Released 12 December 2016

With the support of the Lifelong Learning Programme of the European Union.

Impact Evaluation of Quality Assurance in Higher Education. A Manual

This publication is meant to be an inspirational manual guide with fundamental methodological issues and clarifications next to implementation of methodology and practical application and also with warnings of possible failures to carry out impact analyses of quality assurance in higher education. This manual guide refers to the European Commission-funded project 'Impact Analysis of External Quality Assurance Processes of Higher Education Institutions' (acronym: IMPALA; http://www.evalag.de/impala and http://www.impala-qa.eu).

This manual is based on the work of the IMPALA partner consortium carried out through the three years of project duration and on results published by the partners and invited experts in seven articles of a special issue of *Quality in Higher Education* (Vol. 21, No. 3) in 2015 as well as in three papers published elsewhere. For further and comprehensive analysis and exploitation of the IMPALA data and embedding of the results into the broader context of the social and political role of higher education in contemporary knowledge societies, a special issue of *European Journal for Higher Education* is planned.

IMPRINT

The IMPALA Consortium Partners, November 2016 –

Agència per a la Qualitat del Sistema Universitari de Catalunya (AQU Catalunya; Barcelona, Spain)
Agenţia Română de Asigurare a Calităţii în Învăţământul Superior (ARACIS; Bucharest, Romania)
European Association for Quality Assurance in Higher Education (ENQA; Brussels, Belgium)
European Students Union (ESU; Brussels, Belgium)
Evaluationsagentur Baden-Württemberg (evalag; Mannheim, Germany)
Finnish Education Evaluation Council (FINEEC; Helsinki, Finland)
Jyväskylä University of Applied Sciences (JAMK; Jyväskylä, Finland)
Technical University of Civil Engineering Bucharest (UTCB; Bucharest, Romania)
Universitat Autònoma de Barcelona (UAB; Barcelona, Spain)
University of Oslo (UiO; Oslo, Norway)
Universität Stuttgart (Uni S; Stuttgart, Germany)

See also http://www.impala-qa.eu/impala/ and https://www.evalag.de/international/wirkungsanalyse/the-project/

Editorial: Theodor Leiber (evalag; Mannheim, Germany)

Impact Evaluation of Quality Assurance in Higher Education. A Manual

Content

Executive Summary

This manual is an outcome of the European Commission-funded project 'Impact Analysis of External Quality Assurance Processes of Higher Education Institutions' (acronym: IMPALA; http://www.evalag.de/impala and http://www.impala-qa.eu). The manual presents and discusses the main theoretical perspectives and general methodological elements of impact evaluation of quality assurance (QA) in higher education institutions (HEIs), which should be a cornerstone of quality development in higher education (HE) and contribute to improving the knowledge about the effectiveness of QA. The suggested core methodology of impact evaluation consists in comparative and longitudinal studies carried out simultaneously to the QA interventions and relying mainly (but not exclusively) on surveys and mixed methods. Normally, the surveys should be carried out by online questionnaires and structured interviews. The methodology is applicable to any specific QA procedure and any HEI achievement area (while, for complexity reasons, any specific impact study must make its choices of focussing on certain QA interventions and its intended as well as non-intended goals).

The methodology presentation is supplemented by an overview of causal designs, particularly causal social mechanisms, for impact analyses of QA in HEIs. Then, characteristics of survey questionnaire construction are given and results of the IMPALA project are presented and tentatively analysed. The analysis focuses on students' attitudes towards QA and teachers' assessment of QA expenditures and benefits. The main discussion ends with a quick SWOT (strenghts; weaknesses; opportunities; threats) analysis of impact evaluation of QA in HEIs and the IMPALA project.

The manual is structured in the following way: The Introduction, Section 1, gives the motivating reasons for impact evaluation of QA in HE and core characteristics of the IMPALA project. Section 2 is about basic causality concepts applicable to HE processes. In Section 3 an overview of the contemporary practice of impact evaluation of QA in HEIs is given by the QA agency and HEI consortium partners from Finland, Germany, Romania and Spain (Section 3.1), and general methodological principles, challenges and basic methodological designs of impact evaluation in complex and dynamic social systems are presented (Section 3.2). Here, a special focus is given to the opportunities but also threats of causal social mechanisms. Section 4 decribes basic characteristics of survey questionnaire construction, and Section 5 presents some (still preliminary) empirical results of the IMPALA project. In Section 6 a quick SWOT analysis of impact evaluation of QA in HEIs is carried out. The manual closes with some prospective conclusions for the practice of QA in HEIs and QA agencies.

1 Introduction

1.1 Why Impact Knowledge of Quality Assurance is Needed

During the last two decades a large variety of procedures and instruments of quality assurance (QA)[1] in higher education (HE) was developed.[2] In parallel, an intensive discourse about HE quality including the opportunities and threats of sustaining and developing it was continuously going on (cf., e.g., Brennan, 2012; Harvey & Green, 1993; Harvey & Williams, 2010a; Harvey & Williams, 2010b; Newton, 2013; Rosa *et al.*, 2012; Stensaker, 2008). Despite all these efforts and QA activities, however, until today relevant stakeholders and the scientific community (cf., e.g., Harvey & Williams, 2010b, p. 102; Lillis, 2012, p. 59; Newton, 2013, pp. 9; 11, 13; Shah, 2012, pp. 761, 770) broadly agree that something very important is still largely missing, namely methodologically comprehensive and empirically reliable knowledge about the effects and mechanisms of action of QA measures in HE. Such knowledge is particularly important because various HE stakeholders are striving for evidence-based decision-making such as, for example, research-informed HE quality policy and HE politics. Accordingly, growing interest of higher education institutions (HEIs), QA agencies and other stakeholders can be observed, who want to learn more about the effectivity and efficiency of QA interventions in the sector (e.g., cf. Leiber, 2014a; Leiber, 2014b; Stensaker *et al.*, 2011; Suchanek *et al.*, 2012).

It seems that the main deficiencies of our understanding of QA and its effects in HE can be reduced to three points: One, QA's effects in HE are still 'under-theorised and under-researched' (Newton, 2013, p. 8). Two, the available methodological options for impact analysis in HE have not been exploited so far, i.e., analyses in the past were limited to *ex-post* scenarios (see, e.g., Stensaker *et al.*, 2011; Suchanek *et al.*, 2012). In other words, such approaches were mainly based on after-procedure judgements by selected participants (mostly peer experts). Three, it has been already criticized almost a decade ago that important stakeholders, such as students and teachers, have not been adequately involved in impact studies (cf. Volkwein *et al.*, 2007, p. 253; Westerheijden *et al.*, 2007, pp. 305, 309).

In consequence, a generic design how to implement comprehensive impact evaluations of QA in HEIs is required to overcome the abovementioned deficiencies of our understanding the causality of QA in HE. In particular, such a design must get over the limitation to *ex-post* scenarios and also include relevant stakeholder groups other than peer experts. A corresponding type of impact evaluation methodology and design was worked out and applied in the European Commission-funded project 'Impact Analysis of External Quality Assurance Processes of Higher Education Institutions' (acronym: IMPALA) which was carried out by eleven institutional consortium partners from six European countries between 2013 and 2016 (IMPALA, 2013-16). The impact evaluation methodology and design of the IMPALA project focuses on comparative and longitudinal case studies. The main methodological results and some empirical results[3] of the project are presented and discussed in this manual which is meant to be an inspirational and critical guide with fundamental methodological issues and clarifications next to implementation of methodology and practical application and also with warnings of possible failures to carry out impact analyses of QA in HE.

[1] Throughout this manual QA (quality assurance) is understood as comprising all activities assuring and further developing quality.
[2] Section 1.1 follows (Leiber et al. 2015, pp. 288-289).
[3] A more comprehensive exploitation of the empirical results of the IMPALA project will be embedded in the broader context of sustainable quality development of HE in knowledge societies and will be published as a special issue in the *European Journal of Higher Education*.

1.2 IMPALA Project Partners and Case Studies

As already mentioned above, this manual is based on the project 'Impact Analysis of External Quality Assurance Processes of Higher Education Institutions' (acronym: IMPALA) which was carried out by eleven institutional consortium partners from six European countries, namely Belgium, Finland, Germany, Norway, Romania, Spain. The IMPALA project was co-funded by the European Commission from October 2013 until September 2016.

The project's core was implemented by four partner groups of one QA agency and one HEI each from Finland, Germany, Romania and Spain. In other words, the four case studies of impact analyses were performed by FINEEC (Finnish Education Evaluation Council), Helsinki and Jyväskylä University of Applied Sciences, Jyväskylä; evalag (Evaluation Agency Baden-Württemberg), Mannheim and University of Stuttgart, Stuttgart; ARACIS (The Romanian Agency for Quality Assurance in Higher Education) and Technical University of Civil Engineering Bucharest, both Bucharest; and AQU Catalunya (Quality Agency for the University System of Catalunya) and Universitat Autònoma de Barcelona, both Barcelona.

Centrepiece of the IMPALA methodology is the causal connection of QA procedures and changes in structures and processes on the program and institutional level and attitudes of stakeholders. Each IMPALA partner QA agency together with its university partner carried out a QA procedure – i.e., program accreditation (Jyväskylä); evaluation of program reviews (Stuttgart); combined institutional and program accreditation (Bucharest); program pre-accreditation (Barcelona). In parallel, the agencies implemented simultaneous impact analyses which were based on surveys and document analyses which were carried out in the framework of a before-after comparison approach (see Section 3.2.2).

2 Basic Assumptions and Concepts Related to Impact Studies

2.1 A Working Definition of Causality

When reflecting about impact, two questions normally open the discussion[4]: which impact is under scrutiny, and the impact of what is it? In other words, when impact is on the agenda there is always present the assumption of a cause-effect (or causal) relationship. For such a cause-effect connection, it is usually assumed that, compared to the cause-event(s), the effect-event(s) occur(s) later in time, and that, everything else being equal (*ceteris paribus*), the effect-event(s) would not have occured in the same way without the said cause-event(s) having been there.

A most plausible working definition of causality reads:

> "*C* may be considered a cause of *E* if (and only if) it raises the probability of [the occurrence of] *E* [under ceteris paribus conditions]" (Gerring, 2005, p. 169).

Among other things, this definition comprises the following two fundamental ideas of causal social analysis: first, the event identified as a cause 'makes a difference', and second, that the causal relations of the empirical world typically cannot be adequately modeled by strictly deterministic mono-causal relations – one cause determines one and only one effect –, but only by multi-factorial probabilistic relationships (or causal networks) between the causes and their effects (probabilistic causation). Accordingly, causes can be (formally) classified as necessary, sufficient, and contributing ones. Here, the type of contributing causes signifies the most general and most widespread case: Multiple causes (which may be independent from each other or depend on other) act together in a network to produce one or more specific effects.

[4] Section 2 follows (Leiber et al. 2015, pp. 290-292).

In the context of impact analyses of interventions on a dynamic social system, an effect (E_I) of an intervention may be defined as the difference in the corresponding indicator of interest (I) with the intervention (I_1) and without the intervention (I_0), $E_I = I_1 - I_0$. In this sense, it is said that a cause-effect study attempts to attribute effects to its corresponding cause(s) 'by identifying the counterfactual value' (White, 2010, p. 154) of I, namely I_0, in a systematic manner.

2.2 Types of Effects: Outputs, Outcomes, Impacts

For the purposes of this manual and in accordance with widespread usage, short-term, mid-term and long-term effects are differentiated; they are called outputs, outcomes and impacts, respectively, and are all subsumed under the umbrella term 'effect'. This is in opposition to the fact that many use 'impact' as an umbrella term (as in 'impact evaluation', 'impact analysis' etc.), thus undermining the conceptually preferable alternative. However, this dispute about use of concepts, which ultimately is merely a matter of definition, cannot be resolved here.

2.3 Causal Social Mechanisms

In social science causal networks are also known as a 'causal social mechanism' (Gross 2009; Hedström & Ylikoski, 2010; Little, 2011; Little, 2015a; Steel, 2011). The epistemological idea of causal social mechanisms is

> 'that we explain not by evoking universal laws, or by identifying statistically relevant factors, but by specifying [causal] mechanisms that show how phenomena are brought about' (Hedström, 2005, p. 24).

Such social mechanisms can be understood as

> 'complexes of interacting individuals, [bodies and institutions] usually classified into specific social categories that generate causal relationships between aggregate-level variables. A mechanism will be said to be from the variable *X* to the variable *Y* if it is a mechanism through which *X* influences *Y*' (Steel, 2004, p. 59).

In a nutshell, a causal social mechanism is 'the [social] pathway or process by which an effect is produced or a purpose is accomplished' (Gerring, 2007, p. 178).

Thus, mechanism-based analysis is a basic method for understanding the social world by describing how micro-level and meso-level phenomena (such as attitudes and preferences of individual actors) bring about and produce meso-level and macro-level phenomena (such as common actions; emergence and diffusion of social practices; institutional and organizational changes). This is also called Agent-Based Modelling (Helbing & Balietti, 2015) which looks closely inside the social system's Russian stacking dolls and corresponds to giving a dense, mechanism-based (Astbury & Leeuw, 2010) and therefore explanatory description of the social interactions, its constituents (such as human motivation; cognitive processes; access to information; social relations) and the interactions' effects.

2.4 Complexity and Indispensability of Impact Studies

The basic and big obstacle to impact analysis (or causal analysis) of QA in HEIs based on causal social mechanisms is the shere complexity of the problem: QA interventions, as a rule, do have complex and manifold cross-effects on different subsystems on the micro-, meso- and macro-level of HEIs (e.g., sets of intentional states of individuals; sets

of psychological states of groups; organizational and institutional structures and processes). Particularly, QA interventions in HEIs in total have many different aims and purposes, and they are in competition and interplay with various other causes such as changing environment; other QA procedures; changes in HEI organization; policy measures; etc. (also cf. Beerkens, 2015; Stensaker & Leiber, 2015). Therefore, e.g., non-intended and undesirable effects and long-term effects may occur, and, normally, none of these is easily grasped at all. In summary, it is generally very difficult to adequately model the corresponding complicated (probabilistic) cause-effect, interaction, or cross-impact network.

In view of these basic complexity problems, it is the more important to be clear that in complex social systems (such as HEIs) the identification and measurement of cause-effect relationships, however difficult it may be, in general is of great importance, and ultimately unavoidable from an epistemological as well as societal perspective. Why so? Because organization and understanding of any educational planning and reform, and, in the end, any social life would be impossible without causal mechanisms and attendant regularities (Phillips & Burbules, 2000, p. 92). Or in the words of Daniel Steel:

> 'Causation is one of the most important and contentious issues in social science. Any aspiration for a better social world, whether they concern the allevation of inequities or the promotion of wealth, must explicitly or implicitly rely on beliefs about the causes and effects of government policies, social institutions, norms, or other phenomena that fall within the purview of social science' (Steel, 2011, p. 288).

3 Impact Evaluation of Quality Assurance in Higher Education Institutions

3.1 Contemporary Practice

3.1.1 Expected Impacts

In their work, the QA agencies of the IMPALA consortium hypothetically rely on numerous expected impacts of the various elements of external quality assurance (EQA) procedures. These expectations originate in the broadly applied classical peer review approach the core elements of which are assumed to have effects on the HEIs including their study programs: the assessment criteria; the self-assessment; the exchange with peer experts; an assessment report; and a formal decision with prescribed follow-up.

In the QA agencies' perspective, the expected impacts of EQA originate in the peer review in the following way: The EQA criteria represent guidelines to the HEIs against which the assessments – self-assessments of the institution as well as assessments of external peers – are made during the EQA procedure. Accordingly, one expected type of effects of EQA is that the institution will try to meet the criteria and, if necessary, adjust the institutional and program features; this can already take place just after the dissemination of the EQA criteria, i.e., before the actual EQA procedure starts. During a first self-assessment the HEI can reflect on its structures, processes and activities and adjust according to the results of the self-reflection. Another opportunity for self-reflection is given by the exchange with the peers during the site visit where the external experts give assessments, recommendations and advices. In a more formal (and binding) way the experts' assessment report documents a feedback, which is systematically aligned to the assessment criteria of the EQA procedure. The report's recommendations advise the institution how to improve. If the report, in addition, is also tied to a formal decision as in a certification (such as 'accreditation'), this can affect the HEI directy since it is usually linked to (formal) conditions which must be implemented in order to pass the certifying EQA procedure (see Table 1) (Kajaste *et al.*, 2015, p. 272).

Table 1: Typical (E)QA events and its expected effects (coarse-grained) (adopted from Kajaste *et al.*, 2015, p. 273)

Basic QA events	Expected effects and triggered causal social mechanisms		
	Before QA procedure	During QA procedure	After QA procedure
Dissemination of assessment criteria	Reflection (ideas for change); adjustment	–	–
Self-assessment	–	Reflection; (obvious, direct changes; adjustment)	–
Exchange with peers (e.g., site visit)	–	Reflection; direct recommendations for follow-up	–
Assessment report	–	Reflection; (obvious, direct changes)	Reflection; recommendations for follow-up
Formal decision	–	–	Fulfilment of accreditation requirements (conditional accreditation); follow-up with fulfilment
No formal decision	–	–	Voluntary follow-up

In practice, with respect to when effects of EQA procedures typically occur there seems to be the following difference between certifying (E)QA procedures (such as accreditations) and non-certifying (E)QA procedures (such as evaluations): in certifying procedures many, if not most, effects usually occur before the procedure actually starts, i.e., before the first formal self-assessment because the institution will try to adjust in advance as much as possible to the criteria in order to minimize the risk of fail. The main impact may therefore often take place before the procedure. If there is nevertheless only a conditional certification, however, other important effects may also occur after the formal decision of the procedure. In contrast to this, in the case of non-certifying procedures the (E)QA effects are usually more smoothly distributed across the procedure because the implementation of recommendations follows the rules of voluntariness.

3.1.2 Approaches to Impact Evaluation

The QA agencies of the IMPALA consortium use several methodological approaches for self-assessing the impact of the EQA procedures they carry out. Following the ESG (ENQA 2015), these are applied on a regular basis and based on feedback by HEI members, (E)QA peers and other stakeholders which is collected after each EQA procedure or at least after many of these. The approaches comprise short structured interviews, more comprehensive online questionnaires, workshops on survey results, focus group discussions and qualitative content analysis of assessment reports (see Table 2).[5] In addition to these approaches student satisfaction surveys are carried out in HEIs which comprise students' views on features of learning and teaching. Some of these features are usually also among the intended goals (and non-intended effects) of (E)QA procedures. This information can add to what is gained by the approaches listed in Table 2.

[5] For more details see (Kajaste *et al.*, 2015, pp. 276 ff.).

Table 2: Typology of impact evaluation of QA agencies' EQA (rearranged from Kajaste *et al.*, 2015, pp. 276, 280, 284)

Type of assessment	Methodology and instruments	Strengths/weaknesses
Structured (telephone) interviews with HEI members and peers	Qualitative questionnaires with closed and open-ended questions at the end of the site visit	Strengths - Gather the opinion of all stakeholders involved - Useful for the process of improvement Weaknesses - Few items addressing the impact of the project - Strategy is not outcomes-oriented (but process oriented) - *Ex-post* study: opinions are gathered after the assessment is done - Mostly carried out rather soon after the end of the QA procedure, so that there is not much time for change in the HEIs
	Qualitative questionnaires with closed and open-ended questions after the procedure	
Online questionnaires to HEI members (and peers)	Online questionnaires (with Likert-scale questions and short open-ended questions)	
Workshops on impact surveys results	Presentation and discussion of results with stakeholders	
Focus group discussion with external coordination	An external expert assesses the information collected through questionnaires and conducts focus groups with relevant stakeholders (administration, presidents of external and internal committees)	Strengths - See above - Triangulation of qualitative (focus groups, document analysis) and quantitative data (surveys) - In consequence, larger degrees of reliability and objectivity (confirmability) of the assessments, more trustworthiness Weaknesses - See above - More time-consuming - Costly
Qualitative content analysis	Document analysis (provisional reports) and online questionnaires for academics, technical units and deans	Strengths - Detailed evaluation (vs. global evaluation) - Individual questionnaires: for each requirement and recommendation in the assessment report, it is asked whether they were relevant and suitable - Triangulation of data: document analysis and surveys - Verifying with participants: each report is assessed by the responsible individual(s) or bodies Weaknesses - Highly time-consuming
Follow-up seminars with HEIs	Seminars in the midtime between a EQA certification and re-certification, open to all HEIs in the HE system, including benchmarking exercise	Strengths - Based on interim HEI self-reports - Discussion of EQA recommendation - Discussions between HEIs Weaknesses - Topic of impact of previous certification is not yet addressed
Publication seminar	Seminar with HEI at the end of EQA process to present and discuss findings	Strengths - Discussion of EQA recommendations - Focus on development of HEI Weaknesses - Dfficulty of separating assessment and consulting
Workshops with stakeholders	SWOT analysis of QA agencies' activities with HEI members and independent experts	
Meta-evaluation reports and system analyses of QA agencies	Analysis of several EQA reports with conclusions on the state of QA in HE	Strength - 'Thematic analyses' (ENQA, 2015, standard 3.4)
Internal further education	QA agencies' self-reflection of results of various impact assessments	

The IMPALA project showed that all European partner HEIs already have some instruments and procedures in place for impact analysis of QA, and all of them are striving for further improvement by applying internal QA (IQA) and EQA procedures and impact analyses. A closer look at the recently revised Standards and Guidelines for Quality Assurance in the European Higher Education Area (ESG) (ENQA, 2015) exhibits that the IMPALA partner HEIs are pretty much on their way for (more) coherent QA policies, while there is also room left for further improvement and investigation (for example, in implementing systematic methodologies of impact evaluation and quality policies in general).[6]

In summary, it seems that the IMPALA partner QA agencies' and HEIs' approaches to impact analysis of QA are promising since they use a spectrum of generic methodologies and thereby collect systematic knowledge and carry out 'thematic analyses' (ENQA, 2015, standard 3.4) of their HE systems.[7] There are probably only three aspects in which generic improvement is desirable: first, it seems that, so far, the HE players do not carry out longitudinal panel impact and follow-up studies with sufficiently extended time duration; second, all impact investigations are *ex-post* studies and do not comprise simultaneous effect studies; third, it seems that in the HEIs and QA agencies the methodological clarity of the used approaches seems to be improvable.

3.2 Methodology

3.2.1 Challenges and Basic Principles

The central task of impact analyses[8] consists in the proof that certain observed changes are causally generated by certain applied measures or interventions (the so-called attribution problem). In other words, an impact evaluation must identify effects – outputs, outcomes, and impacts – of some interventions (i.e., the causes) to which the dynamic system is exposed. Core questions are therefore:

- Which intended effects should be achieved by which intervention? When could they be achieved?
- Which non-intended, but desirable and which non-intended undesirable effect(s) might occur?
- Through which causal (social) mechanisms might the interventions be effective?

In the framework of theory-based impact evaluation (cf. White, 2009, p. 274 ff.) these questions are treated by considering the following three principles: devising the causal network; understanding the context; methodological pluralism.[9] In addition, there are three methodological requirements of impact evaluations: One, they must meet quality criteria of social science research (such as relevance and empirical reliability) and evaluation standards (such as utility; realisability; fairness; precision) (DeGEval, 2011). Two, accompanying research towards impact analysis should be incorporated in all phases of interventional programs because strategic program management cannot do without reliable impact knowledge. Three, the choice for an impact evaluation should be based on a benefit-cost analysis.

[6] For details see (Bejan *et al.*, 2015, pp. 367 ff.).
[7] Published material can by found in the references' lists of (Bejan *et al.*, 2015; Damian *et al.*, 2015; Kajaste *et al.*, 2015).
[8] Section 3.2 follows (Leiber *et al.*, 2015, pp. 292-305).
[9] For a description of these principles, see (Leiber *et al.*, 2015, pp. 294-295).

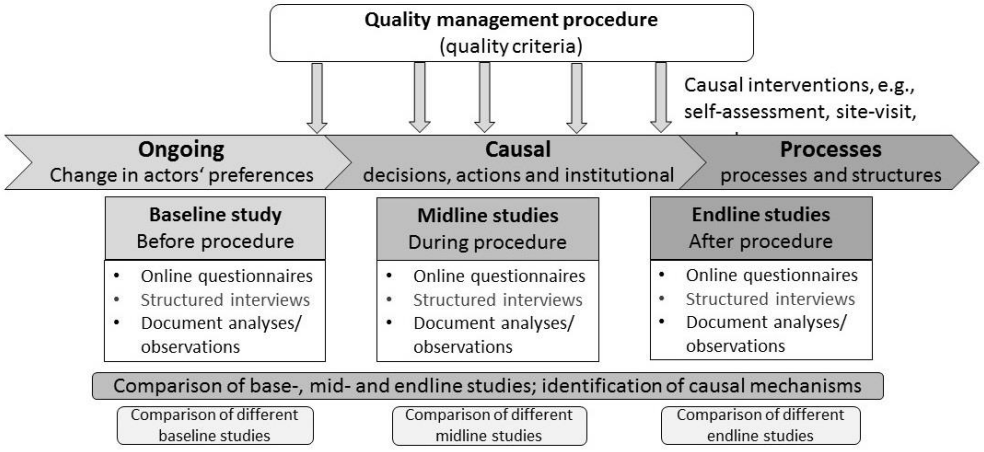
3.2.2 Basic Designs

The next systematic step consists in deciding which methodological design is applicable to impact evaluation of QA in HE. In general, there are four such designs: experiment; control group; before-after comparison; *ex-post* analysis (Leiber *et al.*, 2015, pp. 296-297). The first three of these are based on simultaneous impact evaluation thereby avoiding exclusive dependency on *ex-post* available data.

To make a longer story short, experimental design and control group design, however, cannot be applied to impact analysis of QA in HEIs for two reasons: one, the original state of the experimental system in question would not be recoverable or reconstructable, and, two, there is no control group (i.e., a HEI without the QA measure in question) and treatment group (i.e., a HEI experiencing the QA measure while everything else being equal). In consequence, for impact analysis of QA in HEIs before-after comparison design (or longitudinal research design) and *ex-post* analysis are the applicable approaches.

Before-after comparison design (longitudinal research design): This approach is based on the idea to compare the process state of the investigated system after a certain intervention took place with the *status quo* of the system just before the intervention. The biggest threat of before-after comparison is again the attribution problem, i.e., to identify effects and their relevant causes and to sort out which effects were caused to which extent by the applied interventions, and not by other possible causes. The approach tackles this threat by carrying out a baseline study, which establishes the reference for before-after comparison, and an endline study, which defines the end of the observed processes. In order to carry out a more dense analysis additional midline studies are usually approporiate which establish in-between references for before-after comparison (se Figure 1). Ideally, the baseline study should be carried out before the first interventional measure of the QA procedure in question takes place. This would ensure that the baseline shows the *status quo* of the system before the QA procedure appeared on the scene and therefore the mid- and endline would gather all possible effects of the QA procedure on the system (Leiber *et al.*, 2015, p. 297).

Figure 1: Before-after comparison research design of impact evaluation, adopted from (Leiber, 2016, p. 8)



Ex-post analysis design: This approach relies on attributing assumed effects to certain interventions after the end of the intervention procedure. The typical sources of information of *ex-post* analysis are documents and data as well as assessments of experts and stakeholders (Leiber *et al.*, 2015, p. 297).[10]

3.2.3 Further Design Elements

In general, the following further methodological elements are required for any of the impact analysis designs mentioned in the previous section: analyses of documents and data collections; hypotheses about cause-effect relationships or causal social mechanisms; assessments of intervention effects by participants, key informants, and experts; counterfactual self-estimation of participants; and comparative case studies.

Analyses of documents and data collections: This comprises the analysis of relevant documents of the HEI and its QA as well as collected data (such as performance indicators which are monitored) (Leiber *et al.*, 2015, p. 298).

Hypotheses about causal social mechanisms: Such hypotheses will help to understand the system's dynamics and effectivity of QA interventions. They would raise the study beyond the observation of mere correlations and allow specific questions and system analyses (Leiber *et al.*, 2015, p. 298).[11]

Assessments of intervention effects by participants, key informants, and experts: Another important source of information is the assessment of effects and cause-effect relations by various participants of the QA procedure. For example, this information can be gathered by (online) surveys (with closed and open-ended questions), structured interviews and participants' observation (e.g., participation of experts in SWOT-analysis seminars) (Leiber *et al.*, 2015, p. 298).

Counterfactual self-estimation of participants: Counterfactual self-estimation of program participants (CSEPP) provides a basis for the assessment of individual and average effects of interventions without having to refer to information about either a control group or a before-after comparison. With CSEPP, intervention recipients assess the causal counterfactual for themselves (cf. Mueller *et al.*, 2013) – 'In what state would I be if I had not experienced the intervention?' This implies that CSEPP can only be applied for the assessment of self-reported personal variables such as individual intentional states like the quality of my attitude towards a certain feature of the world. At the same time, intervention effects on other persons or collective outcome variables cannot be assessed by CSEPP (Leiber *et al.*, 2015, p. 298).

Comparative case studies: This approach can be understood as a pragmatic, approximative substitute for a control group design: while a control group in the strict sense of the word may not be 'interpretative comparison of different, but somehow similar (and thus comparable) cases may help to quasi-approximate a with-without comparison to some extent' (Leiber *et al.*, 2015, p. 298).

In general, a reliable and relevant impact evaluation will have to be based on a combination of different methodologies in order to take advantage of cross-check and mutual corroboration of results.[12] The question which methodology could be used in particular cases, is to be answered with respect to the QA goals, the QA measures applied, the concrete aims of the impact analysis, and the intended cost-benefit relation (which is characterized by the available time and money versus the required/desired degree of relevance and precision of results) (Leiber *et al.*, 2015, p. 298).

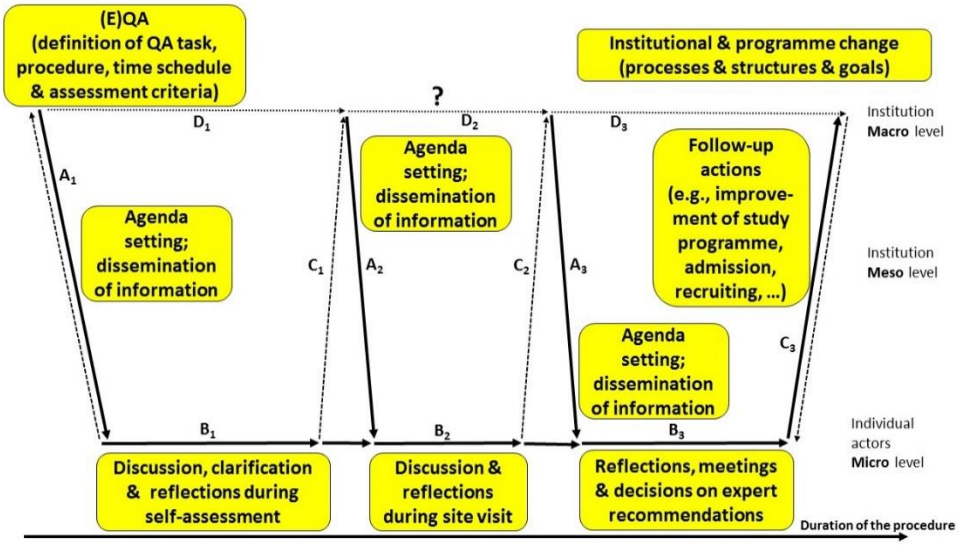
---

[10] So far, to the knowledge of the IMPALA consortium partners only *ex-post* analysis was used for impact analysis of QA in HEIs.
[11] For more details about this important methodological element, see the next section below.
[12] This is also known under the terms of triangulation and mixed methods.

3.2.4 Causal Social Mechanisms: Models, Advantages and Challenges

In this section a short overview of the use of the important methodological element of causal social mechanisms in impact evaluation of QA in HEIs is given.[13] A basic model of causal social mechanisms builds on Coleman's boat (Coleman, 1994, p. 8) which distinguishes three types of possible mechanisms which altogether lead from an intervention on the macro level of a system to the effects of this intervention on the same systemic level (see Figure 2): situational mechanisms, A_i, lead from the macro to the micro level of involved individuals and groups; action-formation mechanisms, B_i, causally connect this micro level with the micro level of decisions and actions of individuals and groups; transformational mechanisms, C_i, generate macroscopic changes on the institutional level from the individuals' and groups' actions. According to sociologist James Samuel Coleman such a cascade is assumed because the direct macro connection, D_i, is accessible only in the form of statistical correlations (see Figure 2) and leaves the causal links in the dark (Leiber *et al.*, 2015, p. 301).

Figure 2: A basic model of causal social mechanisms underlying 'QA meets HEIs' inspired by (Astbury & Leeuw, 2010) (A_i: situational mechanism; B_i: action-formation m.; C_i: transformational m.; D_i: statistical correlation only; i = 1, 2, 3), adopted from (Leiber *et al.*, 2015, p. 294)



Moreover, it is clear that much more causal social mechanisms are possible and often involved in processes such as characterised in Figure 2. Some exemplary mechanisms are listed in Table 3: Following Mark and Henry (2004), four types of QA mechanisms or outcomes (general influence, cognitive and affective, motivational, behavioural) are distinguished each of which has its examples on three different levels of analysis (individual, interpersonal, collective). Typically, a QA procedure may be working by utilising process mechanisms such as, e.g., elaboration, persuasion and policy change which may also trigger each other. Usually, the result is a complexely intermingled network-cascade of mechanisms relating primary (E)QA measures with 'final' effects, whose fine-grained nesting makes a reliable causal analysis often very difficult or even impossible.

[13] For more details see (Leiber *et al.*, 2015; Little, 2015b).

Table 3: Exemplary mechanisms mediating (E)QA influence on HEIs, adopted from (Leiber *et al.*, 2015, p. 293)

Type of mechanism/outcome	Mechanisms/outcomes		
	Individual	**Interpersonal**	**Collective**
General influence	Elaboration; Heuristics; Priming; Skill acquisition	Justification; Persuasion; Change agent; Minority-opinion influence	Ritualism; Legislative hearings; Coalition formation; Drafting legislation; Standard setting; Policy consideration
Cognitive and affective	Salience; Opinion/ attitude valence	Local descriptive norms	Agenda setting; Policy-oriented learning
Motivational	Personal goals and aspirations	Injunctive norms; Social reward; Exchange	Structural incentives; Market forces
Behavioural	New skill performance; Individual change in practice	Collaborative change in practice	Program change; Policy change; Diffusion

A clear example reads as follows (Leiber *et al.*, 2015, pp. 302-303; Mark & Henry, 2004, pp. 42-43): (E)QA measures may lead to an individual's elaboration of thoughts about a study program module; this might change the person's attitude towards the module and thus lead the individual to take on the interpersonal role of a change agent with respect to learning and teaching; further on, this might cause a reconsideration of organizational policy in learning and teaching and ultimately lead to a policy change on the collective-level.

Following these considerations, a schematic model of (E)QA influences in HEIs is given in Figure 3. The model distinguishes the (E)QA inputs (the 'context', such as environment, policy setting, evaluation context), (E)QA activities and (E)QA outputs and outcomes. The various lists of possible inputs, activities and outputs/outcomes in Figure 3 are neither small nor complete. Moreover, the arrows only indicate the basic possible directions of influence between items represented in the boxes, while all of these can be visited iteratively. If one attempts to model a specific (E)QA procedure based on the model scheme, one will easily and quickly find that it is most often very complicated to establish a sufficiently isolated causal mechanism model.

Advantages and Challenges to Causal Mechanism Approaches

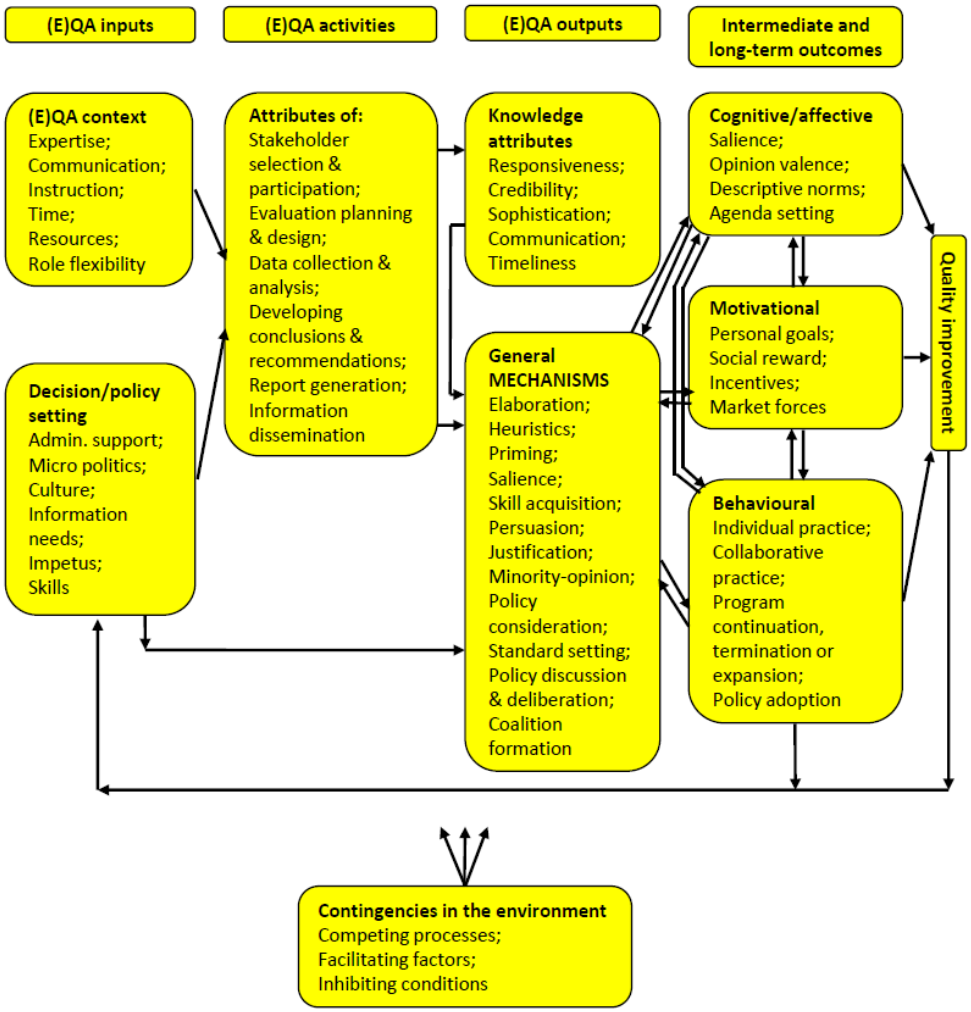
In the present context of impact analysis of QA in HEIs, the main goal of a social mechanism approach (on the operational level) is to allow QA implementers to identify research questions and methods of data acquisition which enable QA implementers to test whether QA procedures are effective. An eloquent example is that survey approaches and structured interviews can be used to analyse information dissemination processes in HEIs because these approaches consider 'the extent to which information is actually getting to individuals' and stakeholder groups 'in a timely [and elucidating] fashion', or they help to understand the type of interpersonal communication network structure that may 'influence information exchange greatly' (Astbury & Leeuw, 2010, p. 373; also cf. Leiber *et al.*, 2015, p. 303).

Unfortunately, for a number of reasons it can be very hard to explicitly conceptualise the mechanisms mediating between cause(s) *C* and an effect(s) *E* (also cf. Gerring, 2010, pp. 1510-1511):

> 'First, social mechanism hypotheses are often not easy to operationalise. For example, it may be difficult to find plausible indicators, or there may be too many of them. [For example, think about the realistic situation that several causal social mechanisms of those listed in the 'general mechanisms' box in Figure 3 are

activated simultaneously and re-iteratively.] Second, normally there is a multiplicity of plausible mechanisms M_1, M_2, … connecting a certain cause *C* and a distant effect *E*. In some cases this multiplicity may not be easily separable, so that it is hard, or even impossible, to empirically test the involved mechanisms separately. Third, the various causal mechanisms M_1, M_2, … may be of different types; they may provide contributing causes for *E* or sufficient causes or necessary causes. Fourth, depending on the context of an event a causal mechanism can have opposite effects. Fifth, not all causal factors influencing the effects can be controlled for ethical reasons. Sixth, the phenomenon of multiple realizability, which is widespread in complex dynamical systems and means that the connection between a cause and an effect can be established in different and equally valuable ways, may further complicate the situation' (Leiber *et al.*, 2015, p. 303-304).

Figure 3: Schematic model of (E)QA influence in HEIs, adopted from (Leiber *et al.*, 2015, p. 304)



In consequence, this leads to a pragmatic attitude which was already summarized by John Gerring a few years ago with the following words:

'What we need is intelligent discussion of plausible causal mechanisms, which should be subjected to empirical testing to the extent that is feasible. What we

should appreciate is that this objective is rarely fully achievable when one is dealing with distal causal relationships, and it is often unnecessary (in the sense of being trivial) when dealing with proximate causal relationships' (Gerring, 2010, pp. 1518-1519).

4 Basic Characteristics of Survey Questionnaire Construction

When constructing survey questionnaires for QA impact in HEIs, in view of the variety of possible effects and their possible causes the question arises what the rationale is behind the questionnaire questions. Here, it is suggested that the following six general topics are relevant for questionnaire construction:

One, the non-trivial insight should be taken into account that, in general, any QA procedure offers a large number of interesting options where to look for its various effects. Therefore, in practice it is indispensable to focus any impact analysis on some achievement area (such as, e.g., research; learning and teaching; administration; etc.) and some treatable (e.g., in terms of time and conceptual simplicity) subarea (e.g., shift from teaching to learning; focus on problem-based learning; management of student life-cycle; fulfillment of specific ESG standards; internationalization in research; etc.) of possible options, and within such an area again to a treatable subset of relevant survey questions. The respective decisions must be made explicit (Leiber *et al.*, 2015, pp. 305-306). (It is probably a truism but nevertheless worthwhile to mention that such complexity reduction, when looking at the final product of the questionnaire trimmed to applicability, may arouse a feeling of unsatisfactoriness, because it may seem that there are too few questions considered, they may seem to be too simple, important questions may seem to be missing, etc. pp.).

Two, the respondents should be asked whether they have observed achievements of intended QA goals; when they have observed them[14]; and what generators of these achievements they assume.[15] Appropriate closed-form questions in the questionnaires will represent the researcher's assumption about causal mechanisms (situational mechanisms, action-formation mechanisms and transformational mechanisms as modeled in Figure 2) without explicitly asking for them. Ideally, it will then be checked by answering the questionnaire several times in a longitudinal study whether these mechanisms, according to the judgement of the respondents, are at work (Leiber et al. 2015, p. 306). When doing this it is recommended to avoid explicit causal mechanism talk in the survey questions in order to prevent that the answers are uncontrollably biased by the respondents' understanding of causality (which is generally unknowable to the researchers).

Three, 'questions should be adapted to the presumed knowledge and engagement of the addressed stakeholder groups and be easily understandable by its members' (Leiber *et al.*, 2015, p. 306). Four, 'respondents should be asked for their suggestions for improvement of QA, specifically desirable effects not achieved so far, and of measures that might presumably generate those effects. Answers can disclose respondents' beliefs about which causal mechanisms are or could be at play in the current QA process and, more important, give insider information for improving causal analysis and future design of QA processes' (Leiber *et al.*, 2015, p. 306). Five, 'respondents should have options for giving comments to open an exploratory dimension to the survey' (Leiber *et al.*, 2015, p. 306). Six, 'questionnaires should be as short as possible. In particular, questions that could be better answered by intensive interviews or documentary analysis should be excluded from the questionnaires' (Leiber *et al.*, 2015, p. 306).

[14] E.g., long ago; recently; after a certain measure of the EQA procedure; etc.
[15] E.g., EQA; IQA; stakeholder interests; HEI governance; etc.

Going one step further in the analysis, the following four groups of leading questions for guiding the survey design can be distinguished:

> '(1) Which aims should be achieved by the impact evaluation? Which are the most important of these? (2) Which intended, non-intended, undesirable effects could be caused by the applied QA procedure? Which are the most important of these? In more detail: What are the effects of QA procedures for the various areas such as governance, IQA, institutional organisation, research programmes, young scientists' development, curricula, administration and so on. What effects of QA do the various stakeholders expect? (3) Which mode of causal analysis is intended and achievable? Which causal social mechanisms could be hypothesised and should be tested by the impact evaluation? Given a certain QA process or measure, which situational mechanisms, action-formation mechanisms and transformational mechanisms (Table 3 and Figure 3) can be hypothesized and put into closed-form survey questions?" (Leiber *et al.*, 2015, p. 306). (4) Which data, information and information resources are required? Who has relevant knowledge and therefore should be surveyed? When, in the course of an EQA procedure, should a stakeholder group by surveyed so that observable changes are reasonably expectable and, at the same time, the possible causes of these changes are overseeable and methodologically controllable? Such conditions should be met to enable the (re-)construction and empirical testing of hypothetical causal relationships' (Leiber *et al.*, 2015, p. 307).

5 Preliminary Results of the IMPALA Project

In this section a preliminary analysis and interpretation of some exemplary data of the impact analyses carried out in the IMPALA project is presented.[16] The project approached the four stakeholder groups of students, teaching staff, institutional leadership and quality managers at the four partner universities in Barcelona, Jyväskylä, Stuttgart and Bucharest with a baseline, midline and endline online survey in Catalan, Finnish, German and Romanian, respectively. The surveys were focusing on the areas of learning and teaching and, to a lesser extent, institutional management (see Appendix). The generic parts of the questionnaires were common to all stakeholder groups and asked for the presence, and in some cases also for recent changes and possible causes of these changes of the following items: course types in study programs; QA instruments used in study programs; alignment of examinations and learning objectives; frequency of development discussions of study programs; observability of QA effects and quality improvements; transparency of responsibilities; attitude towards EQA and IQA; perceived attitude of leadership towards QA; assessment of cost/benefit ratio of QA; plans for major program changes; suggestions for QA improvement (Leiber, 2016, pp. 8-9).

From these items, students' attitude towards QA and teachers' assessment of cost/benefit ratio of QA are selected for the analysis in the next two subsections. For reasons of anonymity, in the following the case study universities are coded as Uni1, Uni2, Uni3 and Uni4.[17] Due to the specifities of the four different QA procedures in the IMPALA project, namely, program pre-accreditation, program accreditation, evaluation of study program reviews, and combined institutional and program accreditation, the engaged

[16] The IMPALA project was formally finished at the end of September 2016; however, data analysis is not finished yet. A final and comprehensive analysis and exploitation of the empirical results of the IMPALA project will be published as a special issue of *European Journal of Higher Education* (online expected towards the end of 2017, printed in 2018).
[17] Participation in the surveys was voluntary. Data are treated completely anonymously and in accordance with the applicable data and privacy protection regulations.

stakeholder groups who can participate in informative surveys, are of different size in the different cases (Leiber, 2016, pp. 9).[18]

5.1 Students' Attitude Towards Quality Assurance

In the online survey, students were asked the following questions:

In general, what is your attitude towards quality assurance in learning and teaching?

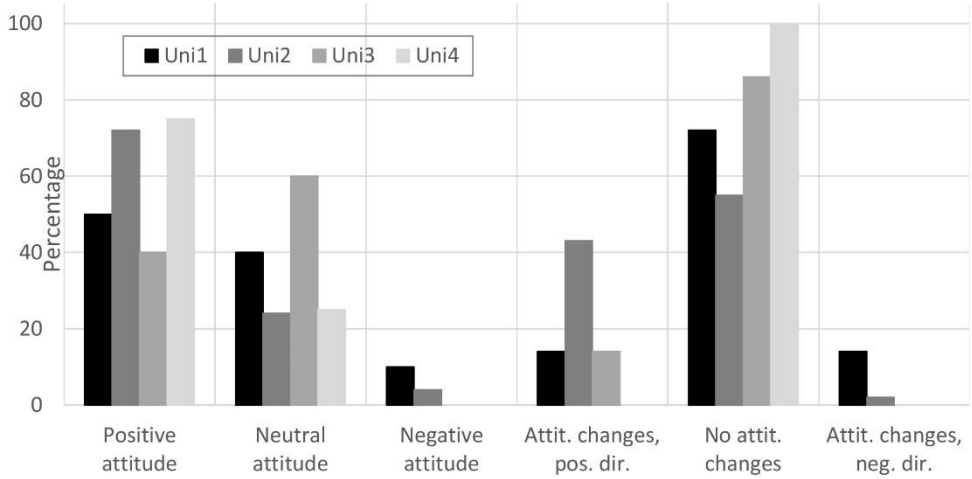
Has your attitude changed in the last year?

What has changed your attitude?

Assigned to the last question were four response options: 'experience with IQA procedures (e.g., course surveys; module evaluations etc.)'; 'experience with EQA procedures (e.g., writing of self-evaluation reports; on-site assessments etc. in accreditations)'; 'taking note of peer reports'; 'other [to be mentioned]'.

According to the baseline data, students' attitudes towards QA are rather positive in all of the four investigated HEIs (see Figure 4): 90 %, 96 %, 100 % and 100 % of responding students of Uni1, Uni2, Uni3, and Uni4, respectively, characterise their attitude towards QA as positive or neutral, while 50 %, 72 %, 40 % and 75 % report a positive attitude, obviously rather high levels in particular in Uni2 and Uni4 (where the 'neutral block' is significantly smaller as compared to Uni1 and Uni3). Notable negative attitudes (10 %) are only reported at Uni1 (see Figure 4). Similar results shine up with respect to recent of changes of attitude towards QA: Notable negative changes during the previous year before the QA procedure under study was initiated are only reported at Uni1 (14 %); during the same period of time, 43 % of student respondents from Uni2 observed changes of their attitude towards QA in a positive direction; students' attitudes towards QA before the IMPALA QA procedure started, are relatively stable, in particular in Uni1 (75 %), Uni3 (86 %) and Uni4 (100 %).

Figure 4: Percentage of students' at Uni1, Uni2, Uni3, and Uni4 reporting their attitudes towards QA, baseline data; adopted from (Leiber, 2016, p. 9)



The most influential factor reported for changes of attitudes in the last year was 'experience with IQA procedures' which is named by 95 % or more respondents as 'main' (40-

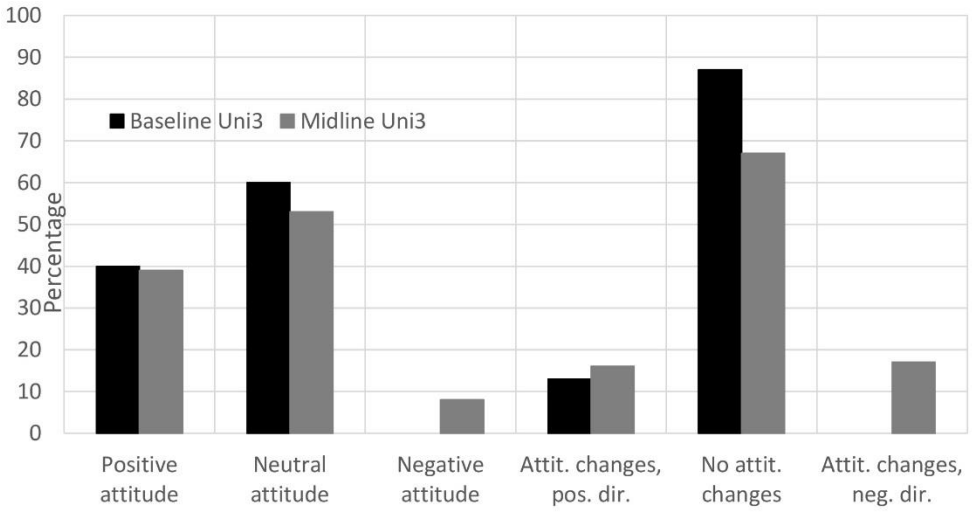
[18] While in the German impact study of program reviews only a few dozens of surveyed persons are sufficiently involved, in the other cases the number of respondents ranges from 60 to 200, depending on the number of students in the assessed programs and the number of involved QA staff or teachers.

60 %) or 'partial' (35-60 %) cause. 'Experience with EQA procedures' and 'taking note of peer reports' were also mentioned, but are seen less influential. The first of these was called a 'mainly' influential factor by 0-20 %, a 'partially' influential factor by 50-80 % and 'not at all' of influence by 0-50 % of respondents. For the answer option 'taking note of peer reports' the corresponding figures are: 10-15 %, 25-30 % and 25-30 % respectively (Leiber, 2016, p. 9).

The just mentioned relative stability of students' attitudes towards QA in the baseline (see Figure 4) suggests to investigate this aspect in the midline and endline. A significant change of the reported level of 'no changes in attitude towards QA' is a plausible indicator for the carried out QA procedure to have had at least some effects. The analysis of Figure 5 below supports this argument (Leiber, 2016, p. 9).

Preliminary results from the IMPALA project are also available for the development of students' attitudes towards QA. For example, a before-after comparison can be carried out with the baseline and midline of Uni3 (see Figure 5).

Figure 5: Percentage of students' at Uni3 reporting their attitudes towards QA, baseline and midline; adopted from (Leiber, 2016, p. 10)



The survey results presented in Figure 5 show that the students' negative attitudes towards QA are very low, at least at Uni3: many students report a positive attitude (baseline: 40 %; midline: 39 %), and even more students tell a neutral attitude (baseline: 60 %; midline: 53 %). At the same time the percentage of students' with a negative attitude towards QA is increasing from the baseline to the midline from 0 % to 8 % (with corresponding decreases in the positive and negative attitude items).[19] Accordingly, student respondents of the midline tell more recent changes of their attitude towards QA in a negative direction as compared to the baseline (from 0 % to 17 %). Simultaneously, a slight increase of changes in a positive direction (plus 3 % from baseline to midline) and a much stronger decrease of 'no changes' of attitude (minus 20 %) is reported.

These data may be interpreted as follows:

> 'The specific experience of […] [a particular] program accreditation (which was carried out at Uni3 for the first time) was a positive experience for a very small portion of responding students of the study program and a negative experience

[19] The survey response rates of the online survey to the question "What has changed your attitude?" are too small to be used here.

for some of them. These changes are balanced by a decrease of the 'neutral majority' which suggests that the observed changes could be caused by a more detailed acquaintance with the program accreditation which was not present before. However, for a more reliable interpretation further data from a deeper system and context analysis are required' (Leiber, 2016, p. 10).

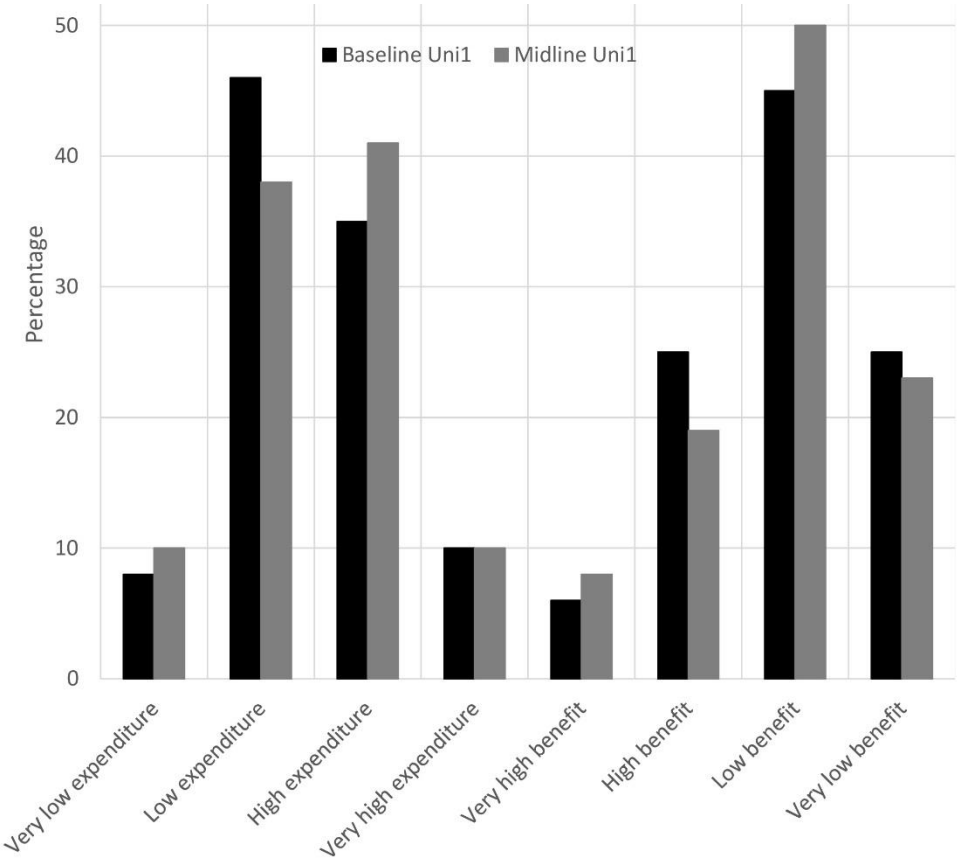
5.2 Teachers on Expenditures and Benefits of Quality Assurance

In the online survey, teachers were asked the following question:

How do you assess expenditure and benefit of quality assurance?

In each dimension, expenditure and benefit, the four possible response options were 'very low', 'low', 'high' and 'very high'. Respondents' data are taken from the baseline and midline of Uni1 ($N_{baseline}$ = 88, $N_{midline}$ = 61) and Uni 3 ($N_{baseline}$ = 13, $N_{midline}$ = 10) (see Figures 6 and 7).

Figure 6: Percentage of teachers' at Uni1 assessing QA expenditure and benefit, baseline and midline; adopted from (Leiber, 2016, p. 10)

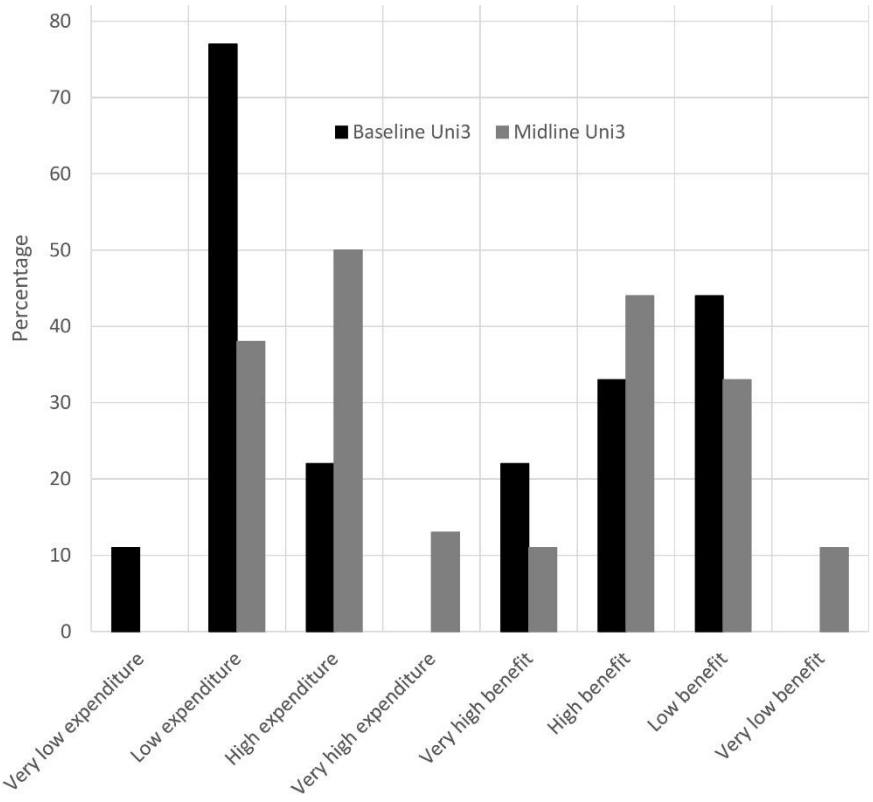


The data can be interpreted by assigning numerical weights to the qualitative response options: 1, 2, 3, 4 from 'very low expenditure' through 'low' and 'high' to 'very high expenditure', and 4, 3, 2, 1 from 'very low' benefit through 'low' and 'high' to 'very high benefit'. Assessment averages can then be calculated for baseline and midlines of expenditures and benefits for each of the two universities. At Uni1, teachers' average assessment of the expenditure is between 'low' and 'high' in the baseline as well as in the midline. If one characterizes the average by the arithmetic meanvalue Ø, formally Ø=2.5

(see Figure 7).[20] This relative stability of expenditure assessment can be seen as a 'weak indicator that at Uni1 the QA procedure does not noticeably affect teachers' assessment of the expenditure of QA. A more detailed and more precise interpretation can only be given after further analysis of the context (e.g., structured interviews with engaged informants), the endline study and the consideration of standard statistical errors' (Leiber, 2016, pp. 10-11).

The data also reveal that at Uni1 teachers' average assessment of the benefit is also constant between the baseline and midline at a 'low' value, Ø=2.9 (see Figure 6). Overall, teachers in QA-monitored programs assess QA expenditure as rather high and QA benefit as low (Leiber, 2016, pp. 11). Causes of this negative overall assessment may be uncovered by further qualitative analysis.[21]

Figure 7: Percentage of teachers' at Uni3 assessing QA expenditure and benefit, baseline and midline; adopted from (Leiber, 2016, p. 11)



At Uni3 teachers' average assessment of QA expenditure grows significantly between the baseline, Ø=2.1 (which is more positive as compared to Uni1), and the midline, Ø=2.8 (which is more negative as compared to Uni1) (see Figure 7). Leaving other possible causes aside for the moment, this increase might be caused by the instance that the QA procedure applied at Uni3 is broader in scope as compared to Uni1 (Leiber, 2016, pp. 11).[22]

The benefit assessment data reveal that, in accordance with expenditure assessments, Uni3 teachers' average benefit assessment decreases by approximately 10 % from the

[20] Strictly speaking, decimal numbers are not defined in the present case because metric data are not available.
[21] Anecdotal conjectures speculate that the longer tradition of the specific QA procedure and its design features could be the reason of some dissatisfaction of some directly involved stakeholders.
[22] A more detailed interpretation is only possible *via* further qualitative analyses of the context and the endline data.

baseline (Ø=2.2) to the midline (Ø=2.4), although the benefit assessment is significantly higher as compared to Uni1 (see Figure 7). Overall, teachers in QA-monitored study programs assess QA expenditure as rather low before the QA procedure was carried out, and (rather) high in the midline (right after the on-site visit of the peers). In the same period of time, teachers' benefit assessment decreases from rather high (Ø=2.2) by approximately 10 % (Ø=2.4, inbetween 'high' and 'low') (Leiber, 2016, pp. 11). Again, causes of this negative overall assessment may be uncovered by further qualitative analysis.

In all possible impact cases discussed above further data and (qualitative) analyses are required to investigate in more detail context-dependent causal mechanisms that could help to better understand the causality of the observed changes. Therefore, in line with the above discussion of causal social mechanisms 'it is desirable to carry out structured interviews with stakeholders, analysis of evaluation reports, analysis of recommendations and follow-up measures' (Leiber, 2016, pp. 10).[23] In summary, further research and a more detailed look at the available data[24] (including the consideration of standard statistical errors) and into the causal networks of the study cases is necessary.

6 SWOT Analysis of Impact Evaluation of QA in Higher Education Institutions

In this section a quick analysis of the advantages and problems of impact evaluation of QA in HEIs is given. This will be done by means of a so-called SWOT analysis, which may be characterised as follows:

> 'A SWOT analysis is a structured assessment method which evaluates the strengths (S), weaknesses (W), opportunities (O) and threats (T) involved in a process or a structure in the most general sense of these terms. A SWOT analysis involves specifying the objectives of the process or structure, identifying the internal and external influences with regard to the degree of achievement of these objectives and, finally the core element, characterising the strengths, weaknesses, opportunities and threats of the process or structure under scrutiny. In general, a SWOT analysis can help developing the assessed entities for further rounds of improved goal achievement, and it usually has an exploratory dimension bringing to the fore aspects which have not been noticed by other means of analysis. This exploratory force originates from the requirement to identify and distinguish explicitly the four different categorisation dimensions of processes or structures' (Leiber, 2017, p. 290).

The SWOTs of impact evaluation of QA in HEIs can be subdivided into the two groups of general methodological SWOTs and SWOTs which are more specific to the IMPALA project. The former SWOTs may be even further subdivided with respect to methodological sub-types such as before-after comparison, *ex-post* analysis, etc.

6.1 Methodological SWOTs

On the generic methodological level, it may be called a weakness of QA impact evaluation in HEIs that an experimental approach and a (quasi-experimental) control group approach cannot be implemented (cf. Section 3.2.2). At the same time, before-after comparison can be named the main methodological strength of QA impact evaluation in HEIs, a statement which also applies to *ex-post* analysis. Further strengths are that QA impact evaluation supports evidence-based QA, governance and strategy. These

[23] For a recent argument in the same direction of exploiting qualitative causal process tracing, see (Bennett & McWhorter, 2016).
[24] This will be done by the IMPALA consortium partners in a future publication which is planned as a special issue in *European Journal of Higher Education*.

strengths are accompanied by certain threats: a general and by no means trivial one being that the chosen methodology must be properly implemented; another threat is to ensure an impact evaluation which is not improperly biased by specific parties' interests; finally, it is a general methodological threat to balance the complexity and effort of impact analysis with the reliability level and benefit of impact assessments (see Table 4).

Table 4: Generic methodological SWOTs (S = strengths; W = weaknesses; O = opportunities; T = threats) of QA impact evaluation in HEIs

S	W	O	T
Before-after comparison Observations & change assessments by participants; Analysis of documents and data; Counterfactual self-estimation; Causal social mechanisms; Methodological principles: devising the causal network; understanding the context; counterfactual and factual causal analysis; mixed methods approach	No experiment	–	Proper implementation of chosen methodology
Ex-post analysis (dto.)	No comparison/control group	–	Independence of impact evaluation from HE politics, HEI leadership, QA agencies, … (impartiality)
Supports evidence-based QA	–	–	HEI performance probably too complex for reliable impact assessments (?); possible/ fruitful to look into micro-mechanisms?
Supports evidence-based HE policy (strategy; governance)	–	–	–

Further methodological strengths of the before-after comparison approach are that the typical memorisation problems of involved persons and other time-lag problems, which occur in *ex-post* approaches, can be avoided; that the analysis is not limited to *ex-post* available data. Ideally, in a before-after comparison the same surveyees should be approached in the different timelines of the longitudinal study (baseline, midlines, endline); it is therefore a weakness of before-after comparison when applied in HEIs that typical universities' stakeholder groups are largely fluctuating, and it is often hard to guarantee that the same stakeholders are responding in the different timelines. The core opportunities are that causal mechanism hypotheses can be applied and dense longitudinal analyses can be carried out. The two main threats of before-after comparison are the ubiquitous attribution problem, and the equally omnipresent question of affordable expenses with respect to time and money (see Table 5).

Table 5: Methodological SWOTs of before-after comparison impact evaluation

S	W	O	T
No memorisation problems of involved persons	**Fluctuating stakeholder groups**	Causal mechanisms; analytical models	**Attribution problem** (which effects are caused by QA and not by other causes)
No other time-lag problems	–	**Dense longitudinal analyses** (e.g., several ‚midlines')	**Expenses** (time, money)
No relegation to *ex-post* available data	–	–	–

As already mentioned above, the ex-post impact analysis suffers from memorisation and time-lag problems as well as from fluctuating stakeholder groups and the ubiquitous attribution problem. On the positive side counts the strength that *ex-post* analysis is always

applicable without specific methodological preparation and effort (see Table 6); in the end, this makes *ex-post* analysis indispensable in practice although it is not the approach which is methodologically most advanced.

Table 6: Methodological SWOTs of *ex-post* analysis impact evaluation

S	W	O	T
Always applicable, without specific methodological preparation and effort	Memorisation problems of involved persons	–	**Attribution problem** (which effects are caused by QA and not by other causes)
–	Other time-lag problems	–	–
–	Relegation to ex-post available data	–	–
–	**Fluctuating stakeholder groups**	–	–

Refering to (change) assessments by participants and analysis of documents and data in before-after and *ex-post* approaches, main methodological strengths are the usability of standardised (online) surveys and structured interviews which allow for complete acquisition of target groups. Another strength consists in including participant observation which can incorporate original views into practice that may not be achievable by other methodologies. A major, non-trivial and omnipresent threat is that the survey and interview instruments must be qualitatively adapted to social, organisational and cognitive contexts of the individuals who are surveyed (see Table 7).

Table 7: Methodological SWOTs in impact evaluation refering to (change) assessments by participants and analysis of documents and data (before-after & *ex-post*)

S	W	O	T
Standardised (online) **surveys** with target groups – complete acquisition	–	–	**Survey instruments must be qualitatively adapted to social, organisational, cognitive context of persons surveyed**
Intensive in-depth/ **structured interviews** with target groups	–	–	**Dito.**
Participant observation (e.g., in status seminars, final presentations)	–	Original view into practice	–

The main and important advantage of the methodology of so-called counterfactual self-estimation (Mueller *et al.*, 2013) is the reference to a counterfactual statement ('What would have been the case, if the cause event had not appeared?') which is not available in a before-after comparison. This advantage is balanced by the restriction of the approach to one's own intentional states and the threats that memorization problems and deficits in self-analysis of intentional states could negatively affect the self-estimation (see Table 8).

Table 8: Methodological SWOTs in impact evaluation refering to counterfactual self-estimation (before-after & *ex-post*)

S	W	O	T
Counterfactual available	**Restriction to own intentional states**	–	Memorisation problems
–	–	–	Deficits in self-analysis of intentional states

An interesting methodological aspect in impact evaluation refering to causal social mechanisms consists on the opportunity to solve the attribution problem by introducing

mechanism hypotheses. Of course, the major threat is then to identify cause-effect mechanisms which do the job. If the approach is successful its major strength will be manifest: the explanation of QA effects by mechanisms instead of law-like relations or statistical correlations. Another, non-negligible threat is to come to grips with the corresponding expenses (see Table 9).

Table 9: Methodological SWOTs in impact evaluation refering to causal social mechanisms (before-after & *ex-post*)

S	W	O	T
Explanation by mechanisms instead of law-like relations or statistical correlations (causal effects of QA measures)	–	**Solve attribution problem by mechanism hypotheses**	**Identification of cause-effect mechanisms, micro-dynamics (high complexity)**
–	–	–	Expenses (time, money)

6.2 IMPALA SWOTs

In addition to the above-mentioned methodological SWOTs which may affect any impact evaluation of QA in HEIs, a few strengths, weaknesses, opportunities and threats can be named which mainly relate to the IMPALA project (IMPALA, 2013-16) and are listed in Table 10. Among the important strengths are: the IMPALA methodology is applicable to any HEI performance area and stakeholder group; IMPALA presents diverse case studies; it impacts perception and assessment of different HEI members; students, teachers, QA staff and leadership are considered; IMPALA contributes to competence development in impact analysis and meta-evaluation in QA agencies and HEIs. The restriction, so far, to four qualitative case studies is accompanied by the opportunity that the approach might be applied to larger samples with larger groups of engaged stakeholders. Among the important methodological weaknesses of the IMPALA project are: a relatively small time window of 36 months for the development, application and analysis of the methodology in relation change processes in HEIs which are relatively slow; a relatively small staff-days budget as compared to the desirable surveying efforts, in particular a sufficient number of structured interviews; missing target group interviews and target group workshops; no metric data (only nominal and ordinal scales). Obviously, several of these weaknesses imply corresponding threats to overcome these, an additional threat being the adequate timing of (E)QA interventions (see Table 10).

In summarizing, it can be stated that impact evaluation of QA in HEIs is indispensable because evidence-based governance and evidence-based QA are necessities in HE which represents a crucial area and innovative power in modern education societies and knowledge economies. It also turns out that impact evaluation cannot be reduced to one puristic strand of methodology such as control group, before-after, and *ex-post* analysis but must rely on a mixed-methods approach which combines the different methodologies which, due to their complementarity, complete each other. Mixed methods will generally also increase the validity and reliability of results. In particular, for impact analyses a mix of questionnaire surveys and structured interviews is required to combine more generic and less exploratory information on the one side with more individual-case and exploratory information on the other side, especially when it comes to causal analyzes. A further general experience is that there are usually no easy solutions to the weaknesses and strengths, i.e., these are unavoidably with us to a certain extent, and HE researchers as well as practitioners have to deal with them discursively. Another general thing is that for impact evaluation of QA in HEIs it is hard to achieve metric data which would be interval scaled. It is also hard, may be even impossible, to do more than case studies because different HEIs in different HEI systems are hardly comparable due to their high complexity and dynamics.

Table 10: Methodological SWOTs in impact evaluation refering to features of the IMPALA project

S	W	O	T
Applicable to any HEI performance area	**Relatively small time window**	–	Extend time window
Applicable to any stakeholder group	**Relatively small staff-days budget**	–	Extend budget
Diverse case studies	Diversity of case studies	–	–
Impact perception and assessment of different HEI members	**Possible biases because of multiple social roles of participants (QA agencies; HEI QA managers; students)**	–	**Adequate timing of (E)QA interventions**
Students, teachers, QA staff, leadership considered	**Restriction of questionnaires** to certain items (unavoidable because of complexity)	–	Extend questionnaires
Four qualitative case studies, to be multiplied	**Low number qualitative case studies (but due to specific QA activities)**	**Apply approach to larger samples** (e.g., large study programs with many engaged/ informed people)	Clarification of relation (E)QA criteria & intended/ desired effects
Contribution to competence development in impact analysis and meta-evaluation in QA agencies and HEIs	**No metric data** (only nominal and ordinal scales)	–	Generate metric data (intervall scales)
–	Missing target group interviews	–	Target group interviews
–	Missing target group workshops	–	Target group workshops
–	–	–	**Expenses** (time, money)
–	–	–	Continuing motivation of partners

Conclusions

A general conclusion for impact studies of QA in HE is that methodological preferences must be related to comparative and longitudinal studies which include assumptions about causal mechanisms and careful survey designs. The methodological overview of impact evaluation of QA interventions in HE shows that the applicable methodology of preference consists of a simultaneous impact analysis with baseline study and before-after comparison. Longitudinal studies should also comprise the practical completeness of the used surveys, meaning that all relevant stakeholder groups, HEI-internal and HEI-external, should be included (e.g., students; academic staff; QA managers; leadership; employers; etc.). Experimental design and with-without comparison design (also known as control group design) will, as a rule, not work in the complex, multivariate and dynamic field of HE.

A main advantage of simultaneous impact analysis based on before-after comparison design and causal social mechanisms is that such approach is not based solely on ex-post surveys which, in real-world situations, are often temporally rather delayed. Furthermore, with online surveys comes the option to include large portions of stakeholder groups – although it must be admitted that it is not always very easy to identify and activate large numbers of stakeholders in the university world of loosely coupled, specific organisations.

Obviously, at present the challenge is that the suggested methodology is applied, tested and possibly improved in case studies. Thus, it can be recommended that impact evaluations as characterised in this manual should be carried out by QA agencies and HEIs by taking the following steps: (1) They should start with a 'thorough analysis of the EQA

procedure one wants to look at: What are its intended goals and what intervention measures are brought to the HEI? What are possible unintended, desirable and undesirable effects of the applied QA measure?' (Leiber *et al.*, 2015, p. 308). (2) They should 'formulate survey questions for various relevant stakeholder groups the answers of which would help to understand how (E)QA affects the HEI' (ibid.). (3) 'If applicable and arguable, they should formulate causal social mechanism hypotheses related to the investigated (E)QA, and try to test them empirically' (ibid.). (4) They should 'carry out document analysis and longitudinal survey studies (e.g., online surveys; structured interviews) simultaneously with (E)QA procedures' (ibid.).

The application of the proposed methodology would improve the following abilities of the concerned stakeholders: understanding the causal networks and effects of QA measures; understanding the connections between QA, quality improvement in various achievement areas, organizational change in HEIs, and establishment of quality culture. These abilities, in turn, 'could well make valuable contributions to further improve QA and quality in HEIs in a strategic and systematic way, methods and instruments of QA agencies and HEI-internal QA, and evidence-based and research-informed HEI governance and HE policy' (ibid.).

Overall summarizing across all methodological considerations, empirical results so far, and SWOTs of impact evaluation of QA in HEIs it can be asserted:

> 'Impact analysis in general, and impact analysis of quality management in HE in particular are very complex social science matters. These are social systems which are highly dynamic open systems and have diverse and complex constituents (e.g., institutional and organisational structures and processes; rather complicated emotional and intentional states of decision-makers; complex processes of decision-making and action-taking). The basic problem of impact analysis is the well-known attribution problem: attributing definite and measurable cause events to observed empirical changes (thus identifying them as effects)' (Leiber, 2016, pp. 11).

References

Astbury, B. & Leeuw, F. L., 2010, Unpacking black boxes: mechanisms and theory building in evaluation, *American Journal of Evaluation*, 31(3), pp. 363-381.

Beerkens, M., 2015, Quality assurance in the political context: In the midst of different expectations and conflicting goals, *Quality in Higher Education*, 21(3), pp. 231-250.

Bejan, S. A., Janatuinen, T., Jurvelin, J., Klöpping, S., Malinen, H., Minke, B. & Vacareanu, R., 2015, Quality assurance and its impact from higher education institutions' perspectives: methodological approaches, experiences and expectations, *Quality in Higher Education*, 21(3), pp. 343-371.

Bennett, E. E. & McWhorter, R. R., 2016, Opening the black box and searching for smoking guns. Process causality in qualitative research, *European Journal of Training and Development*, 40(8/9), pp. 691-718.

Brennan, J., 2012, Talking about quality. The changing uses and impact of quality assurance, pp. 1-11. http://www.qaa.ac.uk/en/Publications/Documents/impact-of-quality-assurance.pdf (access: 13/05/2015).

Coleman, J. S., 1994, *Foundations of Social Theory* (Cambridge, Harvard University Press).

Damian, R., Grifoll, J. & Rigbers, A., 2015, On the role of impact evaluation of quality assurance from the strategic perspective of quality assurance agencies in the European higher education area, *Quality in Higher Education*, 21(3), pp. 251-269.

DeGEval, 2011, *Standards of Evaluation* [in German] (Mainz, Gesellschaft für Evaluation e.V.). http://www.degeval.de/degeval-standards/standards (access: 13/05/2015).

European Association for Quality Assurance in Higehr Education. (ENQA), 2015, *Standards and Guidelines for Quality Assurance in the European Higher Education Area (ESG),* Approved by the Ministerial Conference in May 2015 (Brussels, ENQA). http://www.enqa.eu/wp-content/uploads/2015/05/ESG_endorsed-with-changed-foreword.pdf (access: 17/06/2015).

Gerring, J., 2005, Causation. A unified framework for the social sciences, *Journal of Theoretical Politics,* 17(2), pp. 163-198.

Gerring, J., 2007, The mechanismic worldview: Thinking inside the box, *British Journal of Political Science*, 38, pp. 161-179.

Gerring, J., 2010, Causal mechanisms: yes, but …., *Comparative Political Studies,* 43(11), pp. 1499-1526.

Gross, N., 2009, A pragmatist theory of social mechanisms, *American Sociological Review*, 74, pp. 358-379.

Harvey, L. & Green, D., 1993, Defining quality, *Assessment and Evaluation in Higher Education*, 18(1), pp. 9-34.

Harvey, L. & Williams, J., 2010a, Editorial: fifteen years of Quality in Higher Education. Part One, *Quality in Higher Education*, 16(1), pp. 3-36.

Harvey, L. & Williams, J., 2010b, Editorial: fifteen years of Quality in Higher Education. Part Two, *Quality in Higher Education*, 16(2), pp. 81-113.

Hedström, P., 2005, *Dissecting the Social: On the Principles of Analytical Sociology* (Cambridge, Cambridge University Press).

Hedström, P. & Ylikoski, P., 2010, Causal mechanisms in the social sciences, *Annual Review of Sociology,* 36, pp. 49-67.

Helbing, D. & Balietti, S., 2015, How to do agent-based simulations in the future: from modeling social mechanisms to emergent phenomena and interactive systems design. 55 pages. http://www.researchgate.net/publication/228843530_How_to_Do_Agent-Based_Simulations_in_the_Future_From_Modeling_Social_Mechanisms_to_Emergent_Phenomena_and_Interactive_Systems_Design (access: 17/06/2015).

IMPALA, 2013-16, Impact analysis of external quality assurance processes of higher education institutions. Pluralistic methodology and application of a formative transdisciplinary impact evaluation. https://www.evalag.de/impala / http://www.impala-qa.eu (access: 04/11/2016).

Kajaste, M., Prades Nebot, A. & Scheuthle, H., 2015, Impact evaluation from quality assurance agencies' pPerspectives: Methodological approaches, experiences and expectations, *Quality in Higher Education*, 21(3), pp. 270-287.

Leiber, T., 2014a, Zur Methodologie der Wirkungsevaluation von Qualitätssicherung an Hochschulen, in Benz, W., Kohler, J. & Landfried, K. (Eds.), *Handbuch Qualität in Studium und Lehre*. Ausgabe 46(3), E 7.13, pp. 41-80 (Berlin, Raabe).

Leiber, T., 2014b, Evaluation of the success or failure of quality assurance in higher education institutions: methodology and design, *Journal of the European Higher Education Area*, 2, pp. 39-74.

Leiber, T., 2016, Impact evaluation of quality management in higher education. A contribution to sustainable quality development of the knowledge and learning society, *Qualität in der Wissenschaft*, 10(1), pp. 3-12.

Leiber, T., 2017, Computational social science and Big Data: a quick SWOT analysis, in Wernecke, J., Pietsch, W. & Ott, M. (Eds.), *Berechenbarkeit der Welt? Philosophie und Wissenschaft im Zeitalter von Big Data*, pp. 287-301 (Berlin, Springer).

Leiber, T., Stensaker, B. & Harvey, L., 2015, Impact evaluation of quality assurance in higher education: Methodology and causal designs, *Quality in Higher Education*, 21(3), pp. 288-311.

Lillis, D., 2012, Systematically evaluating the effectiveness of quality assurance programmes in leading to improvements in institutional performance, *Quality in Higher Education*, 18(1), pp. 59-73.

Little, D., 2011, Causal mechanisms in the social realm, in McKay Illari, P., Russo, F. & Williamson, J. (Eds.), *Causality in the Sciences*, pp. 273-295 (Oxford, Oxford University Press).

Little, D., 2015a, Mechanism and method, *Philosophy of the Social Sciences*, pp. 1-19. http://pos.sagepub.com/content/early/2015/04/16/0048393115580420.full.pdf+html (access: 15/05/2015).

Little, D., 2015b, Guiding and modeling quality improvement in higher education institutions, *Quality in Higher Education*, 21(3), pp. 312-327.

Mark, M. M. & Henry, G. T., 2004, The mechanisms and outcomes of evaluation influence, *Evaluation*, 10(1), pp. 35-57.

Mueller, C. E., Gaus, H. & Rech, J., 2013, The counterfactual self-estimation of program participants: Impact assessment without control groups or pretests, *American Journal of Evaluation*, published online October 03, 2013, pp. 1-18.

Newton, J., 2013, Is quality assurance leading to enhancement?, in Crozier, F. et al. (Eds.), *How Does Quality Assurance Make a Difference?*, pp. 8-14 (Brussels, European University Association).

Phillips, D. C. & Burbules, N. C., 2000, *Postpositivism and Educational Research* (Lanham, Rowman & Littlefield Publishers).

Rosa, M. J., Sarrico, C. S. & Amaral, A., 2012, Academics' perceptions on the purposes of quality assessment, *Quality in Higher Education*, 18(3), pp. 349-366.

Shah, M., 2012, Ten years of external quality audit in Australia: evaluating its effectiveness and success, *Assessment and Evaluation in Higher Education*, 37(6), pp. 761-772.

Steel, D., 2004, Social mechanisms and causal inference, *Philosophy of the Social Sciences*, 34(1), pp. 55-78.

Steel, D., 2011, Causality, causal models, and social mechanisms, in Jarvie, I. C. & Zamora-Bonilla, J. (Eds.), *The SAGE Handbook of the Philosophy of Social Sciences*, pp. 288-304 (Thousand Oaks, Sage Publications).

Stensaker, B., 2008, Outcomes of quality assurance: A discussion of knowledge, methodology and validity, *Quality in Higher Education,* 14(1), pp. 3-13.

Stensaker, B., Langfeldt, L., Harvey, L., Huisman, J. & Westerheijden, D., 2011, An in-depth study on the impact of external quality assurance, *Assessment and Evaluation in Higher Education,* 36(4), pp. 465-478.

Stensaker, B. & Leiber, T., 2015, Assessing the organisational impact of external quality assurance: Hypothesising key dimensions and mechanisms, *Quality in Higher Education*, 21(3), pp. 328-342.

Suchanek, J., Pietzonka, M., Künzel, R. & Futterer, T., 2012, The impact of accreditation on the reform of study programmes in Germany, *Higher Education Management and Policy*, 24(1), pp. 1-24.

Volkwein, J. F., Lattuca, L. R., Harper, B. J. & Domingo, R. J., 2007, Measuring the impact of professional accreditation on student experiences and learning outcomes, *Research in Higher Education*, 48(2), pp. 251-282.

Westerheijden, D. F., Hulpiau, V. & Waeyten, K., 2007, From design and implementation to impact of quality assurance: An overview of some studies into what impacts improvement, *Tertiary Education and Management*, 13(4), pp. 295-312.

White, H., 2009, Theory-based impact evaluation: Principles and practice, *Journal of Development Effectiveness*, 1(3), pp. 271-284.

White, H., 2010, A contribution to current debates in impact evaluation, *Evaluation*, 16(2), pp. 153-164.

Acknowledgement

The work on this manual was carried out in the context of the project “Impact Analysis of External Quality Assurance Processes of Higher Education Institutions. Pluralistic Methodology and Application of a Formative Transdisciplinary Impact Evaluation” (acronym: IMPALA; http://www.evalag.de/impala / http://www.impala-qa.eu), which was co-funded by the European Commission (Grant no. 539481-LLP-1-2013-1-DE-ERASMUS-EIGF) between 2013 and 2016. The IMPALA consortium would like to thank the European Commission for valuable financial support which made the IMPALA project possible. The consortium partners would also like to express their gratitude to all engaged experts, in particular (in alphabetical order): Maarja Beerkens (Leiden University), John Brennan (Open University), Lee Harvey (Copenhagen Business School), Daniel Little (University of Michigan-Dearborn), Barbara Michalk (German Rectors’ Conference), Philipp Pohlenz (University of Magdeburg), Markus Seyfried (University of Potsdam), Bjørn Stensaker (University of Oslo), Jussi Välimaa (University of Jyväskylä). They enriched the project by their critical analyses, contributions in internal meetings and presentations at IMPALA conferences in Mannheim (May 2014) and in Barcelona (June 2016) and at an IMPALA Workshop in Mannheim (September 2016). The experts also added a lot to the exploitation and dissemination of the project’s results by various papers.

This publication reflects the views only of the IMPALA project partners, and the European Commission cannot be held responsible for any use which may be made of the information contained therein.

Appendix: An Exemplary Survey Questionnaire

Introductory Remarks

In the following, an excerpt from a generic core questionnaire and a study case-specific questionnaire for surveying impacts of QA in learning and teaching in HEIs is presented.

Concerning the construction of such a questionnaire some remarks are in order because there are certain serious limitations to the genericity strived for. Of course, in general it would be very much desirable to have a generic questionnaire which captures diverse performance areas and various stakeholder groups within a single, maximally integrative methodological approach. Due to the complexity and diversity of HEIs and HE systems in general, it turns out, however, that it is practically impossible to formulate transparent and content-wise adequate questionnaires across different performance areas and the various stakeholder groups. Therefore, the pragmatic decision seems to be unavoidable that a concrete survey questionnaire, first of all, must be focused on a specific stakeholder group (such as students; teachers; QA staff; HEI leadership members; etc., see Table 11)[25] and on a specific achievement area of HEIs (such as learning and teaching; research; Third Mission; etc., see Table 12), because otherwise there is a real danger that the procedure will be unmanageable in terms of the amount of efforts of those being evaluated and the evaluators; the number and available competencies of involved peers; the available and meaningful period of time over which the observed quality process extends.

Table 11: Stakeholder groups in QA in HE, adopted from (Leiber, 2016, p. 5)

Stakeholder groups
Students
Academic staff in teaching and research
Peers
Employers
Quality assurance agencies
Study program managers
Higher education institution managers
Governments
Society
International community

Table 12: Five main areas of QA influence in HE, adopted from (Leiber, 2016, p. 4)

Main areas of influence
TEACHING AND LEARNING
RESEARCH
THIRD MISSION
Internationalisation of higher education
Inter- and transdisciplinarity of higher education
INSTITUTIONAL MANAGEMENT
NATIONAL HIGHER EDUCATION AND QUALITY ASSURANCE SYSTEM
Satisfaction with quality assurance processes

Moreover, in practice it turns out that even within a focused performance area further focusing is unavoidable. This is due to the complexity of the fields as well as the una-

[25] Fortunately, it seems to be a practice experience that there are reasonable similarities and matches between questionnaire questions relevant for different stakeholder groups so that the various questionnaires will not be entirely different.

voidable focuses and also intrinsic limitations of QA procedures applied and the necessity to generate questionnaires which are readable and workable by the surveyees within reasonable periods of time.[26]

For sake of brevity and because of similarities between questionnaires for different stakeholder groups, in this Appendix excerpts of exemplary generic core questionnaires for impact analysis of QA in learning and teaching for students and teachers are given. In addition, an example of a non-generic, i.e. study case-specific questionnaire part is presented. For sake of brevity and readability, the questionnaires are documented without question filters (which are, of course, indispensable in practice).

When having a closer look at the questionnaires below, in view of the variety of options quite naturally the question arises what the rationale is behind the specific questions included in the questionnaire. The following considerations and reasons have been taken into account when constructing the questionnaires:

1) First of all, questions should ask for intended aims of certain QA procedures in learning and teaching[27] (such as external program accreditation; instutional accreditation; audit of quality management system).
2) The questionnaire should also ask for unintended desirable and unintended undesirable outcomes. Of course, the corresponding items must be first identified (for example, either by previous exploratory investigations or by open questions of the questionnaire).
3) Questions should take into account the contemporary move from teaching to learning (e.g., student-centered teaching; alignment of learning objectives and learning outcome assessments).
4) In some relevant cases surveyees should be asked whether they have seen changes of certain features in the foregoing time, which changes they have seen and who or what initiated these changes. – Such questions seem to offer possibilities for identifying correlations and causal social mechanisms.
5) Questions should be easily understandable by the members of the addressed stakeholder group.
6) Questions that could be well or better answered by (structured) interviews and/or documentary analysis should be excluded from the questionnaires and treated by the alternative methods.
7) Surveyees should be asked for their suggestions for improvement of QA procedures.
8) Surveyees should have options for giving additional comments.

Excerpts from a generic core questionnaire

Information on Survey Participants

To which stakeholder group do you belong?

Please choose one of the following answers:

o Students; o Teachers; o QA Managers; o Leadership

[26] Usually not much more than ca. 20 minutes.
[27] Other achievement areas are excluded for pragmatic reasons only.

For which study program did you contribute to the self-assessment report?

Please choose all relevant answers:

o [*insert name of study program* n_1]; o [*insert name of study program* n_2];

o [*insert name of study program* n_3]; o [*etc.*]; o None

Are you a member of the study commission of your study program?

Please choose one of the following answers:

o Yes; o No

General questions about QA in learning and teaching

QUESTIONS for STUDENTS

What kinds of courses are used in your study program?

Please choose the relevant answer for each item:

	Mainly	Partially	Not at all
Frontal teaching (e.g., lectures)	O	O	O
Interactive courses (e.g., seminars based on students' presentation; group work)	O	O	O
Courses with practice-related elements (e.g., internships in HEIs; internhips in business companies)	O	O	O
Project-based courses (e.g., project seminars)	O	O	O
Online courses	O	O	O

Have you in the last year seen changes in the used kinds of courses?

Please choose one of the following answers:

o Yes; o No

Which changes in the last year have you seen in the used kinds of courses?

Please choose the relevant answer for each item:

	Increase of type of courses in curriculum	Decrease of type of courses in curriculum	No change
Frontal teaching	O	O	O
Interactive courses	O	O	O
Courses with practice-related elements	O	O	O
Project-based courses	O	O	O
Online courses	O	O	O

Who or what initiated these changes?

	Mainly	Partially	Not at all
Initiatives of students	O	O	O
Initiatives of teaching staff	O	O	O
Initiatives of HEI management	O	O	O
External QA (e.g., accreditation)	O	O	O
Initiatives of HEI management	O	O	O
Legal requirements	O	O	O
External stakeholders	O	O	O
Other, please give details here:			

Which instruments of QA are used in your study program?

	Regularly	Occasionally	Never
Survey of students	O	O	O
Survey of teaching staff	O	O	O
Survey of employers	O	O	O
Meetings with focus QA and quality development (e.g., Study Commission; faculty councils)	O	O	O
Observation of performance indicators	O	O	O
Written reports	O	O	O
Other, please give details here:			

How do the examinations relate to the learning objectives in the module handbook of your study program?

o Module examinations largely reflect the learning objectives of the modules;

o Module examinations partially reflect the learning objectives of the modules;

o Module examinations little reflect the learning objectives of the modules

In general, what is your attitude towards QA and quality development in learning and teaching?

o Positive; o Neutral; o Negative

Has your attitude towards QA and quality development in learning and teaching changed in the last year?

o Yes, in a positive direction (more approval);

o Yes, in a negative direction (less approval);

o No, no change in my attitude.

What has changed your attitude towards QA and quality development?

	Mainly	Partially	Not at all
Experience with internal procedures of QA (e.g., course surveys; module evaluations etc.)	O	O	O
Experience with external procedures of QA (e.g., writing of self-evaluation reports; on-site assessments etc. in accreditations)	O	O	O
Taking notice of peer reports	O	O	O
Other, please give details here:			

Do the procedures of QA and quality development in learning and teaching (e.g., course evaluations, module evaluations, student surveys, etc.), which are carried out in your HEI, have effects which are observable for you?

o Yes; o No

In your view, have these effects changed in the last year?

o Yes, in a positive direction (increasing effectivity);

o Yes, in a negative direction (decreasing effectivity);

o No, no change

Do you have suggestions for improvement of procedures of QA and quality development at the *[insert name of surveyed HEI]*?

Additional comments?

Please add additional comments.

QUESTIONS for TEACHERS

How do the examinations relate to the learning outcomes in the module handbook of your study program?

Please choose one of the following answers: see *QUESTIONS for STUDENTS* above

Has the relation between examinations and learning objectives in the module handbook changed in the last year?

Please choose one of the following answers:

o Yes, the examinations are now stronger oriented on learning objectives;

o Yes, the examinations are now weaker oriented on learning objectives;

o No, no changes

Who or what initiated the change?

Please choose the relevant answer for each item: see *QUESTIONS for STUDENTS* above

How often do teachers of your study program meet in order to discuss the further development of the study program?

o At least once every three months; o At least once a year; o Less than once a year

Have you in the last year seen a change with respect to the frequency of teachers' meetings for further developing the study program?

o Yes, the meetings became more frequent; o Yes, the meetings became less frequent;

o No, no changes

Who or what initiated the change?

Please choose the relevant answer for each item: – see above –

In general, what is your attitude towards external QA and quality development in learning and teaching? (External QA is carried out by HEI-external agencies, e.g., program or institutional accreditation and evaluations.)

o Positive; o Neutral; o Negative

Has your attitude towards external QA and quality development in learning and teaching changed in the last year?

o Yes, in a positive direction (more approval);

o Yes, in a negative direction (less approval);

o No, no change in my attitude.

What has changed your attitude towards external QA?

Please choose the relevant answer for each item: – see above –

How do you assess the expenditure and benefit of QA and quality development?

	Very low	Low	High	Very high
Expenditure	O	O	O	O
Benefit	O	O	O	O

How do you assess the attitude of the HEI leadership towards QA in learning and teaching?

o HEI leadership actively supports QA; o HEI leadership is affirmative against QA;

o HEI leadership is neutral against QA; o HEI leadership is unfavorable against QA

Have you in the last year seen a change in the attitude of HEI leadership against QA?

o Yes, more support; o Yes, less support; No change

Do you currently plan major changes in your study program?

o Yes; o No

Which changes are planned?

Please choose all relevant answers:

o Optimisation of equipment (financial, material, personell) of learning and teaching;

o Introduction/ modification of student admission procedures;

o Check of the adjustment of workload and ECTS credit points;

o Revision of the definition of learning objectives;

o Adjustment/ alignment of examination formats and learning objectives;

o Optimisation of modules;

o Personel development measures (e.g., further education offers for teachers);

o Changes in view of Employability;

o Introduction of instruments and processes of QA;

o Other, please give details here:

Do you have suggestions for improvement of procedures of QA and quality development at the *[insert name of surveyed HEI]*?

Additional comments?

Please add additional comments.

Further generic questionnaire questions used in the IMPALA project can be requested from the project coordinator at leiber@evalag.de.

Excerpts from a study case-specific questionnaire

The following questionnaire items are adapted to a specifc QA procedure at a certain university.

QUESTIONS for STUDENTS and TEACHERS

How did you experience the handling of the study program report?

Please choose the relevant answer for each item:

	Applies	Rather applies	Rather does not apply	Does not apply	No answer/ Do not know
There were constructive discussions in the Study Commission about the further development of the study program during the preparation of the study program report.	O	O	O	O	O
The students of the Study Commission were intensively involved in the preparation of the study program report.	O	O	O	O	O
The strengths and weaknesses of the study program are well described in the study program report.	O	O	O	O	O
Because of the forthcoming study program review the study program was reflected form a different perspective.	O	O	O	O	O
In the context of the preparation of the study program report the forthcoming study program review was picked out as a central theme.	O	O	O	O	O
As a peer I could gain a comprehensive picture about the strengths and weaknesses of the study program on the basis of the study program report.	O	O	O	O	O
The study program review will contribute to the further development of the study program.	O	O	O	O	O
I am curious what will be written about my study program in the external peer report.	O	O	O	O	O

How did you experience the handling of Study Program Review results in the Study Commission?

Please choose the relevant answer for each item:

	Applies	Rather applies	Rather does not apply	Does not apply	No answer/ Do not know
The external peer report was discussed in a study commission meeting.	O	O	O	O	O
The external peer report was discussed in a study commission meeting.	O	O	O	O	O

How did you experience the handling of the external peer review in the Study Commission?

	Applies	Rather applies	Rather does not apply	Does not apply	No answer/ Do not know
The external peer report triggered discussions about the study program.	O	O	O	O	O
Because of the external peer report weaknesses of the study program were identified.	O	O	O	O	O
Because of the external peer report improvement needs of the study program were identified.	O	O	O	O	O
The external peer report confirmed strenths of the study program.	O	O	O	O	O
Because of the external peer report it became obvious that the study program is on the right path.	O	O	O	O	O

What do you expect from the recommendations of the Program Review Commission?

	Applies	Rather applies	Rather does not apply	Does not apply	No answer/ Do not know
The recommendations of the Review Commission concerning technical and content-wise aspects will contribute to the further improvement of the study program.	O	O	O	O	O
The recommendations of the Review Commission concerning formal and organisational aspects will contribute to the further improvement of the study program.	O	O	O	O	O
The conclusion of the Review Commission could be well understood.	O	O	O	O	O

The study program review made new aspects of the study program visible to me.

Please choose one of the following answers:

O Yes: O No; O No answer

In case of "Yes", please choose the relevant answer for each item:

	Yes	No	No answer/ Do not know
The study program review made strengths visible.	O	O	O
The study program review made weaknesses visible.	O	O	O

Through the review of the study program, my assessment of the study program has changed positively.

Please choose one of the following answers:

O Applies: O Rather applies; O Rather does not apply; O Does not apply;

O No answer/Do not know

Do you have suggestions for improvement of procedures of QA and quality development at the *[insert name of surveyed HEI]*?

Additional comments?

Please add additional comments.