Chapter 17
Uniformity is No Virtue

Ekkehard Kappler

On ambivalence in university evaluation and accreditation.

Revealing the main point of an essay in the title itself is as ill-advised as naming the culprit on the first page of a detective story. But perhaps this is a case where the exception proves the rule. In a competitive society, uniformity is not a virtue. And it is especially dangerous when an exceptional product becomes the subject of inflationary exploitation. Just think of the more than 2,500 MBA programs that now exist. Regardless of their quality, they all take advantage of the highly positive image of the first MBA programs. The column of lemmings heads for the abyss. The problem that has developed in the last few decades from the spread of evaluation and accreditation for universities is in fact the danger of uniformity prescribed and implemented for quality assurance reasons in the higher educational sector in Europe. That would be the exact opposite of competition between universities as a key to improved performance and the opposite of a reform that—so we are told—is designed to replace input control for universities by output control with the aim of increasing the production of ideas and enhancing the quality of the university.

17.1 Retrospect and Point of Departure

It is not so long ago that grades—with the necessary reservations—were seen as a certain measure of student performance. A record of good grades over a longer period of time for the subjects tested has always had a certain currency, and a history of negative grades is definitely not an advantage. The fact that Thomas

E. Kappler (✉)
Department of Organization and Learning, Innsbruck University,
Elisabethstraße 12, 6020, Innsbruck, Austria
e-mail: ekkehard.kappler@uibk.ac.at

A. Altmann and B. Ebersberger (eds.), *Universities in Change*, Innovation, Technology, and Knowledge Management, DOI: 10.1007/978-1-4614-4590-6_17,

Mann was awarded a 'fail' in German and the eminent German surgeon Ferdinand Sauerbruch took home a disastrous school record for the fourth year does not disprove this point. But when people started to realize or at least suspect that grades and school examinations were a key factor in determining the students' futures, the picture began to change. While opportunities in life and the process of selection are based on grades, there is increased pressure, not only on the students in the schools and universities, but also on the teaching staff and faculty. As a father put it to a teacher he was on friendly terms with, 'It's got to be at least a 'B' for my girl in your subject!'—as if grades were something to be negotiated like a pay rise.

The new perception of educational opportunities and the general promotion of education ushered in the heyday of middle-class society. But the greater the awareness became the significance of education, the more the conventional grading system was eroded, with good grades becoming an inflationary product in many cases. Parents became 'trade unionists', representing their children's interests in the schools. And at times of big student populations, the same can be said of the universities. The hardliners continued to defend traditional standards, but not all faculty members are hardliners. And in the long term, a rearguard action is neither a sensible nor a creative strategy. That makes it increasingly difficult to interpret the grades awarded at university.

The rapid increase in student numbers naturally led to an increase in the number of examinations. That triggered several phenomena. Although many personnel managers claim that examination results are not decisive in the case of a first-time application, they are not irrelevant, either—a fact that has led not only to a race for good grades but also to certain bargaining scenarios and practices. Many an overworked university teacher would have found it difficult to maintain their standards over time. 'If students fail, I have even more of them begging for grades in my overcrowded office hours and even more candidates at the next session, which is already full to bursting!' As long as university teachers were not evaluated in any way themselves, they could do what they wanted. At least, that was a widespread public belief, and that belief was reinforced by the not infrequent cases of malpractice unduly generalized.

It is naturally good that university teachers should also be assessed. Normally, they have demonstrated their competence in their respective fields through their various works and publications and in the course of a long appointment process. But even today, the situation is different with regard to their didactic abilities; often enough they do not rise beyond amateur status. And most university examiners do not know the meaning of the term 'examination didactics', as the examinations they set show. In fact, many oral examinations are not far from mental assault. But examiner behavior is not evaluated. It is now general practice to try to assess the quality of the teaching with the help of anonymous questionnaires filled in by the students. That admittedly covers only one standpoint and one perspective, but is an important one. However, it rarely guarantees that the teachers will make sustainable changes in response to the results of the evaluation.

However, clear the need for teacher assessment may be, the methods applied come nowhere near to solving the problems inherent in them. In the big degree programs, the great majority of the teachers are suffering from acute overload in the teaching situation and are trying to make the best of it. Their professional ethos prevents them from kicking over the traces, but appeals to professional ethos are simply a veiled form of exploitation.

There is one negative observation that can be made, however, where inappropriate incentives cause individual examiners to adopt an approach that does not lead to enhanced levels of performance. Offering examiners a certain sum of money per candidate, for example, may cause them to take advantage of examinations to increase their total income. Of course, that is intended, and examinations with large number of candidates are handled as a result. On the other hand, some examiners have gone too far in their desire to generate extra income. When one examination suddenly becomes five without increasing the demands made of the examinees, the incentive is misguided.

In countries with a state university system in which a credit obtained at one university is automatically recognized at all the other state universities, a kind of 'exam tourism' sets in. Coaches are chartered to take students to a university where the examination in a certain subject is a 15-minute oral compared with the 4-hour written examination to be sat at the university the coaches come from. Word quickly spreads as to where the examinations in a given subject are easy and where they are difficult. The students returning to their home university from the easier examination with the better grades in their pockets not only contribute to the examiner's income (and to the economy via their travel expenditures); with their better grades they also increase their chances of receiving a grant to study abroad and other forms of financial assistance.

The increase in the number of students studying abroad—however positive that may be—has also created a number of problems. The differences between the various school and university systems in Europe and also in the USA and all the other countries raise questions of credit recognition, for which answers can be found but without solving the problems at the level of content. Students from aboard sometimes have language problems and sometimes are admired because they do not. That can influence the final grade in either case. Here again, there is a latitude that is difficult to quantify.

Evaluation in the field of research is not without its temptations, either, as in the case of the faculty that improved its results by placing those teachers who had not published anything in the assessment period on the payroll for non-academic staff.

One might say that these cases are excesses of the system, abusive interpretations of the rules, carelessness or even punishable offences, and that they do not offer a true picture of the university. In many cases that would be true, although one probably has to be an insider to see it. The public perception, however, is mainly dominated by reports about problems and poor results in the rankings, for example at the international level. It is this perception that leads to calls for uniform evaluation and accreditation systems. And the more populist the arguments, the greater their effectiveness at election time, in particular when combined

with calls for more supervision and tighter controls—a fatal equation, which merely shows that, apart from an authoritarian attitude, the critics are ignorant of the true tasks and techniques of organizational control. The general public, for its part, is not normally capable of reflection on and criticism of the dubious nature—in terms of the theory of numbers and measurement theory—of such demands and the proposed standards of comparison. On the contrary, people tend to assume that the comparisons, based as they are on numbers, will be valid. That is why uniformity is desired or assumed in the context of evaluation. But numbers and figures are not objective. They, too, 'only' tell a story with a meaning that is far from being unambiguous and with vested interests that are not communicated.

In the view of many university teachers, state regulatory measures have proved incapable of guaranteeing the uniformity of the measurement standard, and that has led to calls for other control mechanisms. Attempts have been made to transfer the successful results achieved with the rules of the market economy to the public service sector, but the supposed autonomy that could have led to much greater variety among the universities was hijacked and restricted by the predefined evaluation criteria. Ultimately, that is understandable as state universities are still unable to generate the income needed to finance themselves; as in the past, they are dependent on a big budget provided by the state. If the universities were to charge the fees needed to cover their costs, open university admission would no longer be available to most of today's students, and a guaranteed contribution from the national budget would be essential. And university research could not be financed through student fees in any case. Nevertheless, we still hear the myth of the superior performance achieved by private universities that receive no money from the state. If we take a look at the performance of the universities around the world, we quickly see how unfounded this myth is—quite apart from the measurability myth.

A true story from the world of trade and industry and management consulting reveals similar problems of quantification. A German city is the location of the research center of an international corporation. The Executive Board decided to take a look at the performance of the research center and appointed a world-famous company of management consultants to perform an evaluation. According to the CEO, the consultants' oral report began as follows: 'I can only say with any precision what comes out of the research center every day at 5 p.m.—4,000 employees.' Ar a university can much probably be measured more easily than that: the essays and books produced by faculty, the hours of classes held, the graduates produced, the ratio of graduates to beginners, the students who switch programs, the dropouts, the actual and average length of study, the available places for work and reading relative to the number of students, the maximum occupancy of the lecture halls and classrooms, etc. More difficulty would be encountered with the number of patents registered, hours worked in secondary employment, work performed for the university, the quality of the teaching and examination work, etc. And real problems arise when people think they can measure the quality of scientific work and publications by classifying the journals in which they appear. The journals with the biggest circulations are the mainstream journals. But the mainstream is not normally the place for creativity, imagination, and innovation.

The dominant paradigm is first taken on board until there is no room left. Articles are expected to have 'connectivity'. And a lack of 'connectivity' is often cited as the reason for rejecting a 'deviant' manuscript, without realizing that criticism is always directed at the object of its passion, whereas the mainstream can largely be satisfied through assimilation and comparatively boring texts. Thus, the peer review is always in danger of degenerating into a self-defense system for the establishment.

The buzzword of university reform in the 1980 and 1990s was output orientation. The aim was to measure what the universities produce and to measure efficiency, i.e., the relationship between input (e.g. budget) and output. And it soon becomes clear that measuring output is not meaningful if it is not related to input. At university, the input traditionally comprised the process of qualification as a scholar (doctorate and post-doctoral qualifications) and the appointment process, on the one hand, and the allocation of government funds on the other. Without reference to both the qualifications and the funding, the outcome is uncertain. As Wilhelm von Humboldt put it in 1810, 'Among the professors there may be many who are good teachers but do not further science; that is perhaps only achieved by a few. But in order to find them, the others must be accepted, too'. This statement is still valid today. For a university that is mainly responsible for providing teaching it may even be sufficient, although the teachers really belong in school, i.e., where the pupils are (Humboldt 1982, p. 256).

At university we only find adults of various ages who together produce the university. That is not easy, but it would still be worth the attempt today. 'We can achieve what is called thought when we think ourselves (…)' (Heidegger 1992, p. 3). But also: 'Only when we want what is in itself though-worthy do we have the power for thought' (Heidegger 1992, p. 3).

That is the point of departure. The university as a loosely coupled system is of necessity a system in which its members are free, freer at least than many other salary earners. University teachers are not badly paid and for many years had tenure (and thus economic independence) and were on the whole free to make use of their time as they saw fit. For many observers, however, they produced too little. But that is a verdict that derives from a concept of closely coupled systems and clear causal chains. In a loosely coupled system with causal chains that are not always clear, and with complex, paradoxical, circular, and mutually dependent relationships, the output does not follow automatically from the input. That is the price to the paid for the chance and for the potential to generate innovation. Loosely coupled systems are more flexible, more adaptable than closely coupled ones, less risky, more imaginative, more creative. They permit a wide variety of potential to develop at the same time. As the positive effects of such opportunities for development can be widely dispersed in time and place, efficiency measurement based on closely coupled causal chains may generate false results for loosely coupled systems. Today's 'crazy guy' may become a famous innovator the day after tomorrow and beyond.

According to Kant, the freedom of loose coupling is the key to the required critical discourse between what he called the lower faculties (especially

philosophy) and the norm-setting role of the higher faculties (theology, law, medicine) (Kant 1968). In the eyes of Kant, the conflict of the faculties was not war but antagonism, i.e., a "conflict of two parties united in (their striving toward) one and the same final end (Concordia discors, discordia concors)" (Kant 1968, p. 35). This "Conflict of the Faculties" is a conflict on the basis of necessary change and difference. The uniformity of the norm setting must be overcome again and again in the interest of the required diversity of practical developments. And the "one and the same final end" is only uniform in this abstract formulation. What it means in concrete terms is itself a subject of scientific discourse and must always remain so. This requirement in no way clearly determines the modes of action, behavior, and decision making at the university and thus cannot define the input adequately either. In view of the multifaceted character of scientific enquiry and historical developments, the "one and the same final end" calls first of all for a leap of faith. That it can be abused cannot be denied. But if there is no leap of faith for the management of the university as a loosely coupled system, the development of the university's potential will be arrested by the nets of mistrustful evaluation. The fact that there has been exploitation of the freedom granted cannot be denied, either, as more than the above examples show. Such facts are quoted as arguments for extensive evaluation and accreditation processes, and also for other restrictive measures including the abolition of tenure (with a consequent increase in opportunism, because who will be so stubborn as to oppose the evaluation criteria if that could cost them their job?!). The argument is logical, but does that mean it is proven? Does the output (!) of the measures selected for evaluation and accreditation follow from the input involved?

The buzzword of university reform in the 1980s and 1990s was output orientation. The aim was to measure what the universities produce and also to measure efficiency, i.e., the relationship between input (e.g. budget) and output. But the measurands are not predefined and cannot be taken for granted. All forms of measurement are dependent on the calibration involved, i.e., on a system of created and defined quanta. Once the system has been established, it gradually loses the character of an artifact, and the convention that has been agreed becomes a truth. The trick of the rule is that it relegates its own genesis to oblivion.

"Much has to be done to render diverse phenomena countable quanta in the first place, namely an abstract from many specific qualities by establishing categories of similarity. Measurement is based on classification systems that ignore 'messential' differences and reduce complexity. Once accepted, such systems appear natural and incontestable (Bowker and Star 2000)." (Power 2004, p. 767).

In the case of the established measurands and ratios, their historical dimension is long since forgotten. We have no measure for innovation, for example. At best we could fall back on the difference compared with what existed before, but that guarantees neither a practicable process nor a meaningful result. Things new derive not infrequently from what has been overlooked, discarded, ignored, and defamed, from what is considered worthless, from waste and dirt (Groys 1992). Let us consult Michael Power again: "Today we regard qualities such as weight, hardness and temperature as obviously quantifiable because of the existence of

trusted technologies for doing this. But there was a time when the ambition to measure heat was regarded as no more different in principle from measurement of virtue: '…if you can manage to think of measuring heat before the invention of the thermometer, then why should we presumptively exclude certitude, virtue and grace?' (Crosby 1997, p. 14).'(Power 2004, p. 768).

The consequences of this thinking are obvious enough. We do not measure things, nor do we represent the 'truth' or the 'facts', but rather we signify. We establish a relationship in a certain way. In doing so, we willy-nilly exclude other modes of presentation, which may naturally relate to other interests. To clarify this point, let us look at what Victor Burgin[1] has to say about the relationship between image and reality.

[The expression 'measurement', outside of a strictly technological application, may need some explanation. What I am proposing as the object of measurement is not restricted to measuring or other ratios, considered as a set of techniques (although, certainly, technique is to be accounted for within the theory); it is, rather, measuring considered as a practice of signification. By 'practice' here is meant work on specific materials, within a specific social and historical context, and for specific purposes. The emphasis on 'signification' derives from the fact that the primary feature of measuring, considered as omnipresence in everyday social life, is its contribution to the production and dissemination of meaning. To argue that the specificity of the object to be constituted in measurement theory is semiotic is not to restrict the theory to the categories of 'classic' semiotics. Although semiotics is necessary to the proposed theory, it is not (nor would it ever claim to be) sufficient to account for the complex articulations of the moments of institution, text, distribution and consumption of ratios, ledgers, calculations, balance sheets, management information systems, systems of knowledge management or balanced scorecards. Confronted as it is with such heterogeneity, it is clear that measurement theory must be 'inter-disciplinary'; there can, however, be no question of simply juxtaposing one pre-existing discipline with another.

For example, at the moment perhaps the least developed aspect of the emerging theory is the sociological component. Measuring is most commonly encountered in sociological texts as 'evidence', the sociologist operating with the common-sense intuition of measuring as a 'window on the university or other organizations'. This type of sociological encounter with measuring is quite simply irrelevant to the project of measurement theory, which must take into account the determinations exerted by the means of representation upon that which is represented. More pertinent is the sociological description of managerial corporations and institutions. Here again, however, the criterion of relevance applies: a description of, say, the hierarchical structures of command governing the accountants in the field

[1] The basic source for the section between [] is the article 'Thinking Photography' (Burgin 1982)—with friendly agreement to any use of his paper by Victor Burgin to whom I have to say many thanks. I transformed his article in a radical interdisciplinary change from the conceptualization of representation by photography to representation by measuring. Of course, I am responsible for any misunderstandings and misinterpretations.

would be less relevant to the theory than a description of the discourses by which the institution inducts its functionaries, irrespective of rank, into a common belief system, constituting them as 'measuring people'. Certainly, we may expect structures of decision-making to be imbricated within beliefs, but it is the beliefs which are the 'sharp end' of that which informs the social effects of measuring. (Nor is this to suggest that accountants' beliefs are simply 'communicated' to their audiences.)

Measuring theory is not exempt from the call made upon any theory to identify observable systematic regularities in its object which will support general propositions about the object. This is already to establish that theory may be taught, and certainly the elaboration of measuring theory constitutes an intervention, at least in principle, in the field of education. In speaking of education in measuring we should distinguish between two quite different pedagogic practices. In the first, a vocational training is given for some particular staff of industry and/or commerce—as when a school trains people to become accountants or managers. In this type of course academic studies will tend to be pragmatic—their content being determined by its practical bearing on the specific form of measuring being taught. It is teaching and training conventions. In the second type of course no particular vocational training is imposed; the student is asked, rather, to consider measuring in its totality as a general cultural phenomenon, and to develop his or her own ideas as to what direction to pursue. Academic studies in the context of this latter type of course are presented as heuristic—aiming to provide the student with a wide range of facts, and a number of critical tools, in the interests of developing an informed capacity for independent thought. Contrary to their declared intent, the majority of those courses whose concern is with measuring as art belong in the first category rather than the second. They offer a vocational training for that staffs of industry, commerce and educational institutions whose products are plans, ratios, budgets, objectives, calculations. The academic content of such courses tends overwhelmingly to take the form of an uncritical initiation into the dominant beliefs and values prevailing in the economic institution as a whole or in the economy of the institutions as a whole. On such courses 'criticism' and 'history' stand in place of a theory that is reconfirming practice.

Measuring criticism, as it is most commonly practiced, is evaluative and normative. In its most characteristic form it consists of an account of the personal thoughts and feelings of the critic in confronting the work of an accountant or planner or controller or evaluator or accreditator, with the aim of persuading the reader to share these thoughts and feelings. Free reference is made to the biography, psychology and character of the person in question, and even to the critic him/herself. The 'arguments' advanced in criticism are rarely arguments, properly speaking, but rather assertions of opinions and assumptions paraded as if their authority was unquestionable. The dominant discourse of such criticism is an uneasy and contradictory amalgam of Romantic, Realist and Modernist aesthetic theories. The 'history of measuring' predominantly supports such criticism in that it is produced within the same ideological framework. In such 'history' the unargued conventional assumptions to be found in 'criticism' are projected into the

past from whence they are reflected inverted in status—no longer mere assumptions, they have become the indisputable 'facts' of history.

I have described the dominant mode of history and criticism of measuring, in which the main concern is for reputations and objects, and in which the objects inherit the reputations to become commodities: a history and criticism to suit the saleroom. Neither history nor criticism is, a priori, committed to this course, and there are indications which follow of alternative approaches to history and criticism. Such alternative approaches reject the tendency to confine discussion of measuring to some narrowly technicist and/or aesthetic realm of ideas; they aim, rather, to understand measuring not only as a practice in its own right but also in its relation to society as a whole. This holistic project has traditionally been that of Marxist cultural theory, which of late has become increasingly engaged with precisely that topic of the production of meaning with which I began.]

In the above quotation [between brackets], the term 'measuring' can easily be replaced by the phrase 'evaluation and accreditation'. To pursue this line of thought further and to see how it can be applied to the question of the extent to which uniformity can be a problem in the context of evaluation and accreditation and how the problem might be addressed, we first need to consider what evaluation and accreditation are.

17.2 Evaluation and Accreditation

There is evaluation and evaluation, and there is accreditation and accreditation. These measures can be seen as both norm-setting in Kant's meaning of the term and also as examples of the antagonism he considers necessary, i.e., the process of setting and criticizing norms together. The recent evaluation debate in particular has shown what that means in concrete terms and that it has practical consequences. The central points of that development are interconnected with the term 'fourth generation evaluation'.[2]

Guba and Lincoln analyze three main problems of the first three generations of evaluation: 'Managerialism', 'Monism of', and 'Cartesianism' (Guba and Lincoln 1989, p. 31–32 and 90). When managers believe to have the right values and are privileged decision makers while other people do not know what to do, we talk about *Managerialism*. Concerning *Monism of Values* 'we would argue that conventional science is as a result a force for disenfranchisement and disempowerment, for maintenance of the status quo (Guba and Lincoln 1989, p. 125).' The attitude of *Cartesianism* believes in a world organized on fundamental principles scientists are able to decode. From a constructivist perspective these positivistic assumptions are far too strong and restrictive. The constructivist approach follows

[2] For more details on the following, see also Habersam (1996); and especially Guba and Lincoln (1989).

weaker assumption, a hermeneutic approach, and dialectical dialogs: 'Meaning is constituted in use' (Hedlin 1996)—e.g. in using measures, ratios, indices and standards to create the governable person (Miller and O'Leary 1987). Whenever we talk about evaluating, quality-defining, performance measuring or auditing, we talk about conventions, not truths. In practice, we are confronted with complex problems, competing interests, and the necessity for judgment. We try to resolve problems emerging from practical situations by impersonal, objective, theoretically generalized prescriptions of behavior. But this strategy has consequences. Michael Power's book on the 'The Audit Society' with the subtle subtitle 'Rituals of Verification'[3] shows the ambiguity of this attempt. Translated into German it would mean 'Die geprüfte Gesellschaft'. At first glance, the word 'geprüfte' means 'audited'. But 'geprüfte' is also a word of the Old Testament which means 'the ordeal sent by God.' Is auditing and evaluation one of the seven plagues, sent—not by god but—by enlightenment and its technocratic success during the last three centuries?

Guba and Lincoln take Robert Stake's 'stakeholder approach' to create a *responsive* evaluation procedure conscious of a concrete situation to be handled, its micro- and macro-political circumstances, and the inherently value-laden need for justifications (e.g., why to spent money on higher education).

'Responsive evaluation will be particularly useful (…) when the staff needs help in monitoring the program, when no one is sure what problem will arise (…), when audiences want an understanding of a program's activities, its strengths and shortcomings, and when the evaluator feels that it is his responsibility to provide a vicarious experience (Stake 1983/1973, p. 303).'

The pragmatic argument for *stakeholder involvement* seems to be the following: If stakeholders are not allowed to participate in this process, there is no enough support and acceptance for a solution. Evaluation becomes an interactive process that never ends.

'Responsive evaluation has four phases, which may be reiterated and may overlap. In the first phase stakeholders are identified and are solicited for those claims, concerns, and issues that they may wish to introduce. In the second phase, the claims, concerns, and issues raised by each stakeholder group are introduced to all other groups for comment, refutation, agreement, or whatever reaction may please them. In this phase many of the original claims, concerns, and issues will be resolved. In the third phase, those claims, concerns, and issues that have *not* been resolved become the advance organizers for information collection by the evaluator. … *The information may be quantitative or qualitative.* (…)In the fourth phase, negotiation among stake holding groups, under the guidance of the evaluator and utilizing the evaluative information that has been collected, takes place in an effort to research consensus on each disputed item. Not all such items will be resolved; those that remain become the core for the next evaluation that may be

[3] Cf. Power (1997). In a similar way Hansson (2006), refers to the consequences of an 'evaluation society', e.g. concerning social control at universities.

undertaken when time, resources, and interest permit (Guba and Lincoln 1989, p. 42).'

Consensus is neither result nor norm, but an *option*. Not all conflicts can be finished. But in most cases they can be handled in a confidential manner if consensus is not a norm but an option. It is also important to understand that this does not imply a container view of information but an emergent view.[4] Information is produced by the selectivity of the individual collection processes and the judgment of users. For this reason, the 'Fourth Generation Evaluation' is not only an alternative approach within the field of evaluative practices but also within the field of quality assessment , measuring/controlling, or even performance management practices.[5]

Putting it all together, Guba and Lincoln stress seven characteristics in their 'Fourth Generation Evaluation' (Guba and Lincoln 1989, p. 252–256):

1. The socio-political process of, if you want to extend the perspective, 'assessing quality', 'evaluation', 'performance measurement and management' is infiltrated and framed by culture, politics, and self-interest.
2. It is a joint collaborative process, which results sometimes in consensus, and sometimes in an agreement to disagree.
3. Each member of this process serves as both learner and teacher. The role of the evaluator is not to be the boss of the setting. S/He is willing to be taught by all other stakeholders: '(…) undoubtedly a new species in the ecosystem' (Guba and Lincoln 1989, p. 254).
4. The process is unpredictable and emergent. 'Every step, therefore, is contingent on the previous steps and can be unfolded only serially' (Guba and Lincoln 1989, p. 255). Design, setting, methods etc., which can be described only retrospectively are not really descriptive but again pieces of a new construction.
5. Not only the process is unpredictable in itself but the outcomes as well. They are not determined by input but by negotiation. 'It is not less politics in assessing quality that is needed but a finer-grained understanding of the values that underlie the various political and policy positions' (Guba and Lincoln 1989, p. 255).
6. This process is not producing a picture of the reality, but a new reality; not findings, but literal creations of all participants.

Within this process all the results, conclusions, and recommendations are joint results, conclusions, and recommendations—or nothing. The tragedy is that in everyday life the details of evaluated organizations, their organizational processes are not understood by many politicians and administrators in many universities. That there is a dominant use of cost accounting and control is the attempt to reduce

[4] For some impressive cases cf. Jönsson (1996).

[5] There is no fundamental difference between these approaches and their 'internal debates' on how to measure and to manage organizations; see Habersam 1997, referring to 'controlling' (i.e. the German term for the Anglo-Saxon term measuring and control), Bohni and Ejler (2008), referring to performance management.

the confusion about quality by tools that are a myth, too. Unfortunately, the creative potential of the members of universities is never really developed and put into action. The prevalence of cost accounting and of vulgar interpretations of business administration leads to bureaucratic administration of quantities as the (one and only) indicators for quality (Boyle 2001).[6]

Undoubtedly, there is a necessity of performance measurement and management in and for universities. But a machine model of the university decreases social competence and motivation. It does not lead to insights into the fabric of society. The arguments concerning the conventional and ritual character of counting and calculating, and its built-in danger, are fundamental. Is there a new kind of *Taylorism* coming back called *Evaluationism*? Will mere cost accounting and financial reporting produce very adapted graduates and more or less conventional research? To define quality quantitatively is not enough. It ignores the aspects of quality as a social construction and process of negotiations. For nearly a thousand years the quality of universities has had a polyphonic and recursive character, but this not guaranteed to last.[7]

Consequently, Guba and Lincoln try to avoid the disadvantages of a bureaucratic model. "The world is no longer seen as a closed system operating by immutable laws, which, once discovered, lay an inescapable mandate for behavior on us all" (Guba and Lincoln 1989, p. 256). From my point of view, to overcome the machine model of a university is to be done every day from scratch, because "experts" or "science" are indispensable parts of our society. When we can no longer hide ourselves behind their cloaks, the consequence is that "(…) accountability yields to shared responsibility" (Guba and Lincoln 1989, p. 256). But how are professors, administrators or students held accountable for the state our universities are in? "Much so-called evaluation (quality assessment, performance measurement and management; E.K.) is directed to making this accountability manifest (…)" (Guba and Lincoln 1989, p. 257). Europe-wide testing programs and rankings used to blame teachers, students, or other stakeholders are examples of what happens when evaluative processes in favor of accountability do not obviate the exploitation of the evaluated institutions and their members. "Conventional evaluators do virtually nothing to improve either knowledge or sophistication for any except a few privileged stakeholder groups, mainly clients and sponsors. But fourth generation evaluation shares knowledge and works at improvement of sophistication for all. (…) It is an educative experience for all"

[6] It is not by chance that the first cost accounting for universities was implemented by the "Technische Universität München" in Germany. Technicians are really good cost accountants but often lack social competence. And the present economic crisis reveals the illusions strongly connected to cost accounting and balance sheets in terms of bankruptcy and the fraudulences of formerly famous corporations and trusted banks.

[7] Even institutions with a longstanding polyphonic history may introduce evaluation programs "(…) strongly inspired by ISO 9000:2000 standards" (Darchini et al. 2006, p. 15), as the case of the University of Bologna shows.

(Guba and Lincoln 1989, p. 258). Quality and the process of studying means evaluating adults in interaction.

The ideas of Guba and Lincoln, presented more than 20 years ago, lead to a new understanding of the role of evaluators responsible for assessing quality in higher education: Facilitator, story teller and listener, illuminator, mediator are more a political role, although s/he must have technical skills of conventional evaluators. The controller becomes a collaborator, supporting a process of emancipation, not of an investigation. Her/his responsibility is to set the stage for this process, to sharpen the awareness of "reality", to act as change agent and to reconstruct the existing reality constructions. Reflecting the changing nature of "reality", these changes in the role-model of evaluators seem to be even more necessary than before.

"In a changing world, an age of 'liquid modernity' as it has been called (Baumann 2000), the design of a curriculum invokes two kinds of responsibilities. First there are the tacit responsibilities of the students. Should the curriculum be developing among students responsibilities toward (a) a discipline and its standards; (b) the world of work; (c) the wider society; (d) the student her- or himself? It may be said that it is all of those, but then large and possibly intractable problems arise. (…) The second set of responsibilities falls on the curriculum designers themselves. How do lecturers and others in universities who find themselves in positions of responsibility towards curricula understand those responsibilities?" (Barnett and Coate 2005, p. 43).

17.3 Conclusion

In the above discussion, the focus is on evaluation. If no explicit mention is made of accreditation that is primarily because the considerations of the theory of numbers and measurement theory apply in equal measure to accreditation. The motives for the introduction of accreditation processes are also comparable with the case of evaluation. Appointment processes are similar, too. At all events, deconstruction of the conditions and processes involved would be appealing, bordering in some cases on indiscretion perhaps.

"Although all this suggests that we should regard the spirit of performance measurement not as monolithic, but rather as fractured, incomplete and evolving, this spirit is also powerful. Specific measurement systems may be defective and fail, but they also constantly reproduce and reinvent an institutional demand for numbers. Political power can be understood as the ability to make even controversial counting and measurement systems appear natural and unavoidable, preventing the widespread institutionalization of distrust in numbers and supporting a variety of schemes for monitoring and control" (Power 2004, p. 769).

Although controversial and deserving of criticism, evaluation, and accreditation processes are now well established. If we are not willing to abolish them in response to their ambiguity and uncertainty (because it is no longer possible), we

have to learn to deal with the complexity and uncertainty of such processes. Then we will find that, when implemented correctly, they deliver useful results. "Performance measurement systems are both contingent and powerful, of varying precision, and are simultaneously democratic and tyrannical as forms of 'normalized transparency'" (Power 2004, p. 780).

Tendencies like uniformity and the excessive reduction of complexity, the signification of certain content orientations, and spoon-feeding programs cannot be ruled out and are indeed to be feared where evaluation and accreditation are handled bureaucratically. Where, in the context of "normalized transparency", norm setting gains the upper hand as normalization, evaluation cannot work. The fact that comparison is of great importance for the further development of the institutions assessed in the framework of evaluation and accreditation programs is almost self-evident. But that must not be permitted to lead to uniformity. The ideas of Guba and Lincoln could counter that with an enlightened process based on discourse.

To apply that process, the autonomy of the evaluated systems must first of all be restored. A system that cannot decide on its own reactions will not change in response to evaluation, either. Since, in a competitive system, diversity—the diversity of ideas—is a prerequisite for the adaptability of the system, and evaluation and related incentive and accreditation systems lead to a reduction of complexity, variety, imagination, creativity, and other strategic potentials, the initial response is likely to be a proposal to abolish evaluation. And that proposal is often voiced. But the response is misguided. Now that they are the fashion, evaluation and rankings will continue to exist and spread. Exact measuring processes deliver punctual trains and flights. The hope of being able to make comparisons and thus find superior yardsticks for our decisions encourages us to use similar processes to evaluate organizations and social relationships. The fact that such an approach can be inadequate will not eradicate the desire to measure and compare. Nor will the question whether the benefits are greater than the costs. And since the decisive factor is always who is doing the measuring, we have to decide how to respond. Who asks about the benefits, benefits for whom, and costs to be borne by whom? What benefits or costs are is not inherent in the measuring process. "It reflects the power and robustness of values other than an unabated need for precision and control" (Power 2004, p. 780).

The increase in evaluation methods and the differences in the results and conclusions are not just cause for criticism; they are there to be exploited. One-off and/or one-sided measurements cannot lead to meaningful conclusions. Conclusions about organizations must be based on a plurality of measurements with different evaluation and accreditation processes. That corresponds—albeit not at the same level of precision—to the techniques of surveying or triangulation. Where evaluation is seen as a complex context-based system while the organizational relationships that constitute the university are treated as multicontextual, the number of evaluations per organization or situation to be assessed must be increased. The differences that then arise make it possible to better recognize the facets and nuances in the overall system evaluated and to possibly process them

further in terms of fourth generation evaluation as proposed by Guba and Lincoln. Positioning the research output of a faculty on the basis of just one list of relevant journals and their weightings as proposed by the mainstream representatives of the discipline (Kappler 2009) would not then satisfy the requirements of a serious evaluation process. Abstractly derived comparability and standardization deny the complexity of individual and social relations, which could be handled through a discursive investigation of the wide range of evaluation processes and findings.

It is astonishing to find similar thoughts in the works of Wilhelm von Humboldt. What the founder of the modern university had to say on the subject of the state still applies today, both with regard to the state and with regard to evaluation and accreditation. Of course, the university does not have to be reduced primarily to teaching. It is therefore appropriate to finish with a closer look at Humboldt's ideas.

"What we call (…) higher scientific institutions are, released of all form within the state, nothing other that the spiritual lives of people guided through external leisure or inner striving to science and research. As such, the one ponders alone and collects, another connects with men of the same age, a third gathers a circle of disciples around himself. To this picture the state (or evaluation through diversity—E.K.) must remain true where these undetermined and, as it were, fortuitous activities are to be given more solid form. The state must aim:

1. to preserve this activity in its liveliest and strongest form;
2. to fully and clearly maintain the distinction between the institution of higher education and school.

The state must always be aware of the fact that it does not and cannot really-achieve this aim itself, that in fact it is a hindrance whenever it interferes, that the whole thing works infinitely better without it, and that the facts are really only as follows:

- that, since all disseminated knowledge must be provided with external form and resources in a positive society, the state has a duty to provide them for the works of science;
- that not only the way in which form and resources are provided can be detrimental to the essence of the matter; the very fact that it provide form and resources for something quite alien must of necessity also have negative effects and pull the spirit down from the heights to the plains of material reality;
- and that the state must therefore always keep the inner being in mind to make good what it, albeit unintentionally, has spoilt or hindered.

Under these assumptions it is easy to see that, with regard to the inner organization of the institutions of higher education, it is all-important to respect the principle that science is something that has not yet been entirely found and can never be entirely found and to seek it unremittingly as such" (Humboldt 1982, p. 256–257, (1810) (italics by E. K.)).

The picture of the university that Humboldt had in mind was doubtless not that of today's overcrowded universities. To that extent, we have to make concessions

with regard to his ideal picture of university programs and their evaluation. But that does not apply to his view of science as an object of search. In both cases, evaluation can help or it can be too much of a good thing. Its meaning can only unfold in concrete action.

Bibliography

Barnett R, Coate K (2005) Engaging the Curriculum in higher education. SRHE and Open University Press, Maidenhead

Baumann Z (2000) Liquid modernity. Polity, Cambridge

Bohni NS, Ejler N (2008) Improving performance? Exploring the complementarities between evaluation and performance management. Evaluation 14(2):171–192

Bowker G, Star SL (2000) Sorting things out: classification and its consequences. MA, Cambridge

Boyle D (2001) The tyranny of numbers: why counting can't make us happy. HarperCollins Publishers, London

Burgin V (ed) (1982) Thinking photography. Macmillan, London, pp 1–14

Crosby AW (1997) The measure of reality: Quantification and western society. Cambridge University, Cambridge, pp 1250–1600

Darchini D, Giannini S, Gola M (2006) Quality assurance and evaluation of programmes at the University of Bologna. In: Orsingher Ch (ed) Assessing quality in european higher education Institutions. Dissemination, Methods and Procedures. Heidelberg, pp 5–22

Groys B (1992) Über das Neue. Versuch einer Kulturökonomie, München/Wien

Guba EG, Lincoln YS (1989) Fourth generation evaluation. Sage, London

Habersam M (1997) Controlling als evaluation. Potentiale eines Perspektivenwechsels, München/Mering

Hansson F (2006) Organizational use of evaluations. Evaluation 12(2):159–178

Hedlin P (1996) Accounting investigations. Doc. Thesis, Stockholm

Heidegger M (1992) Was heißt Denken? Stuttgart (originally Tübingen 1984)

Humboldt Wv (1982) Über die innere und äußere Organisation der Höheren wissenschaftlichen Anstalten in Berlin. In: ders.: Werke, Band 4; Schriften zur Politik und zum Bildungswesen. 3. Aufl. Stuttgart, pp 255–266 (originally 1810)

Jönsson S (1996) Accounting for improvement. Elsevier Science, Oxford

Kant I (1968) Der Streit der Fakultäten. In Kants Werke, Akademie-Textausgabe, Band VII, Berlin, pp 1–116 (originally: 1798)

Kappler E (2009) Die List der Listen. Controller Magazin, Heft 1, pp 32–35

Miller P, O'Leary T (1987) Accounting and the construction of the governable person. Acc Organ Soc 12(3):236–265

Power M (2004) Counting, control and calculation: reflections on measuring and management. Human Relat 57(6):765–781

Power M (1997) The audit society: rituals of verification. Oxford University Press, Oxford

Stake RE (1983) Program evaluation, particularly responsive evaluation. In: Madaus GF, Scriven MS, Stufflebeam DL (eds) Evaluation models. View points on education and human services evaluation. Kluwer-Nijhoff, Boston, pp 287–310