Chapter 4
Purpose and Limits of a National Monitoring of the Education System Through Indicators

Stefan C. Wolter

Swiss Coordination Center for Research in Education, Aarau and University of Bern, Bern, Switzerland

4.1. Introduction

National education systems have traditionally been areas of society and state governance exposed to heavy political and normative control. This is still true today to a varying extent in many countries. Awareness in the political community of the immense guidance and governance issues involved in the education system developed in particular in the post-World War II period with the breakneck expansion of the non-compulsory education system and the growing recognition of the social and economic importance of education. Although the widespread regulatory fervor in the industrialized nations in the 1950s and 1960s was eventually abandoned, a return to a purely normatively guided education policy was no longer imaginable in view of the significant deployment of national economic resources in the education system and the importance of human capital for national advancement and development. Under the prevailing circumstances, it would have been simply too costly to use trial and error as a guiding principle in the political governance of the education system. The same applies to normatively motivated decisions that would contradict an alternative option based on rational and objective analysis.

Although this certainly does not address the full complexity of the issues involved, it is nevertheless possible to single out one event that was crucial to the new government management paradigm in the education system and elsewhere. The Blair government was probably the first to implement a theory-based, institutionally supported philosophy of political

N. C. Soguel and P. Jaccard (eds.), Governance and Performance of Education Systems, 57–84.
© 2008 *Springer*.

management and administration claiming to be based on evidence.[1] Although unadulterated evidence-based policies are not viable in the face of political competition, evidence-informed policies at least represent a significant improvement in terms of the rationality of political decision-making. Incidentally, a decision on the part of governments to base their actions on evidence as far as possible is not entirely altruistic and may reap dividends despite the awareness that a reduction in policy-making freedom arising from the evidence produced may be painful for politicians.[2] It is much harder for subsequent governments to overturn evidence-based decisions than to reverse decisions that are purely normative. Hence, for those political leaders wishing to bring about a long-term and sustained impact in various political areas, the better and more comprehensive their information is regarding the current status, the interrelationships and effects and consequences in the education system, the better equipped they will be to do so.

Evidence or information-based governance and management of the education system is both more urgently needed and more workable today than was the case just two or three decades ago. This is due to a number of intermeshing trends. To meet the growing need for data about the education system, most industrialized countries invested immense sums in education statistics and in administrative information systems. In this, the OECD played an important role to standardize statistics and thus make them internationally comparable. This was paralleled by huge progress in social science empirical methodology. One very significant development was the establishment of new statistical evaluation methods which were first used in other areas of government activity and regulation such as the labor market and healthcare but which can be usefully applied to virtually any area. Improvements in data handling and interpretation methods also helped to improve the statistical basis. Nevertheless, the social science research community also came to realize that observation of real-life phenomena on its own is not a sufficient basis for investigating causal relationships or the impact of new policies. Experimental or quasi-experimental – in other

[1] See, e.g., "Modernising Government White Paper", 1999 (http://www.archive.official-documents.co.uk/document/cm43/4310/4310.htm), or more specific on education: "Educational Research and Development in England", Examiners Report, 2002 (http://www.oecd.org/dataoecd/17/56/1837550.pdf).

[2] This is not a new concern, as the following historical citation shows: "When the Max Planck Society was considering plans for an interdisciplinary education research institute, a German minister for education worriedly noted: 'But that would pave the way for scientifically founded criticism of the minister for education's work.' " (translated quote from Becker 1971:17).

words, deliberately constructed – variations of the kind originally employed as standard procedure in natural science and later in medicine are also necessary as a basis for exploring social phenomena. The advances in the areas of statistics and research methods that were made in the area of education research are, in turn, indispensable for policymaker acceptance of an evidence-based or even just an evidence-informed management paradigm.

This chapter is divided into five parts. Part one (Section 4.2) briefly presents the developments prompting Switzerland to build up a national reporting on the education system within an extensive education monitoring project and the circumstances under which the request to set up this system was issued. Part two (Section 4.3) specifically explores to what extent statistical indicators are viable as a means of managing the education system as well as the limitations statistical indicators are subject to. Part three (Section 4.4) explains the set-up and the thoughts and ideas underlying the first Swiss Education Report. Part four (Section 4.5) uses selected practical examples and issues to illustrate the problems and limitations of using strict indicator approaches. Part five (Section 4.6) summarizes the experience gathered while compiling the first Swiss Education Report and presents the initial conclusions, which may be of benefit in terms of future reporting on the education system.

4.2. Knowledge-Based Governance and Management of the Education System Through Monitoring

In response to the international trends outlined above, Switzerland – albeit with the country's typical tardiness[3] – decided to expand and improve its education governance system through the use of standardized tools of education monitoring and education reporting. Back in the 1990s, ongoing efforts in the area of education statistics, primarily by the Swiss Federal Statistical Office, had progressed to the point where there was a general concurrence that indicator systems should be used to lay the foundations for a better informed and hence more rational education policy. These efforts in the area of education statistics were taken up by the political community toward the end of the 1990s. The idea in some quarters was that a narrowly defined set of indicators numbering no more than a dozen would be sufficient to supply the information needed to guide and govern the

[3]Compared with the Anglo-Saxon countries but not with its neighboring countries.

education system.[4] Then, in 2004, the Swiss Conference of Cantonal Ministers of Education (EDK) decided to launch an education reporting program together with the federal authorities within the scope of a national education monitoring project.

Education monitoring is an extensive process involving, among other things, periodic education benchmarking using tools such as PISA and its younger Swiss counterpart, HarmoS. These individual monitoring projects are intended to generate governance information for specific educational areas and issues. The Swiss Education Report is intended to combine this knowledge with information from other sources – administrative, statistics, research – to give a composite picture. The first Swiss Education Report was published at the end of 2006 and has pilot status. The work done on this first national education report, drawn up by the Swiss Coordination Center for Research in Education (SCCRE), will provide initial and important information on whether and to what extent, and under what conditions and limitations, an entirely indicator-based information system would be able to supply the information needed as a basis for guiding and governing the education system. Without wishing to preempt the conclusions presented later in this essay, it can be stated right here and now that, as a general rule, purely statistical indicators produce governance information that is neither unequivocal nor complete. However, it must also be pointed out that monitoring is intended as a diachronic process. Hence, what matters is not only an inventory of facts at a single point in time, but also an observation of events and developments between different points in time. The education report is intended to be published at 4-year intervals. In these 4 years, the findings from the first report will be processed by the education administration, statistics and research communities and are intended to guide education policymakers in defining issues and items of special interest for the subsequent education report. Thus equipped, the next education report will then present a follow-up inventory of the new status quo, taking a close look at the relative changes versus the prior report as well as presenting the current state of the education system. In order to really be able to evaluate the importance and benefit of the education reporting system, it will hence be necessary to await the completion of at least one full report cycle, i.e. at least two reports.

[4]This idea was in a striking contrast to the similar and longstanding international indicator project of the OECD, which in its annual publication (Education at a Glance) already in its edition of the year 2000 counted almost 400 pages.

4.3. Indicators and Indicator Systems

According to the usual definition, indicators constitute quantitative information on the status, characteristics, proficiency or effects of a system. In other words, they constitute empirically verifiable information which, ideally, will provide a basis for theoretically founded conclusions about a given system. Although it has long been standard international practice to use indicators, in the education system as in other areas, the definition, selection and relationship between individual indicators is still fraught with unsolved issues.[5] An indicator system that satisfactorily resolves all these questions has not emerged to date. The main reasons are briefly presented below:

(a) The various indicators should be operationalized in a manner allowing an unequivocal conclusion about the matter under investigation. Indicators should go beyond the purely informational dimension of statistics (see, e.g., Kanaev and Tuijnman 2001); otherwise, they would merit the term "descriptors" rather than "indicators". To deserve being called "evaluative", indicators must meet two main conditions. First, to ensure that each of an indicator's values permits an unequivocal conclusion with regard to a prevailing state or a need for action, there needs to be a benchmark for the chosen indicator. However, absolute standards (or even consensus standards) that would permit a unique interpretation of a particular indicator value are very rare. For instance, commonly used collections of indicators (Education at a Glance of the OECD, for example) give the reader virtually no precise pointers as to how a specific value should be interpreted, just the raw statistical data (see also Thomas and Peng 2004). One of the few exceptions is international assessment tests (PISA, etc.), which, on the basis of the judgment of experts, define scales within which the observed performance values are matched to specific proficiency levels. The lack of measurable criteria for evaluating indicator values has also to do with the fact that education policymakers (intentionally or unintentionally) have neglected to define precise, i.e., operationalized, goals for what the education system is expected to perform. In the absence of precisely defined

[5]From the German-speaking literature, see in particular section A2 on the status of indicator research from education reporting in Germany (Konsortium "Bildungsberichterstattung für Deutschland" (2003): Bildungsberichterstattung für Deutschland: Konzeption. Frankfurt am Main/Berlin) und den Bildungsreform Band 4 Bericht (Van Ackeren 2003).

criteria for evaluating a particular indicator reading, the observer must rely either upon ipsative (or self-referential) or reference group-related outcome measures. The ipsative approach attempts, on the basis of variation over time, to at least permit a directional conclusion as to whether the indicator has improved or deteriorated since the last measurement. The reference group approach endeavors to generate findings derived from international, national, regional or inter-institutional variations in values. Compared with absolute or consensus-based standards, both these approaches are mere stopgaps as they are not a sufficient basis for truly satisfactory evaluation. Nevertheless, for want of a better alternative, the Swiss Education Report has no option but to resort to these approaches in most areas.

(b) Second, the indicator's relevance in terms of the education policy objective must be theoretically and empirically validated. In other words, the indicator must be relevant to the achievement of a particular goal or as a basis for deciding for or against a political governance option. While the absence of standards or benchmarks imposes limitations in particular on the evaluation of output and outcome indicators, the relation to the output or outcome is crucial in the case of process and context indicators. All too frequently, process or context information, which has to be controlled or taken into account by the education policy community, whose actual relevance to education output and outcome is not conclusively proven, is observed (class size is one example; this will be explored in greater detail in Section 4.5). Mere correlations, particularly on outcomes (such as health, labor market status, crime, etc.), are all too often simplistically interpreted as indicating causality, resulting in the observation of indicators which may be entirely irrelevant in terms of the desired outcome. There is also a tendency to compare and contrast individual indicators (input–output, for example) in a manner that communicates a supposedly clear cause–effect relationship, even in cases where such a relationship is only assumed but has not been proven. In most cases where a relationship with the output or outcome can be assumed to exist, the strength of the relationship is unknown. Consequently, variations in the indicator do not enable an unequivocal conclusion with regard to the change in output/outcome thus produced. If, say, very large changes in the process variables are necessary in order to produce tiny changes in output, caution should be exercised when interpreting changes in the indicator (and vice versa).

(c) At best, indicators show a need for action. However, as a general rule, they produce little information on the available options. Even more rarely do they indicate specific actions that need to be taken. If an indicator shows that the mathematics proficiency of the students of a particular country is not quite up to scratch, this information on its own neither indicates how the students' mathematics skills can be improved (see also Section 4.5), nor does it show the consequences or impact of this lack of proficiency. These limitations in terms of the meaning of indicators are not a problem as long as people are aware of them. The problem is that many indicators are communicated in a manner that creates a different impression, which may lead to poorly thought-out education policy measures.

(d) A single indicator is generally unable to present a full picture of the item in question. Indicators are therefore parts of an indicator system whose composite information needs to be taken as a basis for evaluating a system's state of proficiency. However, as soon as a number of different indicators are projected into a system, it is clear that the interrelationship between the indicators is of key importance. It is easily apparent that a number of different inputs feed into the education process. What is not so readily apparent is how these individual inputs interrelate. The impact of an input may complement that of another input, i.e., the deployment of input A also enhances the impact of input B. Conversely, the two inputs may compete with each other such that increasing input A would concomitantly reduce the impact of input B. Manifold and complex interactions between the indicators mean that the composite information generated from an indicator system must not be understood as merely constituting the sum of the information generated by the individual indicators. The problem is that so little is known about the interdependence between indicators that a greater degree of detail (i.e., more indicators) in an indicator system does not necessarily correlate with a higher degree of evidence or utility for the user.

For reasons of space, other important matters such as data quality, level of aggregation, comprehensibility of the indicator cannot be entered into in greater detail here.

The Swiss Education Report does not solve the problems inherent in using indicators and indicator systems as a basis for guiding government and public policy. However, a threefold approach endeavors to minimize the problems as far as possible:

1. Right from the beginning, it was accepted that a clearly limited small set of indicators would not be able to describe the whole of the education

system in all its complexity. Any such endeavor would have been impossible, merely in view of the multiple goals pursued with the education system. In addition, the indicator set must also be differentiated and diverse enough to reflect all the education levels and types with all their particular specificity. Finally, the indicator set must be flexible enough to keep pace with continually evolving education policy issues and challenges. Although fixed sets of indicators make intertemporal comparison easier, the fact that an indicator was measured in the past is not an adequate guarantee that it addresses an issue that is relevant to the education system in the present.

2. An analytical framework (see Section 4.4) was defined in which the indicators are integrated. However, unlike the policy pursued with most known education reports, the authors did not act according to the principle, "you show what you have". Instead, they asked themselves first which indicator would be necessary to answer the relevant question within the analytical framework. If there was no indicator that met the authors' desires (as was often the case), the indicators used, instead of the desired indicators, were described and flagged as proxy information. Accordingly, much of the information is more akin to descriptors rather than indicators. The consequence of this approach is that the report in particular highlights those areas where we know little, whereas other reports tend to highlight those areas where knowledge is abundant. However, for the first education report in a continuous monitoring project, the chosen approach is more useful in terms of building up governance knowledge.
3. Much importance is placed on using additional information from other areas as a means of improving the understanding and interpretation of the statistical information. The vast majority of relevant statistical indicators are not self-explanatory; nor is their meaning always clear in view of the plethora of complex interdependencies in the education system.[6] Therefore, research findings and education administration data have been processed in order to understand the statistical indicators, to identify their relationship with other indicators and, ultimately, to estimate their impact on the education system. In keeping with the limitations already noted, this approach also reveals what the indicator cannot do or points out potential misinterpretations that one must be beware of.

[6]An example is the rate of return on education as an indicator. Although the return on education is definitely a better outcome indicator than simply salary differentials between various education levels, the indicator itself is commensurately more difficult to understand (see, for example, Wolter and Weber 2005).

4.4. Framework and Objectives of the First Swiss Education Report

Like other national education systems, the Swiss education system is divided into education levels and types.[7] Institutional differences between the education levels and types and differences in education goals, level- and type-specific organization, administration and responsibilities justify structuring the education report on the basis of the various education levels and types. Finally, it is important to remember that, as a rule, statistics and research activities are also based on the specific individual education levels and types. Although this makes it easier to understand how a specific part of the education system functions, it makes it more difficult to compare the various levels. Hence, structuring the education report on the basis of education levels is logical but comes with certain limitations. One such limitation, for instance, is that individual impacts of education (outcomes), for example on people's health or social behavior, are not attributable to a specific unique education level or education type, but constitute an outcome of cumulative education processes. These aspects are, to some extent, accorded too little attention in this education report. However, it must be added that the relationship between the level of education attained by an individual or the entire population and the above-mentioned education outcomes is difficult to determine and is not always unequivocal, in particular in terms of causality.

4.4.1. Context Information

The introduction to the report as a whole gives context information of importance to the education system in general. This context information presents the exogenous framework conditions for the education system as a whole. The sections on the individual education levels and types then take a more profound look at the factors emerging from the context information that are of specific importance in the particular situation. In order to arrive at a comprehensive understanding of the education system, it is important to bear in mind that the education system cannot be evaluated from an internal perspective only. An extensive analysis and evaluation of the education system requires co-analysis of the prevailing interdependencies with other social, economic and political processes and

[7]Adds to that, that there is not really a Swiss educational system but 26 different systems as the governance of most parts of the educational systems lies in the hands of the cantonal authorities.

frameworks. Developments pertaining to family structures, public finances or migration policy may be just as relevant to the success or failure of the education system as the efforts of the players in the education system per se. That said, it must be stated that, for all the importance of these general context conditions in terms of how the education system operates, hard empirical facts about their actual impact on the education system are fairly thin on the ground. The availability of knowledge in this area seems to be in inverse proportion to its importance, a circumstance due not least to the exclusionary "internal-only perspective" adopted for so many years in education policy-making and research.

4.4.2. Chapter Structure

Almost all the sections concerning education types have the same basic structure. Each education level/education type is described in five subsections. The first two subsections show the framework in which the education levels/types operate. On the one hand, one has the exogenous framework conditions (contexts), i.e., the social, economic or demographic trends which have a direct impact on how the specific education level/type operates. These exogenous contexts are derived from the general context for the education system as a whole, as described in the introductory context sections. On the other hand, the internal contexts (institutions) show the institutional characteristics of the education level or type in terms of the set-up, structure, permeability or the coordination between and the decision-making authority of the individual players in each particular area. These internal contexts may vary greatly between cantons, over time, or between Switzerland and other countries. Therefore, a conclusion regarding the proficiency or weakness of an education level or type is possible and admissible only after controlling for the exogenous and endogenous contexts.

The remaining three subsections evaluate the proficiency of the education levels according to three criteria. The first of these three criteria is the *effectiveness* of the education level, i.e., the degree of target achievement of a specific education level or education type in terms of the pertinent and relevant[8] education goals. In practical terms, this involves aspects such as the number of students who achieve or surpass the defined proficiency goals. Another possible measure of effectiveness might be the number of university and technical college students and the number of apprentices in basic vocational training who make a successful transition to the labor market.

[8]The relevance of education goals is determined by educational policymakers.

The second criterion is the *efficiency* of target achievement. Even if there is a high degree of target achievement for a particular education goal, the limited availability of resources in the education system (as in other areas) calls for continual scrutiny as to whether target achievement was efficient, i.e., was the goal accomplished through the use of the minimum amount of resources. Or, conversely, whether an even better target achievement could have been accomplished with the same expenditure of resources. Efficiency in the education system, although rarely explicitly stated as a performance review criterion, is an aspect of evaluation that should not be neglected, also in the interests of the learners.

The third criterion is *equity* in the education system. A high average degree of target achievement and a satisfactory level of efficiency in providing an education say nothing about whether all the people being educated in the system, regardless of their background, have equal opportunities to achieve success in the education system. The fact that different students generate different learning outcomes does not serve as a sufficient indication that the equity principle is being violated. The equity principle is violated only if students' affiliation to a particular group, socioeconomic stratum or gender limits or predetermines their educational outcome.

The identical structure of all the education level sections with the same five subsections is intended to enable a differentiated analysis of the individual education levels and to facilitate comparison between the levels.

4.4.3. Determining Education Goals

It should be clear from the above that any evaluation of education system performance must be based on education goals. In summary, the key issues are: Does the education system achieve the set goals, and to what extent (effectiveness)? What resources are expended in order to realize this degree of goal achievement (efficiency)? Are specific socioeconomic strata or nationalities or is a specific gender at a disadvantage in terms of goal achievement (equity)? All the analyses focus on education goals then. Some of these education goals differ significantly between education levels and education types, which in turn justifies evaluation based on the individual levels.

However, education goals are very often unclear, incomplete or have not even been defined in many areas. And even where general education goals have been defined, they are in most cases not or not fully operationalized. As a result, there is no consensus at the end of the day on a specific and verifiable goal definition approach. Finally, each of the education levels usually pursues multiple goals at the same time. A conclusive evaluation of

the overall degree of goal achievement is therefore not only extremely complex, but also a matter of political judgment. The latter is inevitable because goal hierarchies are necessary in the presence of multiple goals so that one can establish points of reference between degrees of goal achievement for individual education goals. An elementary school pupil should not only be able to read, do arithmetic and write, but also display other intellectual and social skills at the end of the compulsory schooling period. If it were necessary to express the success of education in a single variable, it would be necessary to know beforehand whether, say, performance deficits in reading can be offset by above-average mathematical skills or very good social skills, or whether indeed the latter is in fact more important. Goal hierarchies would be necessary in order to answer these and similar questions. Goal hierarchies of this kind are usually the outcome of political decision-making processes and are determined by the value-judgments prevailing in society at a given time. The authors of the pilot report were not in possession of any such goal hierarchies at the time of writing the report. Hence, both the selection and the presentation of the goals described represent the judgment of the authors and not that of the educational policy. However, the aim of a permanent monitoring process is for the political authorities to guide the process by specifying verifiable goals for future education reporting cycles.

4.4.4. Reference Variables and Dimensions

Apart from a few exceptions (see Section 4.4), education goals cannot be measured in absolute terms because the necessary outcome benchmarks have not been identified. In most cases, however, relative findings or outcomes are possible on the basis of comparisons. Suitable comparisons can be made on an intertemporal basis, between individual educational institutions or between different education systems. There are three dimensions which would be useful comparators for Switzerland and hence for this report. First, you can present the same outcome, averaged for Switzerland, at different points in time and in this manner at least determine whether goal achievement has improved or worsened over time. Second, you can compare values from one canton to another. This at least permits a relative value for each canton in terms of where it stands in relation to the best canton (which is used as a benchmark). However, this method does not generate a direct result for the best canton because this canton might still be operating far short of its theoretical potential. The same applies if you use individual educational institutions (universities, for instance) as comparators. Third, you can compare Switzerland with other countries. The same

possibilities and limitations apply as for intercantonal comparisons and comparisons between educational institutions. A key aspect in all comparisons is that it is assumed that there will be no differences in the circumstances or contexts in the course of time or between institutions that will be so significant as to render a comparison of the relevant variables meaningless.

Hence, each figure can potentially be presented in a very large number of possible comparisons. This report therefore selects and presents only those comparisons which yield a relevant or conclusive finding. In a few cases, comparators/comparisons have been chosen which do not yield a true finding, but which are commonly thought to be appropriate in identifying relevant differences in order to explicitly discuss the uselessness of the comparison.

4.5. Possibilities and Limitations of Indicator-Based System Governance and Management Explained Through Practical Examples

The following part elucidates the possibilities and problems that are encountered with education monitoring and the role indicators play within this context through reference to various practical examples. The examples given are to be construed in an exemplary sense, and care has been taken to ensure that they pertained to different education levels and different areas of education policy.

4.5.1. Class Size

Class size is a classic issue, not only in the realm of education policy but also in the field of education research. It is always amazing how adamantly teachers, parents and, in the wake of these two stakeholder groups, policymakers remain fixated on the issue of class size. If education authorities decide they would like to increase the average class size even only minimally (and there are plenty of examples of this involving cantonal education policymakers), it can be certain that the entire teacher body and many of the parents will rise up in collective protest. Reducing class size will also always figure prominently in the wish lists compiled by teacher unions. Judging by the political explosiveness of the issue of class size, one could easily assume that average class size for a canton and a particular kind of school must be one of the most important indicators within the entire education system. It is also convenient that this indicator

happens to be statistically well covered (although in Switzerland the corresponding data do not go back too far). But what does this indicator really stand for? Turning to educational research for an answer, which has probed this issue for decades and produced enough studies on class size to fill an entire library, the resulting findings can be distilled into one single sentence: School classes that are much too small and much too large have an equally detrimental effect on learning outcomes, but within these two poles (specialist literature mentions a range from approximately 15 to 25 students at compulsory public schools) any variation in the number of children in a school class will have no effect, or at most only a minimal effect, on student achievement (refer, for example, to Averett and McLennan 2004 for an overview of this issue). Considering this minimal effect, then, reductions in class size are in most cases not justified, simply for reasons of efficiency. One might be able to make a case for variations in class sizes by arguing that children who start school from a weak position stand to benefit from having a smaller number of children in their class (see Krueger and Whitmore 2000), but then this would also require a central planner who determines how many children are assigned to each class depending on the individual composition of the classes. Applying this indicator to the effectiveness of the education system, one can safely assume that, within the rather large range mentioned above, the number of pupils in a class does not say anything about the quality of the education provided at our schools.

What is clear, however, is that the indicator of class size is associated with two things that immediately highlight the contrast between educational policy and financial policy on the one side and the teaching body on the other. The first is that, regardless of the actual effects class size may have on student achievement, class size is an important indicator of the cost of the education system and, therefore, it serves as a significant benchmark in education funding. The second is that class size is, of course, an indicator of the work load placed on teachers, because any increase in the number of students per class means additional work for teachers both during and after classroom instruction and vice versa.

Now what should a report on education do with the universally popular indicator of class size in view of the financial-related interests and the vested interests of parents and teachers, not to mention the vast amount of research data? In the Swiss Education Report descriptive reference is made to the average class size in the cantons. According to that information, the average class sizes of all the cantons lie within a rather narrow range of about 17–21 students, so narrow then that one would not expect these differences in class size to give rise to differences in the quality of the cantonal education systems. The education report also reveals that in practically every canton that has

established guidelines pertaining to average class size, these guidelines are either adhered to or the average class size is actually one to three students below the recommended range. Furthermore, in many of the cantons that have fixed a relatively low average class size there is a significant percentage of school classes that fall short of even this recommended minimum. In addition, there were several cantons that – usually for topographic reasons – set the minimum size of classes so low that one would have to say that it was already suboptimal, i.e., that the learning environment itself was detrimental to the effectiveness of the education process. Based on the differentiated information on class size in Switzerland presented in this report, one can conclude that many cantons have cost-savings potential that could be realized without seriously jeopardizing the quality of education.[9] Taking into consideration demographic projections of the school-going population in the coming decade, it appears likely that inefficiency within the education system will increase significantly if no forceful countermeasures are taken. Innovative and comprehensive measures are required and, in some cases, these will have to be implemented against the resistance of the local school organizations.

The education report also allows comparisons to be made across the various education levels, which one can do with class size by comparing recent developments at public compulsory schools and universities. If the increase in average class size at Swiss elementary schools since 1995 had been proportionately the same as the rise in the faculty/student ratio in the field of humanities and social sciences at Swiss universities, then the number of children in the average elementary class would be approximately 26 instead of approximately 20. If the average class size had risen at the same rate as the overall deterioration in the faculty/student ratio at Swiss universities, then the average elementary school class today would number more than 22 children. This comparison raises the question why any attempt to increase the average school class size by one pupil can trigger such heated political discussions about the quality of the education system when an increase six times that number in the humanities and social sciences departments of universities does not provoke any similar public outcry about the quality of university education. Could this be because it is assumed that, unlike changes in the ratio of teachers to pupils at public compulsory schools, changes in the ratio of professors to students at the universities are not relevant to the quality of a university education? This may very well be the case but then this should at least be acknowledged and the universities credited for having increased their efficiency so much over the past 15 years!

[9]One must, of course, also take into consideration the transportation costs that would be incurred if schools in different towns were merged.

4.5.2. The Influence of Classroom Lesson Hours on Student Achievement

Another major issue that educational policymakers and administrators are always confronted with is to what extent the number of lessons given in any one particular subject area has an impact on the scholastic skills in that subject area. Here, too, the general stance is relatively clear: the more lessons, the better. Any reduction in the number of lessons taught immediately arouses concern that student achievement will suffer and, conversely, if student achievement is to be improved in a certain subject area, then, so the general opinion, the number of lessons taught in that subject area will have to be increased. The relatively scant research evidence on how the number of lessons in a particular subject area actually affects student learning levels and learning progress contrasts starkly with the public debate on this issue. That said, this is an admittedly difficult issue to research because it would require experimental variations in the number of lessons taught and the imposition of effective controls to ensure that no compensatory measures are taken in the control group and that the same learning standards are not simply applied to the experimental group, allowing it to cover the same amount of learning material in a greater number of lessons. Natural variations in the number of lessons taught are subject to these same difficulties and lead to other problems as well. For example, the effect of mathematics lessons at upper secondary schools on mathematics skills as presented in a TIMSS study (see Ramseier et al. 1999) was based on a different number of lessons in different branches of Gymnasium. Thus the results were distorted by the fact that there was a self-selection of the students into the different branches of Gymnasium. Consequently, it is completely impossible to ascertain to what extent the better mathematics skills of the students in those branches with an emphasis on mathematics were actually attributable to the fact that these students were receiving a greater number of mathematics lessons than the rest of students in other branches of the Gymnasium.

The PISA 2003 study with its focus on mathematics revealed, however, that there apparently is a relatively strong positive correlation between the varying numbers of cumulative classroom instruction lessons given in the subject area of mathematics in each canton and the average achievement score of each canton in the PISA test. This correlation becomes even more pronounced if the two extremes, Geneva and Ticino, are excluded. While this may initially be welcomed as confirmation that more lessons apparently do have a positive effect on skill acquisition, one can use this same data to illustrate just how far we are from possessing the knowledge necessary to provide educational decision-makers with clear policy inputs. First,

there is no evidence of causality between the number of lessons taught and student achievement scores. Instead, one must simply accept that the given correlation is indeed an indicator of a causal effect. Second, while the relationship between the number of lessons and PISA scores does appear to be rather linear, one must nevertheless assume that, like with any other input/output relationship, marginal returns would eventually decline as the number of lessons is further increased. It would make a difference, then, if one additional lesson were added to a subject area where two lessons a week were taught or if that same increase were made in a subject area where the number of weekly lessons was already much higher. Third, it must generally be assumed that one cannot simply increase the overall number of weekly classroom lesson hours to augment the number of lessons given in one particular subject area, so any increase in one area will always be accompanied by a reduction in classroom instruction in some other area. This makes it clear that, before deciding whether it would be worthwhile to increase the number of lessons in one particular subject area, the opportunity costs in the form of a possible loss of skills in a subject area where the number of lessons would be reduced must be calculated. Since student achievement is currently measurable in only a few areas of the school curriculum, any such decision is immediately shifted from the objective-scientific level to the political-normative level. And even if the overall number of classroom instruction hours were to be increased, the cost of an expanded school program would have to be duly taken into consideration in the decision-making process. The anticipated benefits for the state and society resulting from the increase in student competency would have to at least match the costs of the additional school lessons. This point also makes clear that an abstract variable such as an increase in student competency does not provide an appropriate information base upon which sound educational policy decisions can be reached. An appropriate information base would also always include data on the impact that student competency has in terms of personal, fiscal and social returns. Finally, it must also be noted that other alternatives besides increasing the weekly number of classroom lessons should be explored. For example, would other forms of learning or other learning technologies be more efficient in enhancing skill acquisition within the given lesson plans? The examples cited here clearly demonstrate just how far the currently available statistical information is from producing the management and governance knowledge that is required for making everyday decisions concerning the education system.

4.5.3. Willingness of Companies to Train Apprentices

The third specific example concerns basic vocational education, i.e., the educational programs offered at the secondary level II, which is still the path most Swiss adolescents take after completing their compulsory schooling. This third example is intended not least to demonstrate that education monitoring must certainly not be limited to gathering and assessing information and data on the education system itself but that other areas and stakeholders must also be monitored, depending on which education level and type are involved. Referring to the dual apprenticeship system, it is evident that this system would not even exist if companies were not willing to offer apprenticeship positions and, hence, training and employment opportunities, to Swiss adolescents. The willingness of companies to train apprentices is, therefore, a kind of *sine qua non* for the smooth functioning of the dual vocational education and training system. The question is, what information should the monitoring of this willingness to train apprentices be based on? An indicator that measures the share of companies training apprentices as a percentage of all companies active in Switzerland (which is frequently used by the Federal Statistical Office) is problematic for at least four reasons. First, the number of training firms says nothing about the number of training posts. If the number of apprentices per firm increases, a reduction in the number of training firms would not be problematic. Second, fluctuations in this indicator do not necessarily reflect changes in the willingness among companies to train apprentices, since the number of apprenticeship contracts concluded in any one period is equally dependent on the number of adolescents seeking an apprenticeship. Recent longitudinal studies clearly show that the number of training companies changes in response to demographic fluctuations in the adolescent population (see Müller and Schweri 2006). In the same context, one cannot say with certainty that the adolescents entering the market for apprenticeships always show the same level of scholastic ability and other skills that are required to successfully complete an apprenticeship program. Here, too, the latest research indicates that company willingness to hire and train apprentices fluctuates quite strongly in response to the actual or expected (from the perspective of the hiring firms) quality of school leavers (see, for example, Mühlemann and Wolter 2006). In this case, the "ceteris paribus" assumption would no longer stand and a declining indicator would not mean that companies were less willing to train apprentices. Instead, it would reflect a deterioration in the quality or qualifications of the school leavers. Third, the percentage of companies willing to train apprentices depends just as much on the aggregate number of companies active in

Switzerland as it does on the actual number of companies providing apprenticeship training. If, for example, there is a sudden increase in the aggregate number of companies while the number of adolescents remains stable, then – assuming the willingness to train apprentices is the same among the new established companies as among the older ones – this would lead to a statistically unobservable overhang of companies that want to hire apprentices but have been unable to do so.[10] Fourth, while a simple indicator like this would, despite all the limitations mentioned above, still provide some ex post information about the willingness among companies to train apprentices, educational policymakers would probably be more interested in information that is also meaningful in an ex ante sense. Information about the factors that influence this willingness to train would probably be more appropriate for meeting the latter need. As long as it was presumed to be a natural given that the willingness to train apprentices was governed primarily by the long traditions of apprenticeship training in the corporate sector as well as companies' sense of social responsibility, it seemed pointless to investigate such indicators. In the meantime, however, research has demonstrated quite clearly that, from the companies' viewpoint, the willingness to hire and train apprentices is a decision that is subject to the same business logic as any other decision with a bearing on corporate activities and performance. The cost–benefit ratio of apprenticeship training activities has thus become a decisive factor used by companies in determining whether they should even offer apprenticeship training positions (see Mühlemann et al. 2005; Wolter et al. 2006). A cost–benefit indicator is not only a quantifiable variable but also a variable whose impact on the one variable that is of particular interest to educational policymakers (the willingness of companies to train apprentices) has been scientifically examined and validated.

Another reason why the cost–benefit ratio of a training program is appropriate as an indicator of corporate willingness to provide training is because political decisions made in the field of vocational education often have a direct impact on this ratio. Consequently, continual monitoring of the cost–benefit ratio of such training programs is one means of ascertaining the cumulative effects that the complex supply and demand side factors as well as the political factors have on the willingness to provide training opportunities.

[10] As a matter of fact, the considerable increase in the number of firms in Switzerland, for example, is attributable to a strong increase in the number of one- and two-person companies, which are in no position to train apprentices. This means that the percentage of companies engaged in apprentice training is being measured on a false base.

This example is intended to demonstrate that obvious and easily quantifiable data do not always serve as the best indicator and that the expressiveness of more complex indicators produced through research might be superior to easily quantifiable indicators because the causal connection between the indicator and the targeted objective is proven rather than presumed.

4.5.4. University Rankings and Other Indicators Used in the Tertiary Education System

This section closes with a look at several examples of more or less viable indicators used in the tertiary education system. Comparisons at the tertiary level are generally made between individual universities rather than entire university systems because most persons knowledgeable of the system recognize that, in view of the significant variance between the universities with regard to the quality of education offered, what matters is the achievements of the individual institutions. Given the widespread popularity of national and international rankings of universities, the meticulously compiled country comparisons of university expenditure per student that many institutions use (see for example OECD Education at a Glance) seem to be out of place. What exactly is being compared in comparisons of average university expenditure per student in Switzerland and the corresponding figure in the United States? The only thing these numbers have in common is that they both have something to do with spending on persons who are attending universities. Considering, for example, that most of the university students in Switzerland are attending universities that, according to international rankings, are ranked among the top 200 universities in the world, then, assuming that quality has its price, one would have to select a completely different set of reference variables. While it is true that many of the world's leading universities are located in the United States, the vast majority of university students in America do not study at these universities and a comparison of average expenditure per student at the University of Zurich and at Stanford University would certainly produce a completely different picture than the same comparison between the University of Zurich and the University of Nebraska.

The exact same interpretative difficulties are encountered when comparing variables such as government spending on tertiary education systems as a percent of GDP or the growth rates of spending per university

student over time.[11] The only thing these commonly applied indicators have in common is that they raise more questions than they resolve, not least because the input variables usually cannot be matched against corresponding reference variables for the output.

If one attempts to measure the output or the quality of universities, it is tantamount to opening a Pandora's box. Only a few thoughts on this subject will be presented here for lack of space. Whereas the individual faculties of different universities are generally compared in national rankings, and rightly so, international rankings are usually based on comparisons of entire universities. Exactly what a number of Fields medal winners[12] say about the faculty of law at a university is somewhat puzzling, though, and not only for those uninitiated in the ways of university rankings. In order to glean some meaningful information from these rankings, correlations between what should be measured and what is actually being measured must be presumed which go well beyond the limits of plausibility. These problems arise mainly because many university rankings do clearly state what is being compared but not what the results of this comparison actually mean. The validity of this point is underscored by an ancillary finding in a German study (see Büttner et al. 2003) that revealed that a comparison of all the rankings based on professors, students, personnel directors or expert opinions (CHE ranking[13]) sometimes produced a positive correlation and sometimes no correlation at all and in some cases even indicated a negative correlation. If the same things (quality!) were being measured in all the rankings, then all the correlations that were not significantly

[11] In these comparisons it is not even clear, for example, exactly which cost items are included in the calculations. If research expenditures are included, for example, then a university that successfully competes for research funds will become an "expensive" university in terms of spending per student. The same ambiguity can apply to the increase in spending over time if one cannot observe for which inputs and outputs more funds were appropriated. If, for example, a country neglects the funding of its universities over a longer period of time and must later compensate for this by raising spending levels, then the sudden strong increase in expenditure can be interpreted in two completely different ways. If the quality of the services provided by the universities remains at the old level, then one would have to interpret the corresponding growth figure critically as a deterioration in efficiency and a waste of money, but if quality increases, then the same figure would be a sign that the said country made important and effective investments in its tertiary education system.

[12] The Fields medal is the highest scientific award for mathematicians.

[13] For a critique of the role of this institutional evaluator (CHE, *Centrum für Hochschulentwicklung*) see, e.g., Ursprung (2003).

positive would be indicative of a problem that not even the highly popular summation of individual rankings could resolve.

Looking more closely at the indicators on the quality of universities, one encounters three basic problems. First, with some sub-indicators it is not clear why they can even be used to substantiate claims about the quality of education offered. The quality of university libraries may be important but, considering today's means of gathering information, it is not clear what causal impact libraries actually have on the quality of education given at universities. Second, there are indicators where it is not certain whether they are a cause or a consequence of quality (inverse causality). One can take the faculty/student ratios used in various rankings as an example here. Regardless of the fact that, as mentioned earlier, 20 years of research activity on class size has been unable to produce any conclusive evidence concerning the strength of the influence this ratio has on scholastic performance, the same indicator is accorded significance at the university level for ranking institutions. The problem here is the relatively long time lag before the number of professors in any particular faculty will change in response to fluctuations in student enrolment numbers, and for good reasons that require no further explanation here. In Switzerland the duration of this phase of adjustment (empirically measured from 1990 to 2002) is approximately 5 years. What's more, in Switzerland university faculties are usually so small that the creation of just one additional professorship can have a big impact on the faculty/student ratio. Under these conditions one can now imagine the following hypothetical case. Faculty X at the universities Y and Z each consist of 5 professors and 500 students at a certain point in time *t*. Consequently, the faculty/student ratio at both universities is equally good (or bad). The faculty at university Y is then beset with some quality problems, however, and subsequently loses half of its students to university Z over a period of 4 years. The faculty/student ratio has thus changed from 1-to-1 at the point in time *t* to a ratio of 1-to-3 in favor of university Y. If the faculty/student ratio is used as a measure of quality in a ranking, then the university that shows an improvement is the one whose improvement in its faculty supervision and guidance profile happens to be a consequence of its low quality![14] The third basic problem is that most rankings are based on a large number of sub-indicators, which

[14]See for an example the homepage of the "Swissup Ranking" of Swiss universities where for some disciplines the student/professor ratio is used as an indicator for quality (http://www.swissupranking.com/ranking-result.php?field=1&stats=1&display= ranking).

is problematic because every summarization into a single indicator rests on bold assumptions about the weighting of the individual sub-indicators.[15]

The examples given above are intended to show that the indicators regarding government expenditure on tertiary educational institutions that are so heatedly debated in political circles cannot be interpreted in any meaningful way without corresponding output parameters. However, there is a lack of hard research data that would soundly validate the use of these output parameters (especially those immensely popular rankings). Unfortunately, this is an area in which the doable dominates what would be meaningful in a manner that borders on irresponsibility.

4.6. Initial Findings from the Pilot Report on the Swiss Education System and Conclusions

As posited in the preceding sections, self-evident difficulties become apparent when one attempts to guide and govern the education system with governance knowledge derived from a system of indicators. However, this should not lead one to conclude that it would be better not to use such a system. The only alternatives to the governance and management of the education system based on indicators that are periodically and systematically gathered and interpreted would be relying on political-normative ad hoc decisions or a semi-scientific "trial and error" approach.

The fact that indicator-based governance today does not necessarily guarantee a qualitative advantage over the two aforementioned alternatives is not because an indicator-based approach is generally inferior but because the knowledge required for the successful application of indicators is fragmentary. The main reasons for the incompleteness of such knowledge are briefly explained below because they hold the key to future successful governance and management via monitoring:

(a) Although great achievements have been made in education statistics during the past two decades, major investments are still necessary, both to improve existing statistics and to extend statistical coverage to previously uncovered areas. Two examples particularly important in Switzerland will suffice here. When it comes to educational careers, statistics are still disadvantaged by the segmentation of the education system into different levels and

[15]Particularly resourceful producers of rankings therefore let the users of the information determine the weighting and composition of the various sub-indicators in producing an overall assessment.

different types or pathways. On the input side, monetary costs are still not comparable and there is still very little differentiation in the gathering and recording of real inputs.

(b) Apart from its participation in international achievement tests such as TIMSS or, more recently, PISA, Switzerland does not have a home-grown tradition of administering and conducting achievement tests. As such, an overview of the level of skills achievement within the education system is lacking, from both a cross-sectional and a longitudinal perspective. Participation in PISA did help to determine where Switzerland stands in an international context but, precisely because of its cross-sectional nature, PISA is unsuited as a means of producing knowledge chronicling the origins of the given proficiency levels, yet such knowledge is crucial for governance and management purposes (see also Wolter 2004). The "value-added" approaches this would require can only be implemented after data on individual achievement have been repeatedly gathered and recorded over time.

Together, these two points constitute a knowledge deficit at virtually all levels of education which imposes severe limitations on the assessment of both the effectiveness as well as the efficiency of the education system.

(a) Education research has devoted much of its attention to the internal-only view of the education system during the past decades and paid too little notice to the influence the surrounding environment exerts on the production of education and to education output, to say nothing of education outcomes. Consequently, there is a lack of empirically validated systemic knowledge, which is necessary to adequately grasp the interaction and interrelation between indicators. And that is precisely what is needed before one can even begin to speak of an indicator system.

(b) The long-standing neglect of empirical educational research (see for example Angrist 2004) is a disadvantage when setting up an education monitoring system, because it has led to a situation in which educational policymakers do not have enough knowledge about cause–effect relationships (causalities) and effect size. In real application conditions, knowledge of both is essential, however, and neither theoretical nor experience-based knowledge (expertise or historical comparisons) are a perfect substitute for such knowledge. Another consequence of this long disregard of empirical and hence social-scientific aspects in traditional educational research is that upcoming educational researchers

have received inadequate training in scientific methodology – also when compared with other social science disciplines – and, as a result, there is too little human capital available for future research purposes.

(c) Referring to quantitative-oriented researchers and the statisticians, it seemed that for a very long time they were content with what was merely doable. This led to a situation where the validity of assumptions and interpretations was not established on the basis of stringent scientific analysis but merely inferred on the grounds of plausibility. The misguided use of education indicators in many studies on education mentioned in Section 4.5 of this essay gives credence to this view. Assertions made in this conjunction have eroded the confidence of practitioners in research and statistics and are not entirely blameless for the scanty funding of educational research (again in comparison with other research fields).

(d) Referring to the educational researchers, the relationship with educational policymakers is, unfortunately, still somewhat uneasy and inhibited. A display of interest in research on the part of policymakers is often viewed as a threat to freedom of scientific research. Applied research and hence research geared to political issues was therefore often considered inferior to academic research. On the other hand, however, it is clear that responsible-minded policymakers and education administrators can improve the relationship with research only by acting responsibly when granting research contracts and when applying the findings of research activity. It is certainly possible to build a mutually beneficial relationship between researchers and policymakers, one that produces both academically outstanding quality and knowledge that is of relevance to system governance, as this has already been demonstrated by several other countries in a very convincing manner.

Reflecting on these six points – to which others could certainly be added – one might be tempted to ask whether education monitoring even makes sense under these conditions. Such doubts are justified but they are dispelled by the following two thoughts and observations:

First, a permanent and systematic monitoring and reporting process is needed to improve the knowledge and structures in those areas that constitute today's main problem areas. And this monitoring might also help to curtail the consumption of resources in those areas where unsystematic and redundant knowledge is currently being produced.

Second, one can already observe improvements in all six points mentioned above, so it appears certain that the second education report will have already filled some of the major gaps in governance and management knowledge. In the area of statistics the introduction of personal student identification numbers should enable the collection of more comprehensive data on individual educational careers. Regarding the costs of education, initial results have been produced at the university level and with regard to basic vocational education programs; further improvements will follow. The HarmoS project with the national standards and student achievement tests taken at three different times during compulsory schooling will not only enable effectiveness statements on the quality of public schools but also generate the data education researchers require to improve the understanding of education processes. Other evaluations, EVAMAR 2 for example, or large-scale pilot studies such as the *Basisstufenprojekt* in German-speaking cantons (a basic primary school project spanning 2 years of kindergarten and 2 years of primary school) are an indication of how knowledge can be generated in systematic, large-scale projects that is of practical use for governing and managing entire areas of the education system. In educational research, efforts are underway in the traditional education sciences, for example through structured and inter-university doctorate schools, and other social sciences (sociology, economics, political sciences) are displaying greater interest in education-related issues. Furthermore, one can expect that all of these developments and efforts will put a self-reinforcing process in motion that will prove to be beneficial to the quality of research and thereby strengthen the validity of research outcomes. Confidence in research findings and statistical information is, ultimately, the basic requirement that must be fulfilled before educational policymakers and education administrators will display a willingness to embrace a rational process of "evidence-based or informed policy". A final example here is the innovative research promotion instruments that are being tested in Switzerland, which should enable a more rewarding interplay between the education administration and education research. Reference is made in this regard to the concept of the so-called "Leading Houses" of the Federal Office for Professional Education and Technology, which is designed to address the needs of both researchers and administrators in a simultaneous "top-down" and "bottom-up" approach.

Before closing, the question whether better governance and management will also lead to a better education system must still be addressed. As is so often the case, better governance and management alone will not produce a better education system but it is a necessary precondition!

Acknowledgements

The author wishes to thank those who worked on the "Swiss Education Report 2006" for their input, with special thanks to Miriam Kull. All statements and any errors are the author's sole responsibility.

References

Angrist JD (2004) American Education Research changes tack. *Oxford Review of Economic Policy*, 20(2):198–212

Averett SL, McLennan MC (2004) Exploring the effects of class size on student achievement: what have we learned over the past two decades? In: Johnes G, Johnes J (eds) *International Handbook on the Economics of Education*. Edward Elgar Publisher, Cheltenham, pp 329–367

Becker H (1971) *Bildungsforschung und Bildungsplanung*. Suhrkamp Verlag, Frankfurt am Main

Büttner T, Kraus M, Rincke J (2003) Hochschulranglisten als Qualitätsindikatoren im Wettbewerb der Hochschulen. *Vierteljahreshefte zur Wirtschaftsforschung* 2:252–270

Kanaev A, Tujinman A (2001) Prospects for selecting and using indicators for benchmarking Swedish higher education. Working Paper, Stockholm

Krueger A, Whitmore D (2000) The effect of attending a small class in the early grades on college-test taking in the middle school test results: evidence from project STAR. NBER Working Paper W7656

Mühlemann S, Wolter SC (2006) Regional effects on employer provided training: evidence from apprenticeship training in Switzerland. CESifo Discussion Paper 1665

Mühlemann S et al. (2005) a structural model of demand for apprentices. CESifo Working Paper 1417

Müller B, Schweri J (2006) Die Entwicklung der betrieblichen Ausbildungsbereitschaft, Eine Längsschnittuntersuchung zur dualen Berufsbildung in der Schweiz, Zollikofen. SIBP Schriftenreihe, Band 31

Ramseier E, Keller C, Moser U (1999) Bilanz Bildung. Eine Evaluation am Ende der Sekundarstufe II auf der Grundlage der *Third International Mathematics and Science Study*. Rüegger, Chur

Thomas S, Peng WJ (2004) The use of educational standards and benchmarks in indicator publications. In: Scheerens J, Hendricks M (eds) *Benchmarking the Quality of Education*. (Official Project Report prepared for the EU commission. January 2003)

Ursprung H (2000) Schneewittchen im Land der Klapperschlangen, Evaluation eines Evaluators. *Perspektiven der Wirtschaftspolitik* 4:177–189

Van Ackeren I (2003) Entwicklung von und Forschung über Bildungssindikatoren. In: Van Ackeren I, Hovestadt G (eds) *Indikatorisierung der Empfehlungen des Forum Bildung*. Bundesministerium für Bildung und Forschung, Bonn

Wolter SC (2004) Leistungsindikatoren im Bildungsbereich. Forum der Bundesstatistik, Ökonomische Leistungsfähigkeit Deutschlands, Band 44, Statistisches Bundesamt, Wiesbaden, pp 127–136

Wolter SC, Weber BA (2005) Bildungsrendite – ein zentraler ökonomischer Indikator des Bildungswesens. Die Volkswirtschaft, Oktober, pp 38–42

Wolter SC, Mühlemann S, Schweri J (2006) Why some firms train apprentices and many others do not. *German Economic Review* 7:249–264