapter 3.2).

The **preparatory phase (phase II)** is used for the methodological planning and preparation of data collection. In this phase, we clarify and conceptualise terminologies (see Chapter 5.1). Afterwards, we have to decide how to analyse such concepts, that means, which method is to be applied. At the same time, we have to decide, on whom or what we want to make statements (e.g. people, groups, documents), and, based on this, whom we have to interview in surveys or what we have to examine in content analysis (e.g. module descriptions). Furthermore, we should also define a concrete time schedule and budget plan to facilitate the academic management.

In the **data collection phase (phase III)** we technically carry out the data collection. Depending on our methodological planning of phase II we finish the operationalisation (see Chapter 5.2) of the empirical study, for example, by designing and testing a questionnaire for a survey (see Chapter 5 and 6).

At the beginning of the **data analysis phase (phase IV)** we construct a dataset that can be analysed. This includes, for example, the coding of data (see Chapter 3.3) and the control of mistakes (see Chapter 4.5).

Only after having finished these steps, we can continue with the statistical data analysis (see Chapter 7). Finally, an empirical study is completed with the reporting and dissemination phase (phase V). In this phase we interpret and disseminate the results, e.g. by writing a research report. Based on the results we can deduce and realise further activities
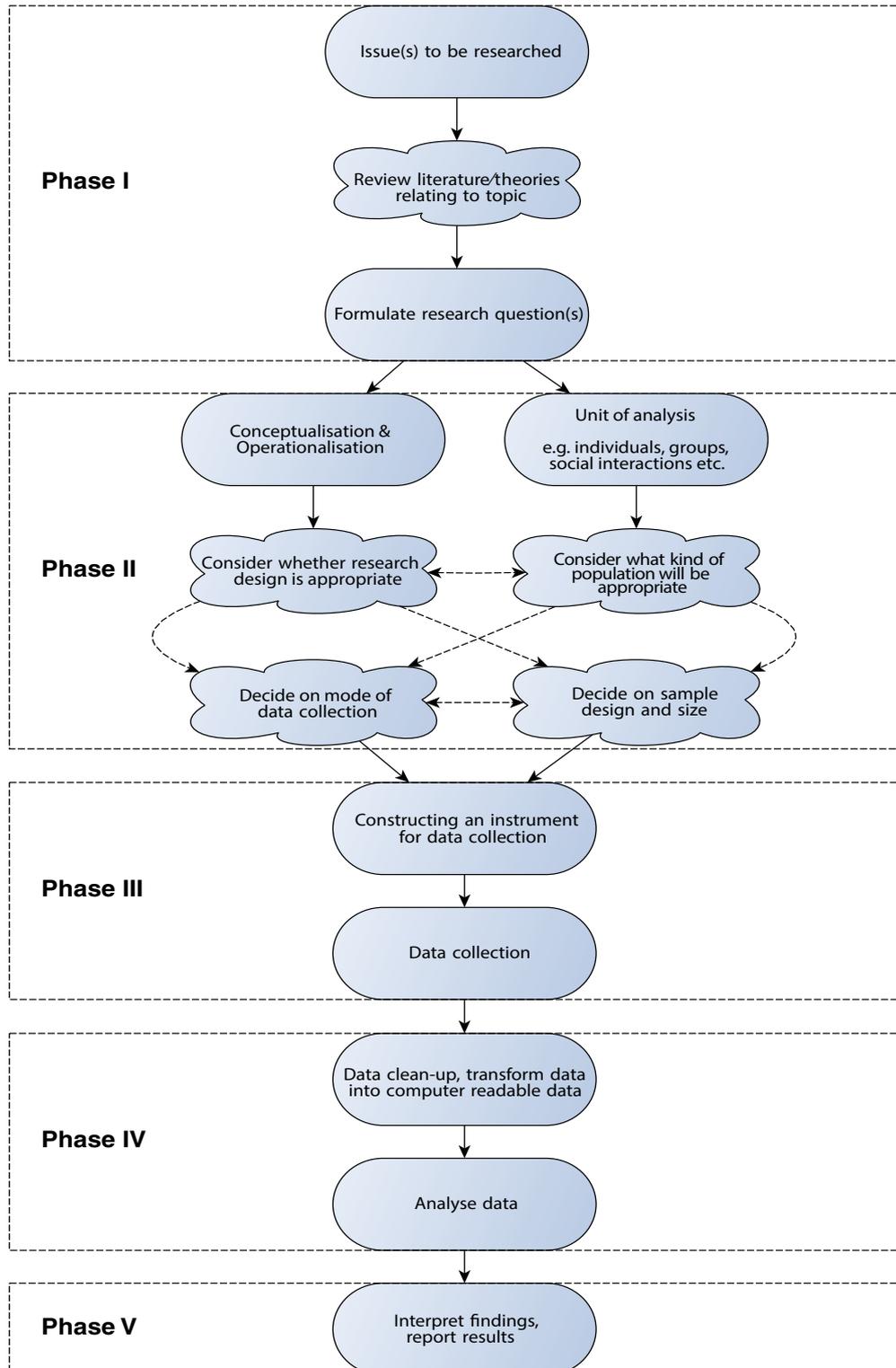
.

Figure 4 *Five ideal phases of an Empirical Study*

⊙ **Questions & Assignments**

1. Please look up an example of an evaluation process at your higher education institution and make a brainstorming or a mind map for structuring the phases I to III.

👓 **Further Reading**

■ Kumar, R. (2005). *Research methodology: A step-by-step guide for beginners* (2nd edition). London: SAGE, 15-25.

4.2 Overview: Empirical Social Study Designs

Having discussed the ideal phases of empirical studies (see Chapter 4.1), we will now look at the preparatory phase (phase II) more in detail: Based on the orientation and determination phase (phase I), we now have to decide on how to proceed with the data collection. We have to clarify: Which methods are suitable to find answers for our research questions? How can we design the different phases of the study process to receive the data needed? Fortunately, social science gives us various research designs that help us to do so.

Basically, research designs have two main functions: First, they help to identify possible methods and techniques for data collection that are appropriate for the respective purpose. Second, they make it easier to guarantee the necessary quality of the collected and analysed data (Kumar 2005, 84). Research designs provide the frame for data collection and analysis. A choice of research design reflects decisions about the priority being given to a range of dimensions of the research process (Bryman 2004, 27).

Two main functions of research design

In the following, research designs – also named study designs – that are often used in evaluation processes will be presented.

The cross-sectional study design is most often used in social sciences. In cross-section studies data is collected on a defined time based on a singular observation or selection. This design is not about collecting single cases, but creating a broad data basis on a certain subject-matter. Exploratory and descriptive studies are often cross-sectional (Babbie 2004, 101 et seqq.).

Cross-sectional study design

An example for a cross-section study design is, when a higher education institution carries out a survey of graduates who belong to a specific study period. The result of such a singular survey can be a snap-shot, for example of the retrospective revision of study conditions. Nevertheless, based on a single cross-sectional study design, it is not possible to examine if study conditions have changed.

Longitudinal study design

Other than cross-sectional study designs, longitudinal study designs consider at least two data collections at different times. That is why they are useful to examine changes chronologically. In sum, there are three particular types of longitudinal studies to be described in the following: trend studies, cohort studies and panel studies (Babbie 2004, 102 et seqq.).

Trend, cohort and panel study design

Trend studies examine changes chronologically within a population (see Chapter 4.4). Thus, trend studies are repeating cross-sectional studies on the same subject-matter. **Cohort studies** also include data collection at different particular times, but they focus on members of a specific group whose composition changes in the course of time. Such specific groups are also known as cohorts. One characteristic of a cohort is that its members share a common attribute (e.g. birth in the year 1986; marriage in the year 2005; study start in winter semester 2011/12) that also influences more or less their respective course of life. **Panel studies** examine phenomena at different times, but they always consider the same group of people. That means, in this case the group composition does not change.

Panel attrition

A challenge of dealing with panel studies is the so-called panel attrition: People who participate in the first survey (wave) of the panel study, do not participate anymore in the second or following waves. This is challenging because we do not know if such people are important for our study, and will no longer be considered, which might change the results.

The following table may help you to distinguish trend, cohort and panel studies:

Study Design	Objective	Issue	Surveys and Populations	Value of Statement
Trend	Changes within a population over time	Conditions for studying	Survey 1 in 2015: first year students of 2015 Survey 2 in 2017: first year students of 2017	Examine, if first year students in 2017 differ in their opinion about study conditions than first year students in 2015.
Cohort	Changes within a cohort over time	Conditions for studying	Survey 1 in 2015: first year students of 2015 Survey 2 in 2017: first year students of 2015	Examine, if the members of the cohort "first year students in 2015" change their opinion about study conditions in 2017 compared to 2015.

Study Design	Objective	Issue	Surveys and Populations	Value of Statement
Panel	Changes within same set of people over time	Professional success	Survey 1 in 2015: graduates of 2015 Survey 2 in 2017: respondents of Survey 1 Survey 3 in 2019: respondents of Survey 1 & 2	Examine the development of the individual professional success of graduates of the year 2015 during the years 2015 until 2019.

Table 13 *Distinction between trend, cohort and panel studies*

Having learned that longitudinal studies examine changes chronologically, we now focus on study designs that examine the outcome of an intervention (e.g. a measure to improve study conditions at a higher education institution). Therefore, we can basically distinguish **"before-and-after studies"** and **"after-only studies"**.

A **before-and-after design** can be described as two sets of cross-sectional data collection points on the same population to find out the change in the phenomenon between two points in time. The change is measured by comparing the difference in the phenomenon before and after the intervention (Kumar 2005, 95). This design is very often used in evaluation processes. For example, we can examine, if the evaluation of study conditions at a certain time 1 changes after an intervention from the evaluation of the same study conditions at a time 2 (see Figure 4).
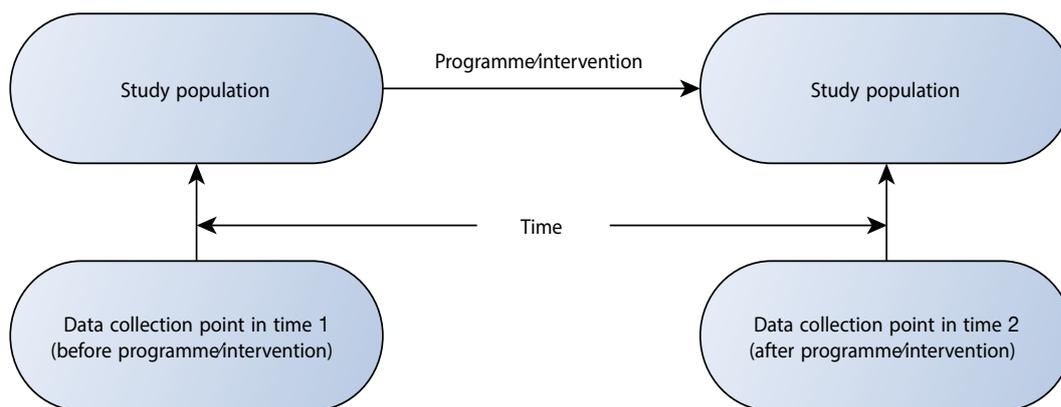
Before-and-after study design

Figure 5 *Before-and-after study design (Kumar 2005, 95)*

We can observe that higher education institutions very often use after-only designs for evaluation processes instead of before-and-after designs. In this case, we measure the outcome of an intervention after it has taken place. Usually, a starting value that represents the status before the intervention (see Figure 4, data collection point in time 1) and that can be used for comparison purposes is not determined. That means, the effectiveness of the intervention

After-only study design

is only evaluated based on data collected after the intervention (see Figure 4, data collection point in time 2), e.g. on the basis of respondent's recall of the situation before the intervention or from available existing data (Kumar 2005, 102 et seqq.). Usually, the reasons to do such after-only studies are rather pragmatic: Since this design only includes one wave of data collection, it reduces time and financial resources that very often are not available.

Case study design The aforementioned studies describe designs that consider a broad data basis, including many cases (e.g. the respondents of a survey). By contrast, a **case study** considers fewer cases, sometimes even only one case. This study design entails the detailed and intensive analysis of a single case which may be a programme, an event, an activity or a set of individuals bounded in time and place (McMillan & Schumacher 2010, 24). They are used to examine a case, e.g. a loan programme on a rather holistic basis. Furthermore, they are used when there are time-space constraints (e.g. an incident occurs only once) or staff constraints (e.g. very few people that are in a certain position). An advantage of such a generally explorative design (see Chapter 3.2) is that we gather detailed results about opinions and relationships among people in rather specific situations. A disadvantage is that it is difficult to transfer the results to other cases.

❓ Questions & Assignments

1. Please describe the two main functions of study designs.
2. What is the advantage of longitudinal studies compared to cross-sectional studies?
3. What are differences between trend, cohort and panel studies?
4. Please describe study designs that examine outcomes.

🔭 Further Reading

- Babbie, E. R. (2004). *The practice of social research* (10th edition). Southbank: Wadsworth, 101-106.
- Bryman, A. (2004). *Social research methods* (2nd edition). Oxford: University Press, 33-54.
- Kumar, R. (2005). *Research methodology: A step-by-step guide for beginners* (2nd edition). London: SAGE, 83-113.

4.3   Overview: Methods of Data Collection

Having discussed qualitative and quantitative research strategies in Chapter 3.3, we know that the choice of a research strategy mostly determines the expected data material in advance. In Chapter 3.4 we have learned that methods of empirical social research such as content analysis, observations or surveys help to plan, structure and realise data collection efficiently. In Chapter 4.2 we got to know that there are many study designs to be realised based on surveys and we discussed respective examples from the higher education context. Since quality managers at higher education institutions often have to deal with survey methods and techniques, we now focus on the most important survey research methods that are usually based on quantitative data. Before doing so, we will first have a quick look at the methods of content analysis and observations, because these are also often used in evaluation processes at higher education institutions.[2] These methods are usually based on qualitative data.

In a **content analysis**, we collect and analyse data from different sorts of text such as radio or TV comments, images, movies etc. The analysis can be quantitative, qualitative or both (Trochim 2005, 127). Evaluation processes at higher education institutions that include content analysis usually consider text such as strategic development reports, module/course descriptions, assessment regulations, but also publication registers etc. For example, if a higher education institution wants to find out to what extent the study programmes have an international profile, it could examine the programme regulations based on a content analysis, asking if they include obligatory curricular semesters to study abroad, or if they consider opportunities to study abroad voluntarily (for a deeper overview on content analysis refer to Babbie 2004, 314 et seqq.).

Content analysis

Observations play a rather secondary role for the work of quality managers at higher education institutions. An observation is a purposeful, systematic and selective way of watching and listening to an interaction as it takes place. We talk about a participant observation, when a researcher participates in the activities of the group being observed in the same manner as its members, with or without their knowing that they are being observed. It is a non-participant-observation when a researcher does not get involved in the activities of the group but remains a passive observer, watching and listening to its activities and drawing conclusions from this (Kumar 2005, 119 et seqq.). We use observations at higher education institutions, for example, when evaluating a didactical concept, examining the interaction between lecturers and students (for a deeper overview on observation refer to Bryman 2004, 164 et seqq.).

Observations

Finally, we discuss **survey research methods**, probably one of the most often used forms of data analysis in social sciences. Many research questions that examine, e.g. opinions or behaviours of human beings can be analysed effectively based on surveys.

Surveys – most commonly used research method

We can differentiate three key aspects with regard to surveys:

First, there exist differences in the level of structuring (or standardisation of) a survey. We can differentiate between a structured (or standardised), a partly-structured (partly-standardised), and a non-structured (non-standardised) survey.

Distinctions of surveys: Levels of structuring, administration and media

---

2 We won't discuss experimental methods in this chapter, because they are rarely used in evaluation processes at higher education institutions (for further information on experiments refer to Bryman 2004, 34 et seqq.).

Second, we can distinguish, if questionnaires are completed by the respondent with or without the help of an interviewer. In the second case, we call it self-administered survey (e.g. paper-based survey). In the first case, it is an administered survey (e.g. a telephone interview, in which the interviewer can help with the completion of the questionnaire or comment on questions). We will see that also computers may obtain the role of an interviewer to administer a survey (e.g. web-survey). Basically, a data collection is an interview, only if a person (an interviewer) asks questions of another (a respondent), including interaction between the interviewer and the respondent (Babbie 2004, 263 et seqq.).

Third, we can differentiate if surveys are completed orally (e.g. telephone interview) or written (e.g. paper-based questionnaire).

Structured surveys

Let's have a closer look at the level of structuring surveys. In **structured surveys** all questions of a questionnaire are formulated in detail and their order and the possibilities for answers is pre-determined. That means, the whole survey is realised based on the same standard. Especially written surveys usually consist of a high level of standardisation. The purpose of structured surveys is to collect data (of a large number). Due to the standardisation, such data can be aggregated (see Chapter 7.1) and evaluated easily, because the comments of the respondents all exist in the same way.

Aggregation of data

> Aggregation of data means to summarise information on a broader level. For example, students from physics, chemistry and biology can be summarised as students from natural sciences.

Semi-structured surveys

A **semi-structured survey** only consists of an interview guideline that helps to structure the survey. If necessary, it is possible to diverge from the guidelines, for example, to better understand a subject-matter by asking further questions.

Unstructured surveys

In an **unstructured survey** we do not use a questionnaire or an interview guideline at all. There are only pre-defined subject-matters or keywords to be discussed during the survey. We often use semi-structured or unstructured surveys for qualitative study designs, while structured surveys are often used for quantitative study designs.

Administered and self-administered surveys

Having discussed the level of structuring surveys, we now differentiate different forms of administrating surveys and the question, if respondents are supported when answering the questions or not.

Paper-based and mail surveys

Paper-based surveys are **administered surveys**, when several respondents complete a questionnaire at the same time (e.g. course evaluation) and an interviewer is present to answer questions. **Self-administered surveys** are those, in which a respondent group receives a questionnaire via mail to be completed by a certain deadline and sent back to the interviewer (e.g. mail survey in case of tracers' studies).

Computer-assisted surveys

Hence, paper-based surveys can be administered as well as self-administered. Electronic surveys are usually self-administered, even though the computer can help the respondents (e.g. based on defined programmed codes to use the standardised questionnaire template). Electronic surveys can be differentiated into email surveys or online surveys. The first refers to questionnaires that are sent via email (e.g. as attachment). The respondent has to answer the survey and resend it to the sender. Online surveys (or also web surveys) are provided on a special website and

are completed by using a web browser. Usually, the data is directly saved in an online data base. Currently, there already exist diverse software solutions for web surveys that are cheap or even free of charge. Very often, these already include an analysis function. That is why email surveys are not that common anymore. Still, it makes sense to invite the respondents and send them the link to a web survey via email (if the respective email contacts can be provided).[3]

Telephone surveys are a rather common form of administered surveys. They can be described as structured, written surveys. By contrast to paper-based surveys, questions are read to the respondent during a telephone call. The responses are usually directly transferred into an electronic data base on a computer by the interviewer.

Telephone surveys

Telephone surveys can also be considered as a special form of **face-to-face interviews**, during which an interviewer reads the question to a respondent, and writes down the answers paper-based or electronically on a computer.

Face-to-face surveys

If several respondents participate, we also speak of group interviews. Such group interviews have to be distinguished from focus groups, which is a guided group discussion. The interviewer acts more as a moderator, only introducing a subject-matter and taking care that all respondents participate in the discussion. The purpose of focus groups is to gather the opinions of the participants with regard to a certain subject-matter. An advantage is that the group interaction and group dynamic help to receive more detailed information than in an interview with only one person.

Focus groups and group discussions

The following table illustrates the level of structuring, the mode and characteristics of different survey forms.

Survey	Mode	Example	Characteristics
Structured	Oral	Individual or group interview, telephone survey (administered)	Rigid structure, rigid contents, rigid questions and wording To collect quantitative data and aspects "Measure"
	Written	Mail survey, online survey (self-administered)	
Semi-structured	Oral	Guided discussion, group interview (administered)	Different levels of flexibility and specificity To collect quantitative and/or qualitative data and aspects "Interpret" and/or "measure"
	Written	Expert survey (self-administered	
Unstructured	Oral	Expert survey, group discussion (administered)	Flexible structure, flexible contents, flexible questions and wording To collect qualitative data and aspects "Interpret"
	Written	Informal survey of experts (administered)	

Table 14 *Characteristics of surveys*

---

3 For more detailed information on web surveys refer to Couper 2008.

(?) **Questions & Assignments**

1. Why might a quality manager prefer to use a structured rather than an unstructured survey method for gathering data?
2. Please discuss to what certain extent survey methods are not applicable at your higher education institution and why.

👓 **Further Reading**

■ Babbie, E. R. (2004). *The practice of social research* (10th edition). Southbank: Wadsworth, 256-280.
■ Neuman, W. L. (2000). *Social research methods: Qualitative and quantitative approaches* (4th edition). Boston: Allyn and Bacon, 271-285.
■ Salant, P & Dillman, D. A. (1994). *How to conduct your own survey*. New York: Wiley, 101-136.

4.4   Sampling and Representativeness

Full census and sample surveys

When doing a survey, sometimes it is not possible to do a full census. One reason for not being able to ask all people, the so-called population of a study, can be that the population is too big and it would take too much time to consider all of them. For example, a survey of all students of a higher education institution is often not possible due to financial and/or time constraints. That is why we often make so called sample surveys. Sample surveys have the ability to obtain information from relatively few respondents to describe the characteristics of an entire population and it takes less time and money to interview few respondents than to interview many (Salant & Dillman 1994, 53). To create samples, we select some units (a sample) from a bigger group of units (the sampling population).
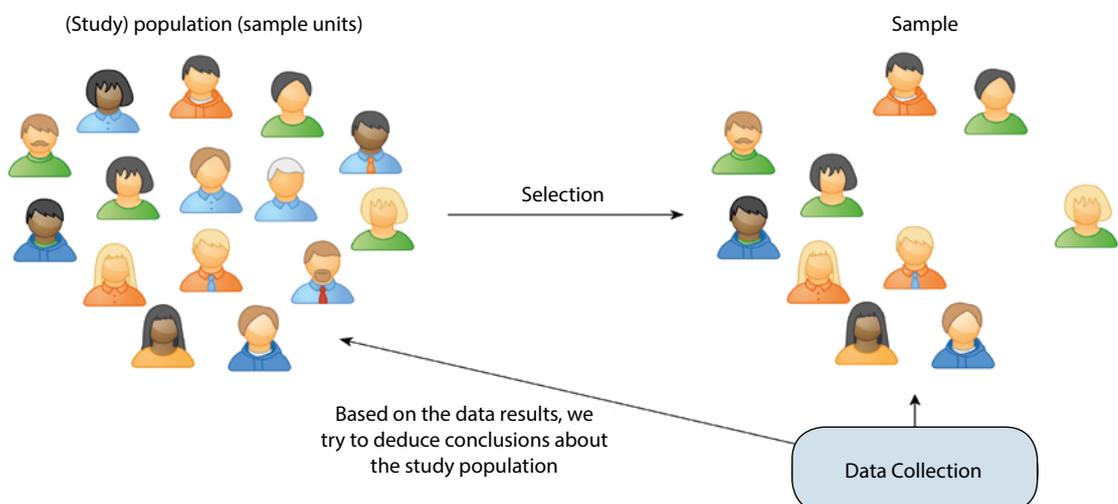

Figure 6 *(Study) population and sample (Kumar 2005, 164)*

To be able to make conclusions on a population without doing a full census, the group to be asked (the sample) has to be similar to the population. That means, the distribution of characteristics among the study population and the sample should be nearly the same. If this is not the case, we call it a sampling bias. Those selected would not represent the study population they have been chosen from. To give an example: In a study population there are 45% women and 55% men. The sample should have the same percentage proportion. That means, a sample is representative when it reflects the population accurately so that it is a microcosm of the population (Babbie 2004, 178 et seqq.).

<div style="float:right">Represen-
tativeness
and sample
surveys</div>

To create representative samples, we can use different methods (Kumar 2005, 171 et seqq.; Babbie 2004, 186 et seqq.). For example, a sample can be based on a random selection. We call it random selection when each unit of the population has an equal or independent chance to be selected. "Equal" in this case means that the chance to be selected should be the same for all, and no units can be selected several times. "Independent" means that the selection is only determined by chance and no other factors (e.g. preferences of the one who makes the selection). If samples are selected based on the mentioned method, the chance to receive a sample that considers the specific and measurable characteristics of the population in a similar proportion is relatively high.

⁇ Questions & Assignments

1. Please define what you understand by the term "representativeness".
2. Which characteristics should a sample have to be as representative as possible?

🔭 Further Reading

- Babbie, E. R. (2004). *The practice of social research* (10th edition). Southbank: Wadsworth, 178-217.
- Bryman, A. (2004). *Social research methods* (2nd edition). Oxford: University Press, 83-106.
- Groves, R. M., Fowler, F. J., Couper, M., Lepkowski, J. M., Singer, E. & Tourangeau, R. (2009). *Survey methodology* (2nd edition). Hoboken, NJ: Wiley, 69-139.
- Salant, P & Dillman, D. A. (1994). *How to conduct your own survey*. New York: Wiley, 53-71.

4.5   Data Collection Methods: Challenges and Difficulties

When collecting data with socio-scientific methods and techniques, there are always various sources of mistakes which are of bad influence on the data collection and with it on the results to be achieved. Referring to the chapter on sampling and representativeness (see Chapter 4.4), it becomes clear that our results can distort and that we should not generalise them when using a non-representative sample.

Besides challenges that refer to the sampling, we also have to consider that even more effects that cause distortions or mistakes can occur with regard to surveys – especially with regard to interviews (see Chapter 4.3).

> *"In this context, mistakes are not understood in the colloquial way but they are deviations of what is desired in the survey process from what is attained."*
> **(Groves et al. 2009, 40)**

Diverse sources of mistakes while collecting data

Catalysts or effects that cause mistakes can be interviewers and respondents, a questionnaire or single questions of a questionnaire, but also technical means to realise a survey (e.g. instable internet access). Basically, we can distinguish various sources of mistakes, as illustrated in the following table (Bryman 2004, 105 et seqq.):

The selection of sources and examples mentioned above shows that at almost any moment of a data collection process there are risks that may lead to mistakes. Sometimes, we cannot completely exclude sources of mistakes (e.g. we should not force respondents to participate in a survey because this also causes mistakes in such a way that respondents give false answers under compulsion). Nevertheless, we have learned that there are various methods and techniques (e.g. with regard to the sampling) that help to reduce such sources of mistakes (see Chapter 4.4). More references on how to reduce mistakes in data collection (see Chapter 6) and data processing (see Chapter 7) will be given in the respective chapters, later on in the course book.

Source	Example(s)
Sampling and sample-relation	Difference between a sample and the study population from which it is selected, even though a probability sample has been selected Respondents completely refuse to participate in an interview (self-selection, unit-nonresponse) Respondents do not answer single questions of a questionnaire (item-non-response)
Data collection	Poor question wording or the way the question is asked by the interviewer. Misunderstanding or memory problems by the respondent. Social desirability: Respondents can show a tendency to answer questions in a way that is regarded as favourably by others.
Data processing	Faulty management of data Errors in coding answers

Table 15 *Sources of mistakes in surveys*

👀 Further Reading

■ Groves, R. M., Fowler, F. J., Couper, M., Lepkowski, J. M., Singer, E. & Tourangeau, R. (2009). *Survey methodology* (2nd edition). Hoboken, NJ: Wiley, 39-211.

4.6 Guidelines for Determining the Appropriate Survey Method

The selection of an appropriate survey method is probably one of the most difficult decisions during a study. In deciding for or against a certain method, it helps to answer the following questions that are oriented on the three fundamental questions of empirical social research (see Chapter 3.2):

■ What do you want to know?
■ What do you want to do with the survey results?
■ Whom do you want to ask? How can you reach them?
■ How do you want to ask? How can they answer your question?

For example, the study population greatly influences which survey method is suitable or not. Online surveys are not eligible, if respondents do not have a computer with internet access, even though they may have an email-address that could be used for the survey administration. Survey methods with interviewers are not possible, if there are no sufficient human and financial resources to do so. If results are needed quickly, semi- or unstructured survey methods are not suitable either.

The following Table 16 illustrates more questions to ask about opportunities and limitations of the survey designs, presented in this course book. Please note: All questions have to be answered against the background of the preconditions at your higher education institution. Some answers to the questions indicated in this table may vary because they are mutually dependent (e.g. a big sample increases the price of a face to face survey).

Question to Ask	Survey Design			
	Mail	Online	Telephone	Face-to-Face
Contact: What do I need to contact the respondent?	Postal address (if necessary, other information to find out address)	Email address (if necessary, other information to find out address)	Telephone number (if necessary, other information to find out the number)	Postal or email address or telephone number
Costs: How much financial resources are available for the survey?	Average (e.g. printing questionnaires, dispatching and packaging, return postage, data collection etc.)	Low (no costs for printing questionnaires, dispatching and packaging, return postage, data collection etc. – maybe costs for online survey-software)	High or very high (e.g. costs for computer assisted telephone interviews software, telephone charges, interviewer wages, data collection etc.)	Very high (e.g. costs for computer assisted staff, interview software, telephone costs, mobile devices, interviewer wages, transport and accommodation costs, data collection etc.)
Support: How much staff do I need to realise the survey? Which facilities are available (software, hardware)?	Some (without software to collect data), little (with software to collect data)	Little	A lot or a great many	A lot or a great many
Sample: How many interviews are to be carried out? How large is the sample?	Average till high/ big	Very high	High	Small till average
Complexity: How complex should the content of the survey be?	Small till average	High till very high	Small till average	High till very high
Length: How long should the questionnaire be? Can we expect the respondents to be able to answer the questions?	Average	Average till long	Average till long	Very long

Question to Ask	Survey Design			
	Mail	Online	Telephone	Face-to-Face
Time: How much time is needed to collect the data?	A lot (without software for data collection), little (with software for data collection)	Very little	Very little	A lot till average
Response rate (see Chapter 6.4): What is the expected response rate?	Average	Low	Low till average	High

Table 16 *Questions for determining the appropriate survey method*

? Questions & Assignments

1. Imagine you wanted to find out which additional services first-year students need during their first semester at your higher education institution. Which study design would you choose and why?

Further Reading

■ Salant, P & Dillman, D. A. (1994). *How to conduct your own survey*. New York: Wiley, 33-52.

■ Trochim, W. M. K. (2005). *Research methods: The concise knowledge base*. Mason, Ohio: Cengage Learning, 94-97.

🎯 **On successful completion of this chapter, you should be able to...**

- name and describe the steps from a concept to its measurement,
- operationalise concepts,
- describe levels of measurement and scale types,
- explain the importance of the quality criteria of social research methods objectivity, validity and reliability,
- describe common types of misuse in evaluation research.

5 Measurement and Operationalisation

5.1 It's a Long Way to Collect Data: From Conceptualisation to Measurement

A survey is often also described as a measurement because it is part of a scientific process to collect data (Trochim 2005, 49). Thinking about measurements in daily life, it is rather easy to do it: Actually, we do it every day, for example, the average rainfall in litre per square meter, the necessary portion of rice that we need for our favourite meal, or the size of our children in centimetres. Such measurements differ from data collection because they are based on a structured design and a structured scale. For example, the corporal size of 1.75 metres can be expressed as 175 centimetres, based on the metrical scale. But how do we proceed, if we want to find out about the teaching quality at our higher education institution?

Measurements in social research

Are we able to measure variables in such an easy way? The answer is: yes, because we can measure everything (Babbie 2004, 119)! This sounds easier than it really is. Very often, we do not know exactly what is to be measured. When talking about "teaching quality", it is an issue that we all link with different perceptions or mental images. Furthermore, we are not able to touch or see this mental image, which means that we cannot directly observe or measure it.

Measurements needs concepts

As already discussed in Chapter 2.8, the fundament of data collection (and with it also of measurements) is a concept. Based on this, the first step of measurement is to develop a mental image (or concept) that is accepted generally. The process of coming to an agreement about what terms mean is conceptualisation, and the result is called a concept (Babbie 2004, 120). The second step is to translate this mental image in such a way that the non-observable concept becomes measurable. For reasons of complexity, this process may involve having to divide our concept into different components or dimensions. As discussed in Chapter 2.8 this "translation" is called operationalisation. During the operationalisation we look for criteria that are observable or measurable, and thereby quantifiable. Such criteria are also known as indicators because they indicate something about a concept and its dimensions, and with it they have a logical link with the concept. Indicators stand for the concept and we use them to tap concepts that are not directly quantifiable.

Three steps from concepts to measurements

To make indicators quantifiable and measurable, we have to convert them in a third step into variables (Bryman 2004, 66 et seqq.; Kumar 2005, 54 et seqq.). Variables can have different values that are connected in a logical way. For example, the variable "gender" can have the values "male" or "female" (see Chapter 5.3).

What are variables?

Figure 7 *Steps from operationalisation to measurement*

The results of these steps are the fundament for conceptualising questions, question guidelines and questionnaires for surveys. Kumar (2005) summarises the importance of a successful operationalisation by recommending the following:

> *"If you are using a concept in your study, make sure you are clear about its indicators and their measurement. (…) Do not use concepts in your research problem that you are not sure how to measure. This does not mean you cannot develop a measurement procedure as the study progress. While most of the developmental work will be done during your study, it is imperative that you are reasonably clear about measurement of these concepts at this stage."*
> **(Kumar 2005, 43)**

To illustrate more clearly the steps of operationalisation, we will proceed with the operationalisation of the concept "student success" in the following Chapter 5.2.

5.2   Example: Operationalisation of the Concept "Student Success"

The Munich Multifactorial Model for Course Quality by German psychologist Heiner Rindermann (see Figure 2) is an attempt to systematise the factors that have an influence on the success – or in other words, on the quality.

Up to now, there is no scientifically-based model for teaching quality. For example, in a model from Rindermann, an indicator of good teaching is when students achieve learning outcomes from a course. Based on this, the success of a course depends on a positive influence on learning outcomes. We will use this influence as a first dimension of our concept.

Dimensions and indicators

In sum, Rindermann assumes that three dimensions contribute to the achievement of learning outcomes and thus to successful teaching and learning: the capabilities and behaviour of the teacher (e.g. speaks clearly and comprehensibly), the capabilities and behaviour of the students (e.g. contribute actively to the discussions in the course) and the framework conditions within which the course takes place (e.g. course has 10 students). These dimensions in turn can be operationalised and transferred to different indicators and thus can be used in a survey (see Chapter 5.1). Therefore, we still have to differentiate the indicators, so that they deliver concrete measurable results. The level of differentiation can refer to the conclusion that a certain aspect exists or does not exist (e.g. question: "Have tools for preparation and revision (such as scripts, slides, e-learning services) been provided? Answer options: "yes" or "no"). It also can refer to determining the degree or extent of a certain aspect (e.g. question: "How much time did you spend on preparation or assignments for the course last week?" Answer option: "Please fill in numbers") (see Chapter 5.3).

From concepts to variables

The following Table 17 gives an overview on how to operationalise the dimensions, defined by Rindermann, and translate them into indicators. Please note that an indicator can be described by different questions. For example, in order to get to know more about the students' commitment in a particular course, you can ask how students' presentations are perceived by their fellows, or how students judge their active participation in the course, or whether group work is perceived to be fruitful etc.

Dimension	Indicator	Example Questions/Statements in a Student Survey/Variables
Teacher's Performance	Structuredness	The learning objectives are made clear. The course requirements are clearly communicated. The content of the individual session is coherent and integrated into a broader context of the subject-matter. ...
	Rhetoric	The lecturer speaks clearly and understandably. The content of the course is presented in an understandable way. Teaching methods and techniques of learning are used appropriately. I can easily keep up with the presentation of course content and the related tasks. ...
	Commitment	The lecturer... ...seems to be well-prepared in the individual sessions. ...shows real interest in students' learning success. ...facilitates students' questions and active participation. ...answers questions or remarks of the students appropriately. ...creates a constructive learning environment. ...is sufficiently available for queries or additional advice (if needed). ...
	...	...
Student	Previous knowledge	The content of the course is related to my previous knowledge. ...
	Students' commitment	I think most of the students in this course... ...attend the individual sessions regularly. ...are prepared adequately for the individual sessions. ...participate actively as far as possible. ...follow the course with interest and attention. ...contribute sufficiently to the course's success. ...
	...	...
Framework Conditions	Room conditions	The number of students in relation to the room size is acceptable. The equipment of the learning area (furniture, media technology, design of a learning platform as part of e-learning etc.) seems to be useful.
	Requirements of the course	I can fulfil the requirements of the course (preparation and post-processing, active participation). Generally, I can fulfil the required preparation and assignments for the course within my time schedule.
	...	...

Dimension	Indicator	Example Questions/Statements in a Student Survey/Variables
Learning outcomes	Competence development	The course helps me to... ...present basic facts and concepts in the subject area of the course. ...be able to work on typical questions or problems of the subject area of the course. ...critically discuss limits and possibilities of the subject area. ...improve my learning methods. ...improve my competency of self-regulated working/studying. ...raise my interest in the subject area

The course content is related to the occupational field and the application of knowledge. The learning arrangement enables me to deepen my understanding of the course content individually. |
| | ... | ... |

Table 17 *Operationalisation of course quality (adapted from Rindermann 1998)*

? Questions & Assignments

1. Why do we need concepts for measurements?
2. Please describe the process of operationalisation.
3. Please develop a mind for a concept, possible dimensions and indicators.

Further Reading

- Babbie, E. R. (2004). *The practice of social research* (10th edition). Southbank: Wadsworth, 119-140.
- Bryman, A. (2004). *Social research methods* (2nd edition). Oxford: University Press, 65-69.

5.3   Levels of Measurement and Scale Types

It can be rather challenging to define indicators or questions that are able to measure our concepts. The way to formulate questions determines the type of variables and how we classify our measurement with a certain **measurement scale**.

We can classify different types of variables for our measurement: Variables can be **constant** and only define one **value** or one **category** (e.g. a student). They can define at most two values (e.g. yes/no). In this case, they are **dichotomous**. Or they can describe more than two categories, then being **polytomous** (e.g. Christian, Muslim, Hindu, and Jew). These **categorical variables** differ from continuous or also known as **metric variables**, such as income or age. Metric variables can refer to any value on the measurement scale (e.g. income: dollars and cents; age: years, months and day). Furthermore, variables can be qualitative or quantitative.

<div style="text-align: right">Types of variables</div>

| | Categorical | | | | |
Constant	Dichotomous	Polytomous	Continuous	Qualitative	Quantitative
Student Lecturer	Yes/no Male/female	*Attitudes* Strongly favourable Favourable Uncertain Strongly unfavourable	*Income* ($) Age (years)	*Gender* Male female *Age* Old Young Child	*Income* $ per year *Age* Years/months

Table 18 *Categorising variables (Kumar 2005, 66) (own table)*

Distinguishing quantitative and qualitative variables means determining whether a variable measures qualitatively or quantitatively a certain characteristic. In doing so, we define the measurement scale of a variable. This differentiation is important because it influences the later data analysis, interpretation and reporting of the results (see Chapter 7).

Basically, we can differentiate four types of measurement scales: Nominal (or classificatory) scale, ordinal (or ranking) scale, interval scale and ratio scale. We have a nominal scale when we have to decide between equality and inequality. Nominal scales name, classify, or number. But the values of the variables cannot be graded according to their size (e.g. gender). Each subgroup has a special characteristic that is common to all within that subgroup (e.g. male). An ordinal scale has the characteristics of a nominal scale. In addition, subgroups have a relationship to one another. They can be arranged in ascending or descending order. Thereby, we do not know the intervals between the individual values (e.g. socio-economic status: upper/middle/low). If we are able to define the intervals between the values of a variable, we call it interval scale (e.g. temperature). It has all the characteristics of a nominal and ordinal scale. In addition, it has a unit of measurement with an arbitrary starting and terminating point. Based on this, we can also calculate average values (see Chapter 7.3). A ratio scale includes all characteristics of an interval scale. In addition, it has a fixed starting point (e.g. income $) (Kumar 2005, 69).

<div style="text-align: right">Types of measurement scales</div>

Scaling techniques and items

In surveys within evaluation processes we often use scaling techniques to measure attitudes or beliefs. Such scaling techniques are able to summarise several items that are connected based on a logical or empirical structure. Items are the smallest elements of a research tool, such as questions or statements to which the respondents can indicate the individual level of agreement or disagreement. For example, instead of asking just one question on competence development, we can ask a series of questions, as shown in the following table:

Please indicate the extent to which you agree or disagree with each statement (please mark your answer).					
Items/Statements	*strongly agree*	*mildly agree*	*undecided or unsure*	*mildly disagree*	*strongly disagree*
...present basic facts and concepts in the subject area of the course.	◎	◎	◎	◎	◎
...be able to work on typical questions or problems of the subject area of the course.	◎	◎	◎	◎	◎
...critically discuss limits and possibilities of the subject area.	◎	◎	◎	◎	◎
...improve my methods.	◎	◎	◎	◎	◎
...improve my competency of self-regulated working/studying.	◎	◎	◎	◎	◎
...raise my interest in the subject area	◎	◎	◎	◎	◎
...raise my interest in the subject area	◎	◎	◎	◎	◎

Table 19 *Using a series of questions to examine a single topic*

The advantage of using such a series of questions is that complex issues are not oversimplified. Thus, the scaling technique helps us to summarise answers on the whole series into one indicator (Salant & Dillman 1994, 87; Babbie 2004, 150 et seqq.).

? Questions & Assignments

1. Why is it important to be able to distinguish between the types of variables?
2. Make sure you are familiar with the differences between the types of measurement scales outlined.
3. Please describe the purpose of items with regard to scaling techniques.

👀 Further Reading

■ Babbie, E. R. (2004). *The practice of social research* (10th edition). Southbank: Wadsworth, 132-177.

5.4   Criteria in Social Research

We have already learned a lot about methods of data collection and measurement. We know that the process of data collection includes different phases and steps (see Chapter 4.1) that have effects on the accuracy and quality of our conclusions. Fortunately, there are **quality criteria** that help to judge quality of social research. Such criteria are the basis to be able to trust in collected data and use it for our work. The three most important criteria are **objectivity, reliability and validity**, and they will be explained in the following.

Main criteria in social research: Objectivity, reliability and validity

We achieve objectivity, when the results of the collected data are independent from the person in charge. That means, a standardised questionnaire should not give the interviewers room to formulate the questions in different individual ways. Furthermore, data analysis should be independent from the persons who do the analysis. That is why we use established statistical procedures that enable a verification of the analysis. Finally, the interpretation of the results should be independent from the interpreting persons. That means that the results of the interpretation should be based on a unitary standard (e.g. a norm).

Objectivity

Reliability is defined as formal accuracy of our results. In other words, the results should be free of mistakes (see Chapter 4.5), so that we achieve the same measurement result when repeating a measurement under the same conditions.

Reliability

The **validity** of a measurement is given when the results of a measurement reflect what should be measured. For example, a survey on "job-success" would not be valid, if it only focusses on study success instead of job-success. In this case, not all dimensions of our concept would have been considered in the measurement.

Validity

A valid measurement also has to be objective and reliable. That means, our operationalisations, measurements and conclusions are only reliable if we use the quality criteria as standard for our own work. We should do that for our own interest.

💬 Questions & Assignments

1. Please give an example of a measurement that is not objective.
2. Please give an example of a measurement that is not reliable.
3. Please give an example of a measurement that is not valid.

5.5 Common Types of Misuse in Evaluation Research

Avoid
misuse
to be
credible

We have now learned the most important quality criteria of empirical social research and are aware of their relevance. Experience shows that especially within evaluation processes methodological principles and standards of empirical social research and with it the mentioned quality criteria cannot be accomplished. For example, participants of an evaluation are not able to work independently because of missing resources (see Chapter 3.1). Even though there are challenges like this, the following types of misuse in evaluation research should always be avoided.

Type of Misuse	Reason/Example
Asking "wrong" research questions	Summative or formative evaluation processes are based on quality criteria or hypotheses that do not fit to the purpose of the research design
Requesting an evaluation study after a decision on a programme has been made	Using the study only as a reason to delay or justify the decision already made
Demanding the use of a research design/data collection technique that does not fit to the purpose of the programme evaluation	Results may be invalid, unethical and not useful; the evaluation would be useless and a waste of time and resources
Influencing the research design or data collection process to ensure that it produces desired results	Results may be invalid, unethical and not useful; evaluation would be useless and a waste of time and resources
Continuing a programme even if the evaluation results unambiguously show it to be ineffective; or closing a programme even if the results unambiguously show it to be highly effective	Evaluation would be useless and a waste of time and resources
Deleting positive results that oppose closing a programme; or deleting negative results that oppose continuing/expanding a programme	Results may be invalid, corrupted, unethical and not useful; evaluation would be useless and a waste of time and resources

Table 20 *Types of misuse in evaluation research (Neuman 2000, 105) (own table)*

The types of misuse mentioned in the table above refer to the abuse of methods and techniques in empirical social research that is done on purpose. It contradicts the paradigm of "good scientific practice" and should be neglected.

Further Reading

- Babbie, E. R. (2004). *The practice of social research* (10th edition). Southbank: Wadsworth, 61-82.
- Neuman, W. L. (2000). *Social research methods: Qualitative and quantitative approaches* (4th edition). Boston: Allyn and Bacon, 89-119.

Chapter 6

Formulating Questions and Survey Design

On successful completion of this chapter, you should be able to...

- name and distinguish several types and functions of questions,
- formulate appropriate questions,
- apply the rules of questionnaire design to your own survey projects,
- implement measures to obtain a high response rate when conducting your own surveys.

6 Formulating Questions and Survey Design

6.1 Functions and Types of Questions

The selection and formulation of adequate types of questions for our research design are very important because they highly influence the quality of collected data. As we have learned in Chapter 5.2, dimensions can be operationalised and transferred to different indicators and thus can be used in a survey. In the following, we will get to know how to translate indicators into good questions. First of all, we have to be aware that questions can be used for different purposes: Either we want to find out about facts and knowledge of the respondents, or we ask respondents about their behaviour and opinions or other interesting aspects.

We can distinguish questions according to their function in a questionnaire: Starting or contact questions are used to introduce a survey and to create a comfortable survey situation. They are the key to making respondents participate in a survey. If the first questions of a survey are boring or misunderstandable, many respondents will already drop the survey. Questionnaires also become boring for respondents if they have to give answers to issues that are not relevant to them. For example, in a survey on study services, students cannot comment on services which they have not used yet. That is why we create so called filter questions that guide respondents through a survey and only lead to the questions that are relevant to them. Sometimes, questions can also have a control function. Such control questions are used to check if the given answers on the questions are consistent. Therefore, we ask two questions on the same issue (if possible, with a time interval). Only if both questions are answered in the same way, has the respondent answered consistently. Such control questions have to be used carefully because they can also include negative effects. For example, a respondent might feel they are not taken seriously.

Functions of questions

In the following, we will get to know different types of questions as they are generally used in surveys (Neuman 2000, 261; Salant & Dillman 1994, 77 et seqq.).

Open-ended questions do not offer fixed answers but the respondents can answer the questions in their own words. Respondents are not forced to answer in the same way, compared to questions that already include response choices. Such open answers enable responses that the researcher might not have considered yet. Open-ended questions are useful for exploring new areas or areas in which the researcher has limited knowledge. They can be used to generate fixed-choice format answers. Open-questions are time-consuming for interviewers to be administered, and they have to be coded. They require a greater effort from respondents.

Types of questions

Example (extracted from a questionnaire on course evaluation)

What aspect of the course did you like most so far? (please specify) _____

Closed-ended questions offer the respondent different pre-formulated answers. When developing such closed-ended questions, the proposed responses have to consider all possible answers as response items. We differentiate between so-called single-choice questions that only offer one response, and so-called multiple-choice questions that offer various response possibilities. If several responses are possible, this should be indicated in the question. We also talk about closed-ended questions, if we want to achieve an opinion on a certain issue by asking a series of questions (see Chapter 5.3, Table 19).

Examples (extracted from a survey on course evaluation)

Single choice

Within what study programme do you attend this course?

◎ B.Sc. Chemistry
◎ M.Sc. Chemistry
◎ B.Sc. Water Science
◎ M.Sc. Water Science

Multiple choice

What are your reasons for taking this course? (Multiple answers possible)
☐ time of the course
☐ special interest in the topic
☐ the lecturer
☐ course is relevant for future employment
☐ research orientation of the course
☐ compulsory course
☐ elective/optional course
☐ preparation for exam
☐ course topic is relevant for final thesis
☐ other reasons

Furthermore, we can also pose **partially closed-ended questions**. These questions are a mix of open and closed questions. Besides pre-formulated answers, respondents have the possibility to answer in their own words. Usually the use of open-ended questions should be avoided when it is possible to formulate a closed-ended question.

Examples (extracted from a questionnaire on course evaluation)

What are your reasons for taking this course? (Multiple answers possible)
- ☐ time of the course
- ☐ special interest in the topic
- ☐ the lecturer
- ☐ course is relevant for future employment
- ☐ research orientation of the course
- ☐ compulsory course
- ☐ elective/optional course
- ☐ preparation for exam
- ☐ course topic is relevant for final thesis

- ☐ other reasons (please specify) _____

Due to the pre-formulated responses, (partially) closed-ended questions mainly deliver quantitative data and are used in structured surveys (see Chapter 4.3). Compared to open-ended questions, advantages refer to better comparability of the responses, and a higher level of objectivity with regard to the realisation and analysis (see Chapter 5.4). Furthermore, it is less effort for respondents to complete the survey and to analyse the collected data (see Chapter 3.3). Nevertheless, closed-ended questions also have some disadvantages: They can suggest ideas that respondents would not have on their own. Respondents with no opinion or knowledge on a certain issue can answer anyway or respondents can be frustrated because their desired answer is not given in the selection options. Furthermore, especially in case of multiple response items, the order of these response items can influence the answer of a respondent.

Types of questions: Assets and drawbacks

❓ Questions & Assignments

1. What difficulties do we have to consider when using open-ended questions in surveys?
2. What are the limits of closed-ended questions?
3. What types of questions are usually used in structured interviews and why?

6.2   Formulating Good Questions

If the questions in a questionnaire are poorly formulated, this may influence the quality of answers given and hence the usefulness of the survey. Experience has shown: Respondents usually answer any question, even if it is very poorly formulated. As researchers, we sometimes only realise that a question was poorly formulated when we see the answers. We become aware of this problem, for example, when many respondents do not answer a specific question or if the answers of open questions do not refer to the question.

It is easy to formulate bad questions

It is rather easy to formulate a bad question, as is shown in the following example from the higher education context. Let us assume that we want to find out more about students' mobility abroad in reference to their studies. Therefore, we have formulated the question: "Did you stay abroad for a certain period while doing your degree?" At first glance, this question seems to be acceptable. But, is it really a good question to get information with regard to our research field on students' mobility during their degree? – The answer is no for at least two reasons:

First, it might be unclear to respondents what exactly the question refers to: Is it vacation time, a language course, a compulsory period abroad as part of the curriculum, or a holiday to visit family abroad? Basically, it could be anything, and it is the respondent who decides how to interpret the question. Secondly, respondents might be unsure about how much time is meant with "a certain period". Is it one day? More than three days? More than a week? Also in this case, it is the respondent who defines the time frame for being abroad. Based on this, it can be that some respondents do not answer the way we have intended. For example, they might have done a three days language course abroad, but they do not mention it in the questionnaire because they think it is too short to be mentioned. Since a respondent is always only as good as a question, we should consider the following fundamental rule:

> Questions always have to be understandable and respondable in the way we intended them to be understood and responded.

Ten golden rules to formulate good questions

In addition to this rule, the most important recommendations can be summarised as follows (Bryman 2004, 152 et seqq.):

"Ten Golden Rules" to Formulate Questions	Example and Problem of Formulation	Possible Revision
1. Use simple, unambiguous formulations which can be understood by all survey participants in the same way.	"If you weren't interested in ET, would you eschew that course?" *Use of abbreviations (ET) & uncommon expressions (eschew)*	"If you weren't interested in Ed Tech, would you avoid that course?"

"Ten Golden Rules" to Formulate Questions	Example and Problem of Formulation	Possible Revision
2. Avoid hypothetical questions	"Imagine you had a 16-year-old son, who wanted to quit his studies to become a soccer professional. Would you support him?" *Probably far from the respondent's reality and hard to imagine.*	Ask questions that are directly linked to what you want to know.
3. Avoid long and complex questions.	"Many people think that, in this day and age, students have too many other obligations at their university or they have time-consuming hobbies. To what extent do you agree?" *Too long and lots of unnecessary information*	"Today, students have too many other obligations (job, committee activities, hobbies) besides their studies. To what extent do you agree?"
4. Avoid double-barrelled questions and negative formulations.	"Please indicate how much you agree with the following statement: I do not feel welcomed by my boss and colleagues." *Negative formulation (I do not feel welcomed) and double-barrelled (boss and colleagues)*	"Please indicate how much you agree with the following statements: I feel welcomed by my boss. I feel welcomed by my colleagues."
5. Avoid allegations and suggestive questions.	"Now that you have experienced the benefits of taking additional courses to improve key competences ("soft skills"), would you choose such a course again next semester?" *Allegation/suggestive (you have experienced the benefits)*	"If you took an additional course to improve key competences ("soft skills") last semester, how helpful was it?" Remember the possibility that no course may have been taken. Include an answer-option like "not applicable".

"Ten Golden Rules" to Formulate Questions	Example and Problem of Formulation	Possible Revision
6. Avoid needless questions.	"Please indicate how much you agree with the following statement: Self-employed people work very hard." *Interesting to know, but what information do you get from this answer?*	Always double-check, whether a question is connected to your survey project.
7. Use questions with a clear-cut (timely) reference.	"How do you rate the counselling at your university in the past?" *No clear-cut reference (counselling by teachers, career service etc.?) and no clear-cut time reference (last 3 weeks, last 3 years?)*	How do you rate the counselling by the central student guidance concerning your course options in the last semester (WS 14/15)?
8. Use response categories, which are complete and disjunctive.	*What was your age at the time of graduation?* ◎ 22 – 24 years ◎ 24 – 26 years ◎ 26 – 28 years *Not disjunctive (what about people aged 24 or 26?) and not complete (what about people older than 28?)*	What was your age at the time of graduation? ◎ 22 – 24 years ◎ 25 – 27 years ◎ 27 years or older
9. Ensure that the context of a question has no influence on the response behaviour.	A survey among renowned scientists has shown that 90% of them don't approve of the "Bologna-process". What is your opinion on this? *Context influence (a survey among renowned scientists has shown…)*	How do you rate the "Bologna-process" generally?

"Ten Golden Rules" to Formulate Questions	Example and Problem of Formulation	Possible Revision
10. Define unclear expressions.	What is your average monthly net household income? *Not everybody knows the definition of net household income.*	What is your average monthly net household income (i.e. the combined income of all members of your household after taxes)?

Table 21 *"Ten golden rules" to formulate questions (Bryman 2004, 152 et seqq.)*

We should consider that these rules are not exceptional, and sometimes they oppose each other. That is why they should be considered as recommendations that cannot always be applied at the same time.

No rule without exception

? Questions & Assignments

1. Please give a negative example for each of the ten rules. In the following, please formulate the example according to the rule.

6.3   Questionnaire Design

The questionnaire is an important tool for data collection in empirical social research. The design of a questionnaire is more than simply formulating good questions (see Chapter 6.2). Depending on mail, online or telephone surveys, we also have to consider aspects such as designing pages, format and printing, graphic design and representations (e.g. on mobile devices) (Salant & Dillman 1994, 101 et seqq.). When designing a questionnaire, first of all, we should consider the following golden rule:

"Do unto your respondents as you would have them do unto you!" (Trochim 2005, 86)

Golden Rule of questionnaire design

Finally, we are imposing on the life of our respondents and we are asking for their time, their attention, their trust and for their personal information. So keep the following in mind:

The **Golden Rule of questionnaire design** in practical terms (Trochim 2005, 86)

- Thank the respondent at the beginning for allowing you to conduct your study - and at the end for participating.
- Keep your survey as short as possible! Only include what is absolutely necessary.
- Be sensitive to the needs of the respondent and be aware of any sign that the respondent is uncomfortable.

Sequencing of questions

Having formulated questions and thinking about how to structure them in the questionnaire, we know: Surveys undergo a dramaturgy that is based on the sequencing of the individual questions in the questionnaire (see Chapter 6.1). If we choose a clever dramaturgy, this can have a positive effect on the survey. By contrast, a survey becomes less successful, if there is an inadequate order of questions. For example, responses to questions can be influenced by previous questions. Questions can be posed too late or too early to gain interest. Or a question does not receive sufficient awareness due to the surrounding questions. In view of these aspects, we should consider the following recommendations that are independent from the respective survey design (Trochim 2005, 86):

- Start with easy, non-threatening questions and put more difficult, threatening and sociographic questions at the end.
- Do not start mail- or online-surveys with an open-ended question.
- Follow in chronological order.
- Ask about one topic at a time and use a transition when switching topics.
- If you want to ask respondents about difficult or uncomfortable subjects, precede such sensitive questions with some easier warm-up questions.

Pre-test your questionnaire

To test the practicality in terms of content and technique and to optimise the questionnaire, you should always do a **pre-test** after having finished the prototype of the questionnaire, even if you copy most of the questions from other questionnaires. Pre-tests give answers on the time frame of a survey and uncover technical problems (such as filtering mistakes) or unclear questions. The easiest way is to ask people from your personal surroundings to complete the survey and to ask them afterwards about their experiences, possible difficulties or problems of understanding when doing the survey (Babbie 2004, 256).

Train your interviewers

Interviewers (see Chapter 4.3) who use a questionnaire in an interview have a particular task within the survey process: They have to motivate the respondents to participate in a survey (until the end), and at the same time they have to keep neutrality on the content of the survey. This neutrality is important to fulfil the required level of standardisation. On the administrative level they have to organise the interviews and consider the respective requirements of the sample. Therefore, it is useful to train the interviewers in advance of a study. Only carefully trained interviewers are able to realise a survey technically and methodologically correctly (Babbie 2004, 263ff). Furthermore, we need a very accurate survey administration that defines realistic requirements, such as the number of interviews and the respective time needed.

> ## ⁇ **Questions & Assignments**
>
> 1. Please explain the golden rule of a questionnaire.
> 2. Please describe a successful sequencing of questions.
> 3. Why should a questionnaire be based on a pre-test?

6.4   Measures for Obtaining a High Response Rate

Another key problem of surveys is low response rates. The response rate (also called completion or returned rate) is the number of people in percentage, participating in a survey, divided by the number selected in the sample (Babbie 2004, 261). Reasonable minimum values for a response rate do not exist because it is strongly determined by the method. But it can be stated: The lower the response rate, the higher the probability that the results are based on self-selection of the respondents or item-non-response (see Chapter 4.5).

For example, a higher education institution conducts a survey among its employees about the issue of "leadership". Only very few of the secretaries participate because they are frightened of sanctions. Such behaviour can falsify the whole survey. If the response rate is not 100%, there is always the risk of sampling bias (see Chapter 4.4), even though the response rate is higher than 50%. Higher response rates reduce the probability of possible sources of mistakes, and they strengthen the possibilities of statistical data analysis. That is why we should always try to achieve the highest response rate possible.

High response rates welcome

To do so, first of all we should consider the recommendations on formulating questions and questionnaire design (see Chapter 6.2 and 6.3). Based on the total-design-method (Dillman 1983), that achieves response rates that are much higher than 50%, and the respective study design, the following aspects should be considered:

Total-design method

- Design of the questionnaire
  - Paper-based surveys: White paper and a brochure format
  - Online surveys: A questionnaire should be able to be illustrated and completed by any browser; this also includes browsers on mobile devices such as smartphones or tablets.
- Invitation letter
  - In general: Clarify the benefits of the study, and underline the importance of the respondents and their participation. Explain, why the respondents have been chosen, and ensure the protection of the given data. Name a concrete deadline for the response and motivate for asking questions, if necessary.
  - Paper-based surveys: Only use official material (e.g. paper with corporate design of the institution); if applicable: correctly stamped and including a stamped self-addressed envelope; sign (digitally) personally.
  - Online surveys via email: The invitation has to be readable with any email-programme. A trustful sender address has to be used. If needed, the paper-based invitation can be attached.
- Distribution
  - In general: Having sent the invitation you can send three more letters as a reminder of the survey. The time schedule for sending these letters is designed based on the survey time-frame.
  - Paper-based surveys: Send the invitation in such a way that it arrives at midweek. After one week, you can send a postcard to thank the respondents for the participation/to remind them in a friendly way to participate. After three weeks you can resend a short letter including the survey again.
  - Online surveys: Reminder-emails can be sent respectively. If you also have postal addresses, you can also send a letter that informs about the online survey. If you have telephone numbers, you can also remind addressees by telephone instead of a letter or email.

■ Incentives

■ In general: You can offer money or other rewards as an incentive to participate in a survey. Using incentives should be reflected critically, because they can also lead to distortions. That is why incentives should not be the key motivation to participate in a survey, and it should be the same for all participants.

⟨?⟩ Questions & Assignments

1. Why can low response rates be problematic, and why are high response rates less problematic?
2. Please describe how you would design a mail survey to achieve a maximum response rate.

👓 Further Reading

■ Babbie, E. R. (2004). *The practice of social research* (10th edition). Southbank: Wadsworth, 242-279.
■ Bryman, A. (2004). *Social research methods* (2nd edition). Oxford: University Press, 144-162.
■ Couper, M. (2008). *Designing effective web surveys*. New York: Cambridge University Press.
■ Dillman, D. A. (1983). Mail and other self-administered questionnaires. In P. H. Rossi (Ed.), *Quantitative studies in social relations. Handbook of survey research* (pp. 359-378). Orlando: Academic Press.
■ Groves, R. M., Fowler, F. J., Couper, M., Lepkowski, J. M., Singer, E. & Tourangeau, R. (2009). *Survey methodology* (2nd edition). Hoboken, NJ: Wiley, 217-258.
■ Salant, P & Dillman, D. A. (1994). *How to conduct your own survey*. New York: Wiley, 77-136.

Chapter 7

Processing and Analysis of Data: From Questionnaires to Survey Results

On successful completion of this chapter, you should be able to...

- have a basic understanding of the process of data clean-up and coding, according to general rules,
- explain functions and name contents of field reports,
- interpret the presented measures of central tendency,
- have a basic understanding of preparing reports of survey results.

7 Processing and Analysis of Data: From Questionnaires to Survey Results

7.1 Data Clean-up and Coding

Having finished our survey (and with it the process of gathering data), the data does not yet exist in a form which can be analysed immediately. Usually, data based on online or phone surveys is saved in such a way that it can be extracted immediately. Nevertheless, even these datasets still have to be examined and checked for mistakes.

If, after an email survey, a data set does not yet exist, it has to be created. Usually, we use special software that allows quantitative data to be managed and analysed and, for example, after an email survey, the acquired data has to be transferred into a dataset. If this is not done based on a software programme that was used for data collection (e.g. within an online-survey), it has to be done by hand. Therefore, we code all questions of a questionnaire and every possible response of (particularly) closed-ended questions with a code, to be able to analyse the data.

Creating a data set

The following data clean-up is the fundament for our data analysis. Any data collection, be it on an automatic or manual basis, can include mistakes. The most common mistakes are wild codes and outliers. Wild codes are values that are outside the defined frame. Let us assume a variable with values between 1 and 5. That means, 1, 2, 3, 4, and 5 are **valid values**. If the data matrix also includes the value 7, this is a wild code (perhaps just being a literal mistake). Outliers are values that are not plausible. For example, a variable on semesters includes a value of 85. This value would not be plausible and should be rejected from the collected data. Existing wild codes, outliers or non-responses (see Chapter 4.5) are non-valid values that are coded with an own code (e.g. 8 = not applicable and 9 = no answer). They are also called **missing values**.[4]

Valid and missing values

Sometimes, it can occur that data in a survey is collected in a more differentiated way than actually needed. As we have already learned, we can summarise different variables to a new variable. In doing so, we aggregate the data (see Chapter 4.3). To give an example, we might summarise the values of a metrical variable (see Chapter 5.3): To be able to illustrate the responses on income graphically, it makes sense to categorise these responses in a new variable, and to recode them. In this case, the variable income that consists of individual numerical values could be transferred into a new recoded variable that summarises the numerical values in different scopes. The new variable might have the following values: 1 (= no income), 2 (= income of 0-250), 3 (= 751$ - 1000$); 6 (= 1001$ and above); 9 (= no answer). Metrical variables should always be categorised like this, because they offer other possibilities of analysis than categorised variables (see Chapter 7.3).

Quality control – check your data

---

4 For more detailed information on post-collection processing of survey data refer to Groves et al. 2009, 329 et seqq.

We already got to know an example of coding qualitative data in Chapter 3.3. A special case is the processing of data in unstructured studies that only collect qualitative data. For complexity reasons, we will not discuss this special case in this course book.[5]

7.2   Codebook and Field Report

The code lists should be included in a **codebook**, or a reference has to be made. This is important if the data entry will be done by someone who is not familiar with the project. There are no formal requirements except that anybody who reads the codebook should be able to understand how the information given in the questionnaire should be treated as data.

Let others understand what you did

A **field report** includes reasons for a survey, the development (and perhaps the theoretical background) of the questions and the questionnaire, the methodological design, the population or sample, as well as the operative and administrative data collection process (e.g. dates to send invitations and reminders). Based on this information, those who order a survey or others who are involved in an evaluation process can get an overview about the realisation and quality of the study. Therefore, a methodological field report should include the following information:

- Stakeholders who commission a study, their reasons, and purposes of the study,
- Design and pre-test of the questions and the questionnaire,
- Target group and dataset-design (if necessary),
- If applicable: measures to train interviewers,
- Response rate, including exact dates for invitations, reminders etc.
- Average time schedule for interviews
- If applicable: Particularities of the survey, e.g. problems and limitations

The field report is part of the survey results reporting (see Chapter 7.3).

7.3   Reporting Survey Results

Having collected and processed the data, the more interesting part of data analysis follows, this usually being a report of the results. Therefore, data is usually processed in such a way that it is aggregated by using tables, graphics or so-called **measured values**. In doing so, results can be illustrated in such a way that they become clear and easily readable for addressees. That is why we preferably use methods of the so-called descriptive statistics. In the following, you will get a broad overview about these statistics and measured values that can often be found in evaluation reports.

---

5 For detailed information about qualitative data analysis (QDA) and learning materials on how to use Computer Assisted Qualitative Data AnalysiS (CAQDAS) packages, please confer e.g. the following website http://onlineqda.hud.ac.uk/index.php

Kumar (2005), 248 et seqq., and Salant & Dillman (1994), 206 et seqq., give helpful examples to illustrate data in tables and graphs.

We can create a **frequency distribution** from coded or quantitative variables. We receive so called absolute distributions and the respective percentage ratio. Creating frequency distributions is recommendable for any variable because this also helps to find mistakes such as wild codes or outliers during data processing (see Chapter 7.1).

Measures of central tendency are used to aggregate gathered information in a number. They can be calculated easily by most analytics software. Such measures give information about the most frequent value (mode), about the value that is in the middle of a distribution (median), and about the average value (mean).

Does correlation indicate cause and effects?

We can already use the **mode** for nominal scale variables. It describes the measured value that exists most often. To calculate a **median**, we need at least ordinal variables. It describes the measured value in the middle, if all values are ordered from the lowest to the highest value. That means 50% of the values are below, and 50% are above this value. The mean should only be calculated for variables of minimum interval scale. Often, it is also used for ordinal scales (see Chapter 5.3). It describes the average value and is calculated by summarising all measured values, divided through the number of responses. There are disadvantages with regard to the extreme values that differ a lot from the other measured values. The so-called **standard deviation** measures how values are spread around a mean. Or in other words, whether they cluster together or are widely dispersed. It is the square root of the sum of the squared deviations from the mean divided by one less than the sample size.

Very often, we want to describe more than only average values or distributions. We are also interested in statistical correlations between variables. If there is, we say that variables **correlate** with each other. For example, a **positive correlation** means that "the more of variable A…, the more of variable B." Looking at the variables "final degree" and "income", a positive correlation would be that the better the final degree, the higher the income. Correlations are indicators, but they are no proof of **causalities**, meaning proofed correlations between cause and effects. To illustrate the common frequency distribution of two variables with the purpose to find correlations, we can create **cross-tabulations** (cross-tabs). For example, when examining the hypotheses "the better the final degree, the higher the income", we check if there is a correlation between the variables "final degree" and "income". In this case, the final degree is the so-called independent variable, because we assume that it influences the income. When doing cross-taps, we have to consider the formal convention that the **independent variables** are illustrated in the column, and the dependent variables in the lines. The results of a cross-tap analysis can be summarised by three ways:

1) absolute numbers
2) column percentages (percentages in each column add up to 100)
3) line percentages (percentages in each line add up to 100)

Computer programmes designed for statistical analysis also routinely calculate a statistic called chi-square. This value explains if two variables correlate. At the same time, we check if the correlation is significant. How-

ever, it is important to consider that the significance does not refer to any possible causal correlation. Therefore, we need even more detailed analysis.

Having concluded our data processing and designed a report we can use a structure according to the following chapters (Salant & Dillman 1994, 203):

- Abstract or executive summary (includes the most important findings)
- Problem statement (explains why the survey was done)
- Field report (see Chapter 7.2)
- Findings (present results that really matter in logical order)
- Implications (draw findings together to answer original questions and explores implications for decision making)
- Appendices (provide supplementary material like questionnaires etc.)

To describe our findings, we can refer to our calculated measures. Furthermore, we can create graphics based on the frequency distributions that are easier to read than tables.

Functions of the report

Based on this, our report follows three functions: It describes and documents our processing with regard to data collection and data analysis, and our results. Such data can be used for different evaluation processes because due to the scientific method it is of high quality and becomes clear and understandable to anybody.

? Questions & Assignments

1. How do you proceed to examine the quality of your data matrix?
2. What is the difference between valid and invalid values? Please give an example.
3. What are categorised variables? Please give an example.
4. Please name functions and content of field reports.
5. Which measures of central tendency are often reported in result reports? Please explain their explanatory value.
6. Please describe the advantages of cross-tabs.
7. Please name and explain functions and structure of result reports.

Further Reading on the Analysis of Qualitative Data

- Babbie, E. R. (2004). *The practice of social research* (10th edition). Southbank: Wadsworth, 369-393.
- Bryman, A. (2004). *Social research methods* (2nd edition). Oxford: University Press, 398-416.
- Neuman, W. L. (2000). *Social research methods: Qualitative and quantitative approaches* (4th edition). Boston: Allyn and Bacon, 417-442.

Further Reading on the Analysis of Quantitative Data

- Babbie, E. R. (2004). *The practice of social research* (10th edition). Southbank: Wadsworth, 395-420.
- Bryman, A. (2004). *Social research methods* (2nd edition). Oxford: University Press, 218-242.
- Neuman, W. L. (2000). *Social research methods: Qualitative and quantitative approaches* (4th edition). Boston: Allyn and Bacon, 313-343.
- Salant, P & Dillman, D. A. (1994). *How to conduct your own survey*. New York: Wiley, 175-200.
- Trochim, W. M. K. (2005). *Research methods: The concise knowledge base*. Mason, Ohio: Cengage Learning, 204-223.

References

Atteslander, P. (2010). *Methoden der empirischen Sozialforschung* (13th edition). Berlin: Erich Schmidt Verlag.

Babbie, E. R. (2004). *The practice of social research* (10th edition). Southbank: Wadsworth.

Bryman, A. (2004). *Social research methods* (2nd edition). Oxford: University Press.

Couper, M. (2008). *Designing effective web surveys*. New York: Cambridge University Press.

DeGEval – Gesellschaft für Evaluation (2001). *Evaluation standards (DeGEval-Standards)*. Retrieved from http://www.degeval.de/fileadmin/user_upload/Sonstiges/STANDARDS_2008-12_kurz_engl.pdf

Dillman, D. A. (1983). Mail and other self-administered questionnaires. In P. H. Rossi (Ed.), *Quantitative studies in social relations. Handbook of survey research* (359-378). Orlando: Academic Press.

ENQA – European Association for Quality Assurance in Higher Education (2009). *Standards and guidelines for quality assurance in the European higher education area*. Retrieved from http://www.enqa.eu/index.php/home/esg/

Groves, R. M., Fowler, F. J., Couper, M., Lepkowski, J. M., Singer, E. & Tourangeau, R. (2009). *Survey methodology* (2nd edition). Hoboken, NJ: Wiley.

Hussein, A. (2009). The use of triangulation in social sciences research: Can qualitative and quantitative methods be combined? *Journal of Comparative Social Work* 2009/1.

Kumar, R. (2005). *Research methodology: A step-by-step guide for beginners* (2nd edition). London: SAGE.

McMillan, J. H. & Schumacher, S. (2010). *Research in education: Evidence-based inquiry* (7th edition). Boston: Pearson.

Neuman, W. L. (2000). *Social research methods: Qualitative and quantitative approaches* (4th edition). Boston: Allyn and Bacon.

Nickel, S. (2010). *Qualität in Hochschulen – was ist das?* Präsentation im Rahmen des Moduls Qualitäts- und Personalmanagement im Masterstudiengang Hochschul- und Wissenschaftsmanagement. 4 June 2010.

Pan, D. (5th ed). (2008). *Learning to teach, teaching to learn: A handbook for NUS teachers*. Centre for Development of Teaching and Learning, NUS: Singapore.

Rindermann, H. (1998). The Munich multifactoral model of course quality: Design, justification and verification. *Beiträge zur Hochschulforschung 3,* 189-224.

Röbken, H. (2012): *Qualitätsmanagement in Bildungsorganisationen*. Oldenburg: Carl von Ossietzky Universität Oldenburg.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th edition). Thousand Oaks: SAGE.

Salant, P & Dillman, D. A. (1994). *How to conduct your own survey*. New York: Wiley.

Secolsky, C. & Denison, B. (Eds, 2012). *Handbook on measurement, assessment, and evaluation in higher education.* New York: Routledge.

Stockmann, R. (2004). *Was ist eine gute Evaluation?* CEval-Arbeitspapiere 9. Saarbrücken: Centrum für Evaluation.

Stockmann, R. (2006). *Evaluation und Qualitätsentwicklung. Eine Grundlage für wirkungsorientiertes Qualitätsmanagement*. München: Waxmann.

Stockmann, R. & Meyer, W. (2010). *Evaluation: Eine Einführung*. Opladen: Budrich.

Trochim, W. M. K. (2005). *Research methods: The concise knowledge base*. Mason, Ohio: Cengage Learning.

Yarborough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd edition). Thousand Oaks: SAGE.

List of Tables

List of Figures

Annexes

Annex 1 – Example Course Evaluation Report at National University of Singapore

<div align="center">STUDENTS' RATINGS ON TEACHER</div>

Faculty Member: XXXXXXXXXXXXXX

Department: XXXXXXXXXXXXXX   Academic Year: 20XX/20XX

Faculty: XXXXXXXXXXXXXX   Semester: 2

Module: XXXXXXXXXXXXXX

Activity Type: LECTURE

Class Size/Response Size/Response Rate: 9 / 1 / 11.11%

Contact Session/Teaching Hour :   3 / 6

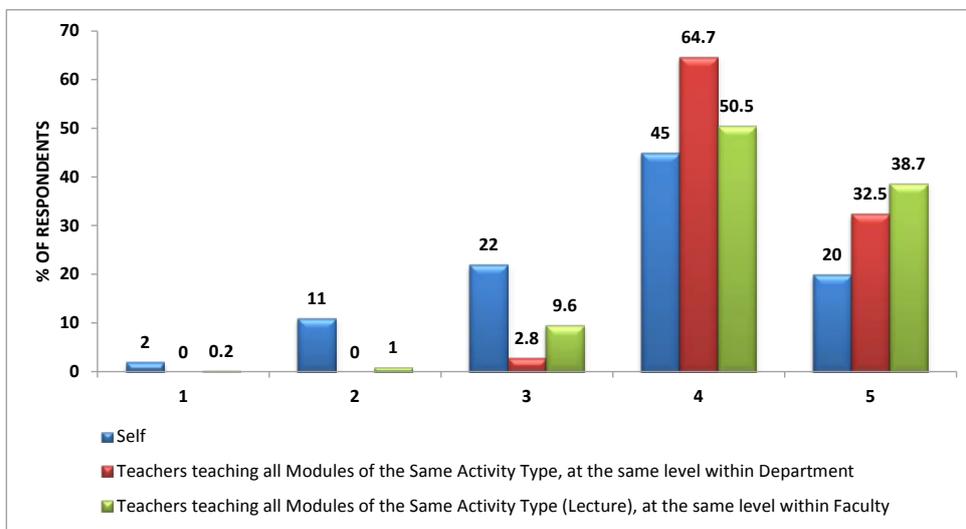| Qn	Items Evaluated	Fac. Member Avg Score	Fac. Member Avg Score Std. Dev	Dept Avg Score (a) (b)	Fac. Avg Score (c) (d)
1	The teacher has enhanced my thinking ability.	4.000	0.000	4.056 (4.225)	3.990 (4.203)
2	The teacher has increased my interest in the subject.	3.000	0.000	4.060 (4.225)	3.963 (4.186)
3	The teacher provides timely and useful feedback.	4.000	0.000	4.092 (4.338)	4.015 (4.264)
4	Department-specific question	4.000	0.000	4.092 (4.338)	
5	Department-specific question	4.000	0.000	4.026 (4.225)	
6	Department-specific question	4.000	0.000	4.053 (4.225)	
	Average of Q1 to Q6	3.833	0.000	4.074 (4.261)	
	Computed Overall Effectiveness of the Teacher.	4.000	0.000	4.132 (4.296)	4.054 (4.265)

Note:

1. A 5-point scale is used for the scores. The higher the score, the better the rating.

2. **Computed Overall Effectiveness of the Teacher** = 0.0866 + (0.4276 * Q1 score) + (0.3150 * Q2 Score) + (0.25 * Q3 Score).

3. **Fac. Member Avg Score:** The mean of all the scores for each question for the faculty member.

4. **Fac. Member Avg Score Std. Dev:** A measure of the range of variability. It measures the extent to which a faculty member's Average Score differs from all the scores in the faculty member's evaluation. The smaller the standard deviation, the greater the robustness of the number given as average.

5. **Dept Avg Score:**
(a) the mean score of same activity type (Lecture) within the department.
(b) the mean score of same activity type (Lecture), at the same module level (level 6000) within the department.

6. **Fac. Avg Score:**
(c) the mean score of same activity type (Lecture) within the faculty.
(d) the mean score of same activity type (Lecture), at the same module level (level 6000) within the faculty.

FREQUENCY DISTRIBUTION OF RESPONSES ON TEACHER

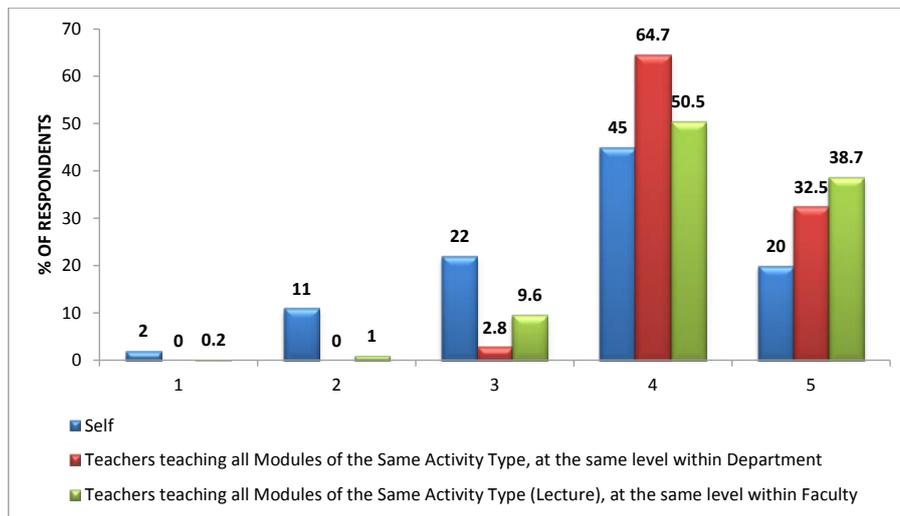| Faculty Member:	XXXXXXXXXXXXXX		
Department:	XXXXXXXXXXXXXX	Academic Year:	20XX/20XX
Faculty:	XXXXXXXXXXXXXX	Semester:	2
Module:	XXXXXXXXXXXXXX		

Frequency Distribution of responses (Qn 1: The teacher has enhanced my thinking ability.)

Nos. of Respondents(% of Respondents)

ITEM\SCORE	5	4	3	2	1
Self	0 (.00%)	1 (100.00%)	0 (.00%)	0 (.00%)	0 (.00%)
Teachers teaching all Modules of the Same Activity Type (Lecture), at the same level within Department	23 (32.39%)	46 (64.79%)	2 (2.82%)	0 (.00%)	0 (.00%)
Teachers teaching all Modules of the Same Activity Type (Lecture), at the same level within Faculty	257 (38.70%)	335 (50.45%)	64 (9.64%)	7 (1.05%)	1 (.15%)

Frequency Distribution of responses (Qn 2: The teacher has increased my interest in the subject.)

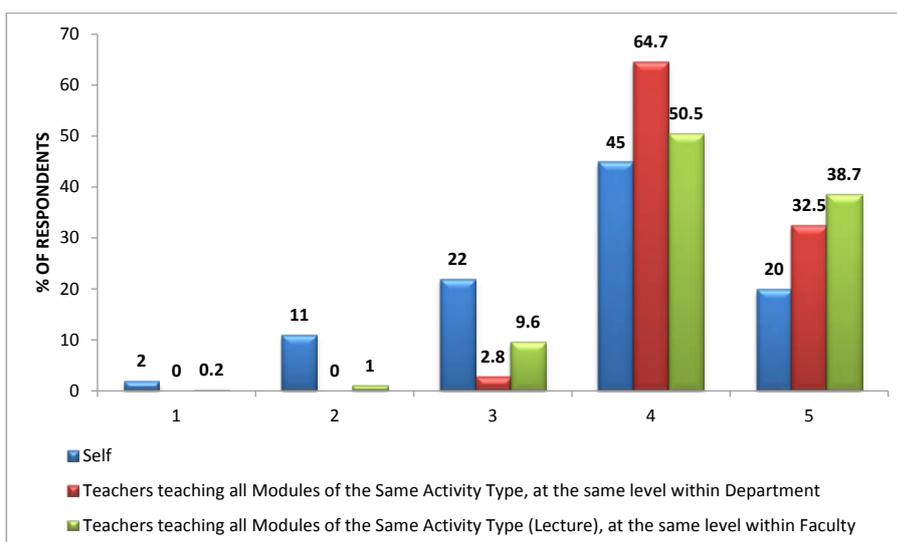|	**Nos. of Respondents(% of Respondents)**				
ITEM\SCORE	**5**	**4**	**3**	**2**	**1**
Self	0 (.00%)	1 (100.00%)	0 (.00%)	0 (.00%)	0 (.00%)
Teachers teaching all Modules of the Same Activity Type (Lecture), at the same level within Department	23 (32.39%)	46 (64.79%)	2 (2.82%)	0 (.00%)	0 (.00%)
Teachers teaching all Modules of the Same Activity Type (Lecture), at the same level within Faculty	257 (38.70%)	335 (50.45%)	64 (9.64%)	7 (1.05%)	1 (.15%)

Frequency Distribution of responses (Qn 3: The teacher provided timely and useful feedback.)

		Nos. of Respondents(% of Respondents)			
ITEM\SCORE	5	4	3	2	1
Self	0 (.00%)	1 (100.00%)	0 (.00%)	0 (.00%)	0 (.00%)
Teachers teaching all Modules of the Same Activity Type (Lecture), at the same level within Department	23 (32.39%)	46 (64.79%)	2 (2.82%)	0 (.00%)	0 (.00%)
Teachers teaching all Modules of the Same Activity Type (Lecture), at the same level within Faculty	257 (38.70%)	335 (50.45%)	64 (9.64%)	7 (1.05%)	1 (.15%)

<div align="center" style="color:green">STUDENTS' COMMENTS ON TEACHER</div>

Faculty Member:	XXXXXXXXXXXXXXX		
Department:	XXXXXXXXXXXXXX	Academic Year:	20XX/20XX
Faculty:	XXXXXXXXXXXXXX	Semester:	2
Module:	XXXXXXXXXXXXXX		
Activity Type:	LECTURE		

What are the teacher's strengths?

Comments from students who gave an average score <u>> 4.5</u> for the overall effectiveness of the teacher
Comments go here…..

Comments from students who gave an average score <u>4.0 - 4.5</u> for the overall effectiveness of the teacher
Comments go here…..

Comments from students who gave an average score <u>3.5 – 4.0</u> for the overall effectiveness of the teacher
Comments go here…..

Comments from students who gave an average score <u>3.0 – 3.5</u> for the overall effectiveness of the teacher
Comments go here…..

Comments from students who gave an average score <u>< 3.0</u> for the overall effectiveness of the teacher
Comments go here…..

What improvements would you suggest to the teacher?

Comments from students who gave an average score <u>< 3.0</u> for the overall effectiveness of the teacher
Comments go here…..

Comments from students who gave an average score <u>3.0 – 3.5</u> for the overall effectiveness of the teacher
Comments go here…..

Comments from students who gave an average score <u>3.5 – 4.0</u> for the overall effectiveness of the teacher
Comments go here…..

Comments from students who gave an average score <u>4.0 - 4.5</u> for the overall effectiveness of the teacher
Comments go here…..

Comments from students who gave an average score <u>> 4.5</u> for the overall effectiveness of the teacher
Comments go here…..

STUDENTS' NOMINATIONS FOR BEST TEACHING

Faculty Member:	XXXXXXXXXXXXXXX		
Department:	XXXXXXXXXXXXXX	Academic Year:	20XX/20XX
Faculty:	XXXXXXXXXXXXXX	Semester:	2
Module Code:	XXXXXXXXXXXXXX	No of Nominations:	30

Comments go here…..