Discrepancies between intended and achieved effects of school evaluation in international comparative perspective: empirical evidence and recommendations

Hans-Georg Kotthoff

ABSTRACT

According to organisational theory and new institutionalism, any intervention into complex social systems will produce unintended or even contra-productive side-effects. Taking the well established theoretical assumptions of these two fields as a starting point, the author argues that the introduction of elaborate evaluation measures into an education system which can be regarded as a massive intervention into the «grammar of schooling» will also produce discrepancies between *intended* and *achieved* effects. This assumption is 'tested' by analysing and comparing the empirical evidence available on the aims and effects of school evaluation in England and Sweden. The analysis shows that the aims of school evaluation can be paradoxical and that they can be achieved in different ways. In addition, there is strong empirical evidence that the introduction of school evaluation into the English and Swedish education system has also generated unpredicted, unwanted and even contra-productive side-effects. Any attempt to introduce elaborate evaluation systems into any given school system should therefore try to minimize discrepancies between intended and achieved effects of school evaluation. The article finishes by giving conceptual and practical recommendations which could help to minimize these discrepancies.

I. Introduction

In contrast to many school effectiveness researchers who share the general assumption that almost any school can be turned into an effective school (cp. Teddlie & Rose, 2000), there is growing scepticism amongst school researchers and educational sociologists with regard to the possibility of creating effective schools in «adverse» circumstances (cp. Coe & Fitz-Gibbon, 1998; Lauder *et al.*, 1998; Ball, 1998). The scepticism is also fuelled by empirical findings generated in other social sciences which strongly suggest that any intervention into complex systems or organisations (like schools) will produce unintended or even contra-productive side-effects. According to organisational theory, educational institutions *can* be changed but they cannot be changed according to an exact plan because they are very complex social units which consist of individual actors who do not always act according to rational principles (Clarke, 1985; Scott, 1995). This view is supported by proponents of the «New Institutionalism» who try to explain the high degree of institutional stability and discrepancies between intended and achieved effects of institutional change by investigating how members of institutions change, break down or «personalize» externally imposed reforms and adopt them to their own routines and attitudes (Dimaggio & Powell, 1991; Crowson *et al.*, 1996).

If we take these well-established assumptions from organisational theory and New Institutionalism as a starting point, it is highly likely that external interventions into school *systems* will also produce discrepancies between intended and achieved effects. Thus, the paradigm shift from preceding bureaucratic «determination» of schools to subsequent control of schools through educational evaluation which must be regarded as a massive intervention into the «grammar of schooling» (Tyack & Tobin, 1995) will –apart from its intended effects– also produce unintended, undesirable or even contra-productive (side)effects. While *intended* effects can be defined as the official aims of school evaluation which feature most prominently in the steering and policy documents of most European education systems where they serve a legitimizing purpose for the introduction of school evaluation into a given education system, the *achieved* effects of school evaluation can be defined as real effects which have been proved empirically in various research studies.

In this article I will critically examine the assumption about discrepancies between intended and achieved effects of school evaluation by analysing the empirical evidence on aims and effects of school evalua-

tion in England and Sweden. The rationale for focusing on England and Sweden is based on the observation that both countries have changed the evaluation of their education system fundamentally during the last 20 years which has generated a wealth of practical experiences and empirical evidence with regards to the effects of school evaluation. Apart from this common feature there are a number of individual features which make Sweden and England very interesting cases particularly from a German point of view: Firstly, the Swedish case is of particular interest because the development of a systematic approach to school evaluation in Sweden took a similar starting point in that it was also initiated during a process of decentralization in education. The local autonomy of Swedish schools has been strengthened considerably since the beginning of the 1990s while a complex system of school evaluation has been developed at the same time. The Swedish education system has therefore quite recently initiated and carried out educational reforms which are similar to those educational reforms which are currently being discussed in Germany. Secondly, the English evaluation system is an equally interesting case because it represents a model of school evaluation which is quite unique and very prominent in the international debate on educational evaluation and can therefore serve as an important reference point in designing evaluation systems. Thirdly, the chosen education systems represent two different educational traditions within Europe (British, Nordic) which means that the effects of school evaluation can be studied in different contexts, i.e. in a traditionally «decentralized» system (England), as well as in traditionally «centralized» education systems (Sweden). Finally, the length of practical experiences with different concepts and instruments of school evaluation varies considerably between the two countries which enables us to study long and short term effects of school evaluation.

2. *Intended* and *achieved* effects of school evaluation: the empirical evidence[1]

Beginning with the intended effects of school evaluation in England three aims seem to be of particular importance: The prime aim of the highly standardized evaluation of schools in England is to increase the accountability of schools and to make their efficiency and quality more transparent and comparable. By doing this, parents should be enabled

to base the choice of school for their child on rational and objective criteria. Secondly, by publishing all evaluation results rigorously the state wants to increase the competition between schools and to motivate the schools to raise their standards. The competition between schools is intensified further by linking the funding of schools directly to the number of enroled students (*per capita funding*) which means in practice that dropping enrolment numbers will lead to a continuous decrease of the school budget. Thirdly, by inspecting all schools with the help of external evaluators on a regular basis the state wants to control the quality of individual schools and stimulate long-term internal school improvement and the professionalization of teachers («improvement through inspection»).

Contrasting these aims of school evaluation in England with some of the achieved effects which have been proved in empirical studies leads to the following results: While the accountability and transparency of the English education system has undoubtedly risen by exposing individual schools to public scrutiny (e.g. through school inspection), the hoped for effect that parents would base their choice of schools on «hard» and supposedly objective evaluation has not been proved yet. On the contrary, Walford points out that even years after the introduction of league tables in 1992 most parents base their choice of schools primarily on the wishes of their children and that the academic quality of a given school is a rather inferior selection criterion (Walford, 1996: 57). In addition, the publication of evaluation data seems to favour in particular those privileged middle and upper class parents who are able and prepared to deal with the complexities of comparative analyses. According to Walford, the increased transparency of the quality of individual schools combined with the free parental school choice does not therefore lead to the general improvement of all schools but to a hierarchization of schools (op. cit., 60). A second (unintended) side effect of increased transparency in education on the one hand and strengthened parental choice on the other is that «schools are tempted to become increasingly selective, both academically and socially» (Whitty *et al.*, 1998: 116) in order to attract children who have got the potential to enhance the school's test scores.

With regard to the increased competition between individual schools as a result of the rigorous publication of evaluation results the empirical evidence is ambivalent. While there is some evidence that the publication of evaluation results stimulates competition between those schools

which are competing for good students (i.e. the «good» schools), there is also evidence that weak or so-called failing schools are primarily concerned with handling their serious problems rather than competing with neighbouring schools (Kotthoff, 2003: 355). With regard to the large number of «average» schools it seems rather unlikely that they will develop a competitive attitude since they can hardly use their mediocre evaluation results to distinguish themselves from other schools. The assumption that competition between schools will, in the long run, lead to higher standards has been repeatedly questioned by English researchers, because it presumes real choices between competing schools – a situation which is most likely to arise only for those parents who live in urban centres (Walford, 1996). The desired effect is also quite unlikely, because it is common knowledge that it is the popular and oversubscribed schools which choose their students rather than the parents selecting the schools (Whitty *et al.*, 1998: 116). While the empirical evidence that competiton will raise standards is still outstanding, there are a number of «undesirable» side-effects which have been attributed to the publication of evaluation results in combination with financial sanctions. The least desirable effect is that so called «failing schools», which are permanently situated at the lower end of the performance tables because of their disadvantaged catchment area, are hardly inspired by poor evaluation results to improve their performance, but feel rather disillusioned and demotivated (Kotthoff, 2003: 357). Another side-effect of the use of raw league tables of performance is, according to Sammons, that it can «increase pressures on schools to exclude disruptive pupils and encourage schools to take older poor attendees or those unlikely to its public examinations off role, since these pupils have an adverse effect on raw league table positions» (Sammons, 2001: 9). Finally, there is also strong empirical evidence that the strong focus on the schools' output in form of their test and exam results and the explicit culture of performativity has had a standardizing influence on teaching learning processes («teaching to the test») and a narrowing influence on the school curriculum: «The emphasis on the "core" subjects [English, mathematics and science] together with the requirement that schools publish tests as the main indicator of educational success, is leading to pressure on non-core subjects, such as music and art [...]. Non-assessed learning appears particularly vulnerable» (Whitty *et al.*, 1998: 87).

Judging from the available empirical data on the effects of school inspections, there can be no serious doubt that school inspections achieve

their intended quality *controlling* effect to a very high degree. This is primarily due to the broad and very systematic conception of the whole inspection process which takes account of numerous aspects of the quality of schools as well as the rather skilful triangulation of different inspection methods (classroom observation, documentary analysis, interviews with pupils, teachers, headteachers), which increase the probability that every aspect of school life is scrutinized in detail. However, while the accountability function of school inspection is certainly achieved, its quality *developing* function is less visible: headteachers report in interviews that it is in particular the preparation time in the run up to the school inspection which has a quality developing function and which increases their teachers sensitivity to quality issues. This seems to be supported by the detailed inspection guidelines which have been published to help schools to prepare for their inspection (Ouston/Davies, 1998; Ouston *et al.*, 1999; Kotthoff, 2003).

There are, however, other undesirable or even contra-productive effects of school inspections which are also reported in the same studies: these are e.g. the extremely high workload and stress of the teachers in the run up to the inspection, the lack of feedback that teachers receive from the inspectors and the increasing gap between the «ordinary» classroom teachers and the headteacher and/or the senior management who are the key figures in the inspection process. Whether school inspections fulfill a quality developing function or not seems to depend largely on the starting point of each individual school: weak schools and in particular so called «schools requiring special measures» which receive long-term support and consultation through specialised «School Improvement Teams» are most likely to improve considerably and over time. There is, however, not enough empirical evidence (yet) to support the claim that school inspections will stimulate long-term internal school improvement processes and support the professionalization of teachers in the majority of schools. This is due to the fact that the vast number of «average» schools have got nothing to win or lose through inspection results and that inspections happen too rarely (usually every three years) to have a permanent quality developing effect on schools. Finally, it has been shown that good schools in particular do not rate inspections very highly in terms of their value for money because they feel that in most cases the inspection results do not come as a surprise and are therefore hardly helpful for internal school development processes (Kotthoff, 2003: 362-363).

Compared with the aims of school evaluation in England, the Swedish evaluation system pursues rather different aims: Firstly, it has to be pointed out that, although the responsibility for the evaluation of schools lies formally with the National Agency for Schools (*Skolverket*), it has been made clear from the very beginning of the National Agency that the main responsibility for the evaluation of schools should be in the hands of the schools and the 289 municipalities (*kommuner*). A direct intervention into the quality controlling and developing function of the municipalities and the schools should therefore be limited to very exceptional cases.[2] Secondly, the main aim of school evaluation in Sweden should not be the control but rather the development of quality (evaluation *as* school improvement). Thirdly, by obliging each school to evaluate its quality on a regular and systematic basis the Swedish state wants to increase the sensitivity of its teachers for quality issues in schools and make a contribution to their professionalization.

If we confront these aims of the Swedish evaluation system with its achieved effects which have been proved in empirical studies the following picture emerges: with regard to the allocation of evaluative responsibilities between *Skolverket* and the municipalities and/or the schools there is a fundamental question of how *Skolverket* can fulfill its supervisory function for the control and development of quality in schools while at the same time respecting the autonomy of the municipalities and the schools in this respect. This dilemma which was already identified in an international OECD report on school evaluation in 1995 is still not solved satisfactorily (OECD/CERI, 1995: 131). An indicator for this key problem is the fact that even in 2000, i.e. almost ten years after the foundation of *Skolverket*, only two thirds of the 289 municipalities had ever written their obligatory annual quality report for *Skolverket* (Skolverket, 2001: 5) which shows quite clearly that one third of the Swedish municipalities had not yet accepted their quality assuring function. A related problem has been established by Swedish researchers who found that the cooperation between the municipalities and the schools with regard to quality questions is quite poor and that more than 90% of Swedish teachers do not trust the results and insights gained through the municipal evaluation of the school system (Granström & Lander, 1999: 39). More recent research suggests that the schools themselves have accepted their evaluative role: self-evaluation seems to be a regular feature in Swedish schools now, although there is still evidence that this evaluation is quite often rather unreflected and not always very systematic due

to the teachers' lack of theoretical and methodological knowledge on evaluation. Negative side-effects of the increasing evaluative responsibility of schools are primarily that classroom teachers complain about increasing workloads which are due to additional meetings to plan evaluations and to write up evaluation reports. In addition, case studies show that exclusive reliance on school internal evaluation may aggravate quality differences between schools because the municipal feedback to internal school evaluation as well as the headteachers' competence in setting up effective evaluation structures will vary substantially between municipalities and schools (Kotthoff, 2003: 315-318).

While the design of the Swedish evaluation system certainly suggests that the accountability function of school evaluation is subordinate to its quality development function (e.g. importance of self-evaluation, renunciation of school league tables, no public «naming and shaming» of individual schools etc.), the question whether the evaluation of schools has helped them to improve their quality is difficult to assess at the moment. On the basis of their analyses of Swedish and other Scandinavian (case) studies on the relationship of evaluation and school improvement Lander & Ekholm conclude rather sceptically: «The examples from research cited above are not success stories about the use of evaluation for school improvement. Teachers have often reacted with suspicion and resistance to such state initiatives. [...] Case studies of successful schools suggest that school evaluation does not play a leading role in the school improvement process» (Lander/Ekholm, 1998: 1131). The results of my own case studies on the effects of evaluation in three Swedish schools do suggest however that evaluation does lead to school improvement in individual schools. This seems to be particularly the case in those schools where evaluation is done systematically and continuously, e.g. in the form of a «quality cycle» (Kotthoff, 2003: 302-314).

The case studies also provide empirical evidence with regard to the third aim of school evaluation, the sensitization of teachers to quality issues and their professionalization: according to the interviewed headteachers, systematic evaluation intensifies the teachers' discourse on quality within schools considerably. One positive side-effect of this intensified discourse is that there is more interchange of information in general which in turn can raise teachers' awareness of areas of concern and can help them to identify good practice within their school. Another positive effect, is according to the headteachers, that schools can use evaluation reports to document their work, make their school life and

the results of their pupils' work more transparent to the parents which is a prerequisite for their cooperation. Finally, the headteachers report unanimously that internal school evaluation has also had a positive effect on the professionalization of their staff. Amongst the activities which generate the most important learning effects and which improve the professional repertoires of their teachers most noticeably the headteachers mention in particular the planning and preparation of evaluations, the collection and interpretation of evaluation data and the development of practical measures to improve quality. When interpreting these rather positive statements we should bear in mind though that the positive assessment of the effects of school evaluation depends largely on the perspective of the stakeholder, which means that headteachers rate evaluation much more positively than those directly involved in the teaching process, i.e. the students and teachers (Kotthoff, 2003: 370-74). A final observation on this point is contributed by Granström and Lander who report that the professionalizing influence of evaluations is felt particularly strongly when the internal evaluation is supported by a «critical friend» or through peer group reviews (Granström/Lander, 1998: 8).

3. Discrepancies between *intended* and *achieved* effects of school evaluation: conclusions

If we try to summarize our findings with regard to the discrepancies between aims and effects of school evaluation in England and Sweden, three points are of particular importance:

Firstly, in both education systems the main aim of school evaluation is the improvement of the quality of schools. Apart from this however, the results of our analysis suggest that the aims of school evaluation do not always have to be complementary. They can even be paradoxical or contradictory to each other, if external evaluation is used to criticize schools negatively, linking external evaluation for control with internal assessment for development. This paradox becomes most evident in the English evaluation system where inspection is supposed to exercise a controlling and normalizing effect on individual schools while at the same time developing the professional autonomy of the school and the teachers. However, non-complementary aims can also exist in internal evaluation systems when internal evaluation data which have been generat-

ed by teachers to promote school development are used in external evaluations, in order to criticize and judge this school.

Secondly, the aims of school evaluation can be achieved in different ways: This can be shown, for example, with regard to the sensitization of teachers to quality issues which is one of the main aims of evaluation in both education systems. On the one hand the English example demonstrates that this sensitization can be stimulated from «outside» through external school inspections. On the other hand the Swedish evaluation system shows quite convincingly that this effect can also be achieved through internal forms of evaluation which intensify the inner school discourse on quality.

Thirdly, the two analysed cases show that measures of school evaluation can fail to achieve their desired effects and that they can have unintentional or even contra-productive side-effects. A particularly striking example for *failed* aims is the apparent difficulty of both evaluation systems to translate the results of internal or external school evaluation into concrete and quality-improving school development processes. Evidence for *unintentional* side effects of school evaluation are for example the increased work and stress burden of English classroom teachers in the run up to school inspections and the observation that school evaluation can exercize a standardizing influence on teaching-learning processes and/or on the school curricula. Finally, the possibility of *contra-productive* effects is indicated by the observation that the evaluation of schools can under certain circumstances, e.g. if it is combined with financial sanctions, disadvantage weak schools and therefore endanger the equity within the school system.

The comparative analysis of the two evaluation systems has clearly shown that there are considerable discrepancies between *intended* and *achieved* effects of school evaluation. Any attempt to provide a comprehensive analysis of school evaluation in different education systems must therefore necessarily take account of the unpredicted and unwanted «side»-effects of school evaluation as well. I have shown elsewhere that the well-established analytical differentiations between «monetary» and «non-monetary» «costs» and «benefits» which have been developed in the field of economics of education (cp. Psacharopoulos, 1987) can serve as useful analytical «tools» in this context and can help to shed some light on the question whether the provable quality improving effects of school evaluation (i.e. «benefits») justify its «costs» (cp. Kotthoff, 2003: 375-390; Kotthoff, 2004). In this contribution however, I want to shift

the focus and try to formulate recommendations which could help to reduce discrepancies between intended and achieved effects of school evaluation and to minimize its unintentional, unwanted or even contra-productive side-effects.

4. Improving school evaluation: recommendations

From an analytical perspective and on the basis of the empirical evidence available, it seems possible to differentiate between three groups of recommendations:

I. Recommendations which aim at raising the level of acceptance of school evaluation amongst teachers

On the basis of the empirical evidence presented it seems most important that school evaluation reaches high levels of acceptance amongst the teaching profession. This in turn can be raised, if schools are granted high degrees of autonomy in the first place. Teachers will only be convinced about the necessity of educational evaluation if they enjoy high degrees of professional freedom in their work place.

Secondly, from the teachers' perspective the credibility and acceptance of (external) school evaluation is substantially raised if the (external) evaluators can bring their own practical school experience and teaching expertise to the inspection. The English example shows clearly that external evaluation is severely mortgaged from the teachers' perspective if it is carried out by privatized evaluation agencies.

In addition, the English example shows that the teachers' acceptance of negative inspection results will be raised if the local particularities of individual schools are taken into account and the individual progress of the school is appreciated from a "value added" perspective.

Thirdly, judging from the Swedish and English experiences, it will be of the utmost importance to convince teachers of the usefulness of school evaluation and to let them experience this usefulness in their daily professional practice as teachers. This requires that the knowledge which is generated through school evaluation becomes more practically relevant, concrete and instructive for the improvement of schools and teaching-learning processes. The establishment of a so-called school internal "evaluation culture" which is currently demanded in many European

education systems and which will lead to a substantial additional work load for the teachers will only be accepted by the teachers if it leads to concrete and perceptible improvement in the quality of the school and if it will help at least in the medium term to make their work easier.

Finally, levels of acceptance will be raised amongst the teaching profession if the school authorities are also prepared to be evaluated themselves and if they can prove to the teachers that they themselves can deal with negative evaluation results in a constructive manner.

2. Conceptual recommendations which aim at raising the quality-improving effect of school evaluation

With regard to the conceptual design of quality-raising evaluation systems, the empirical evidence presented seems to suggest that there needs to be an optimal balance between internal and external evaluation strategies and instruments. This recommendation is also supported by recent evaluation experiences from Switzerland. According to the Swiss school researcher Strittmatter, the empirical evidence seems to suggest that the more elaborate (in terms of time and emotional investment) and important (in terms of sanctions or benefits) external evaluations are, the less energy and time will be made available for any serious, professional internal evaluation. Elaborate and exhaustive external evaluations which are considered as threatening will, according to Strittmatter (2006), always result in internal evaluations being carried out purely as an alibi or not at all.

Similarly, the balance between the external standardization of what is defined as «good» school quality on the one hand and the acknowledgement of the professional autonomy of the teachers on the other needs to be constantly reflected and adjusted. Teachers have got their own legitimate notions on what counts as "good" quality, which they have developed in the course of their professional career. The external definition of quality standards is therefore quite often perceived by teachers as an intervention into their professional expertise and judgement. This is particularly the case in those education systems which value professional autonomy of their teachers very highly.

With regard to the conceptual design of evaluation systems, the presented empirical evidence from England and Sweden seems to suggest that internal evaluations of schools will benefit if they are supported systematically through different forms of "metaevaluation". One type

of metaevaluation are quality audits which are carried out by the school authorities to make sure that the schools' internal evaluation systems fulfil minimum procedural standards.

Finally, evaluation systems should be designed in a way which protects individual schools from so-called «naming and shaming» effects. As long as sensitive data on particular schools are leaked to the public without protection, schools will react defensively by faking positive results and/or by simulating alibi efforts. This is even more true if negative results of external evaluations (e.g. negative inspection results) are indirectly linked to financial sanctions as is the case in England.

3. Practical recommendations

In the face of the unpredictability of the effects of school evaluation which has become obvious in the English and Swedish case, a first practical recommendation should point out that newly developed evaluation systems should be principally open for revisions and can react flexibly to new (international) experiences. Secondly, evaluation criteria should be made transparent. The *Handbook for Inspection* (OFSTED, 1995), for example, which has been developed by the English school inspectorate OFSTED and which gives the teachers a comprehensive list and explanation of evaluation criteria has been warmly welcomed by most headteachers in England who regard the handbook as a very useful tool for school development. A third recommendation underlines the importance of offering the schools concrete help and assistance in putting the recommendations of the external evaluators into practice. Again, with reference to England, the rapid improvement of the so-called *schools requiring special measures* show convincingly how much "weak" schools in particular can profit from an intensive period of coaching and advising after an external evaluation. Finally, the introduction of educational evaluation is a social process which has to be established very carefully in any given education system. The hesitant reaction of many Swedish schools and municipalities with regard to their obligation to introduce regular and systematic (internal) evaluations shows the necessity to introduce new evaluation methods and instruments cautiously and in small and sensitive steps.

Notes

[1] The conclusions in this section are based on the results of my comparative study on school evaluation systems which were published in 2003 (Kotthoff, 2003). The publication also includes a detailed analysis of the main evaluation concepts and instruments in England and Sweden which will not be presented here.

[2] However, in 2004 *Skolverket* started to inspect schools and municipalities directly (Skolverket, 2004). Presently it is too early to assess the effects of this development and it remains to be seen how this shift of responsibility will affect the schools' and the municipalities' autonomy in quality issues.

References

Ball, S. (1998), «Educational Studies, Policy Entrepeneurship and Social Theory», in Slee, R., Weiner, G. Tomlinson, S. (eds.), *School Effectiveness for whom? Challenges to the School Effectiveness and School Improvement Movements*. London, Falmer, pp. 70-83.

Clarke, D.L. (1985), «The context of the paradigmatic shift», in Lincoln, Y.S. (ed.), *Organizational Theory and Inquiry*, Beverly Hills, pp. 43ff.

Coe, R., Fitz-Gibbon, C.T. (1998), «School Effectiveness Research: Criticisms and Recommendations», in *Oxford Review of Education* 24/1998, pp. 421-438.

Crowson, R.L. *et al.*, (1996), «The Politics of Education and the New Institutionalism», *The 1995 Yearbook of the Politics of Education Society*, London.

Dimaggio, P.J. & Powell W.W. (eds.) (1991), *The New Institutionalism in Organizational Analysis*, Chicago, the University of Chicago Press.

Granström, K. & Lander, R. (1998), *The Swedish National Report for the European Pilot Project Evaluating Quality in School Education*, Stockholm, Skolverket.

Granström, K. & Lander, R. (1999), *Läraropinion om grundskolans styrningskultur och lokala utvärdering. Resultat från en riksrepresentativ enkät.* IPD-rapport 1999: 15, Göteborg universitet, Institutionen för pedagogik och didaktik.

Granström, K. & Lander, R. (2000), «Samverkan och planering. Två indikatorer på skolans styrningskultur», in *Utbildning & Demokrati* 2000, Vol. 9, Nr. 2, pp. 75-85.

Holly, P. & Hopkins, D. (1988), «Evaluation and school improvement» in *Cambridge Journal of Education*, 18, 2, pp. 221-245.

Hopkins, D. (1989), *Evaluation for School Development*, Milton Keynes, Open University Press.

Kotthoff, H.G. (2003), *Bessere Schulen durch Evaluation. Internationale Erfahrungen*, Münster/New York, Waxmann.

Kotthoff, H.G. (2004), «The Rise of Educational Evaluation in Europe: Panacea for the Improvement of Quality in Schools? Cases from England, Sweden and Germany», in Buk-Berge, E. *et al.* (eds.), *Education Across Borders: Comparative Studies.* Oslo, Didakta, pp. 189-303.

Lander, R. & Ekholm, M. (1998), «School Evaluation and Improvement: A Scandinavian View» in Hargreaves, A., Liebermann, A., Fullan, M., Hopkins, D. (eds.), *Kluwer International Handbook of Educational Change*, Dordrecht u.a., Kluwer Academic, pp. 1119-1134.

Lauder, H. *et al.* (1998), «Models of Effective Schooling: Limits and Capabilities», in Slee, R., Weiner, G. Tomlinson, S. (eds.), *School Effectiveness for whom? Challenges to the School Effectiveness and School Improvement Movements*. London, Falmer, pp. 51-69.

OECD/CERI (1995), *Schools under Scrutiny*, Paris, OECD.

OFSTED (1995), *The OFSTED Handbook: Guidance on the Inspection of Primary Schools*, London, HMSO.

Ouston, J. & Davies, J. (1998), «OFSTED and afterwards? Schools' responses to inspection» in Earley, P. (ed.), *School Improvement* After *Inspection?*, London, Paul Chapman Publishing, pp. 13-24.

Ouston, J. *et al.* (1999), «Memorandum» from Brian Fidler, University of Reading, Janet Ouston, Peter Earley, Neill, Ferguson and Jackie Davie, Institute of Education, University of London, in Education and Employment Committee, *Fourth Report: The Work of OFSTED: Volume III: Appendice*, London, The Stationary Office, pp. 284-87.

Psacharopoulos, G. (ed.) (1987), *Economics of Education: Research and Studies*. Oxford, Pergamon.

Sammons, P. *et al.* (2001), «Fairer Comparisons of Schools: The Role of Effectiveness Research», in *Promoting Improvement*, paper presented at *The meaning of Quality in Education Conference* in Karlstad, Sweden, 02.04. April.

Scott, W.R. (1995), *Institutions and Organisations: Theory and Research*, Thousand Oaks, Sage.

Scriven, M. (1991), *Evaluation Thesaurus* (4th edition), Newbury Park, Sage.

Skolverket (2001), *Dialogue for improvement: The governing and monitoring of child care and education – the Swedish system*, Stockholm, Skolverket.

Skolverket (2004), *The Swedish National Agency of Education's Educational Inspectorate: Inspecting for Improvement*, Stockholm, Fritzes.

Strittmatter, A. (2006, *in print*), «Schulinspektion zwischen Polizeiradar und Solbad», in Böttcher, W., Kotthoff, H.-G., (Hg.), *Schulinspektion zwischen Rechenschaftslegung und Qualitätsentwicklung: internationale Erfahrungen*. Münster, Waxmann.

Teddlie, C. & Reynolds, D. (eds.) (2000), *The International Handbook of School Effectiveness*, London, New York, Falmer Press.

Tyack, D. & Tobin, W., «The "grammar of Schooling": Why has it been so hard to change?» in *American Educational Research Journal* 31/1994, pp. 453-479.

Walford, G. (1996), «School Choice and the Quasi-market in England and Wales», in *Oxford Studies in Comparative Education*, Vol. 6, 1, pp. 49-62.

Whitty, G. *et al.* (1998), *Devolution and Choice in Education: the School, the State, and the Market*, Buckingham, Open University Press.

ΠΕΡΙΛΗΨΗ

Αποκλίσεις στόχων και αποτελεσμάτων στη σχολική αξιολόγηση: Ερευνητικά ευρήματα και προτάσεις

Σύμφωνα με την Οργανωσιακή Θεωρία και το Νέο Θεσμισμό (New Institutionalism), κοινωνικοί θεσμοί μεγάλης πολυπλοκότητας —όπως τα εκπαιδευτικά ιδρύματα— δεν ανταποκρίνονται πάντα σε μεταρρυθμιστικούς προγραμματισμούς, καθώς, συχνά, οι ξεχωριστοί actors που τους απαρτίζουν δρουν ανορθολογικά. Οι θεωρίες αυτές επιχειρούν να ερμηνεύσουν τις αποκλίσεις μεταξύ τεθέντων στόχων και πραγματικών αποτελεσμάτων, διερευνώντας τους τρόπους με τους οποίους οι εισαγόμενες σε έναν οργανισμό μεταρρυθμίσεις απενεργοποιούνται ή διαφοροποιούνται, προσαρμοζόμενες στις κατεστημένες «νόρμες», διαδικασίες και πρακτικές λειτουργίας του.

Με γνώμονα τις παραπάνω θέσεις, το άρθρο διερευνά τις συνέπειες που είχε για την εκπαιδευτική αξιολόγηση η «αλλαγή παραδείγματος», από το γραφειοκρατικό έλεγχο των εισροών στη σημερινή τάση επικέντρωσης στα αποτελέσματα. Ειδικότερα αποτιμάται η απόκλιση στόχων (αυτών που καταγράφονται στα επίσημα έγγραφα των εκπαιδευτικών αρχών ως δηλωμένη επιδίωξη των μεταρρυθμιστικών μέτρων) και πραγματικών αποτελεσμάτων (όπως εξάγονται από τα ευρήματα των ερευνητικών μελετών). Ως μελέτες περίπτωσης χρησιμοποιούνται οι αλλαγές στον τρόπο αξιολόγησης των σχολείων στα εκπαιδευτικά συστήματα της Αγγλίας και της Σουηδίας· δύο χωρών που στη διάρκεια των τελευταίων ετών μεταρρύθμισαν ριζικά το σύστημα εκπαιδευτικής αξιολόγησής τους και ταυτόχρονα αντιπροσωπεύουν δύο διαφορετικές εκπαιδευτικές παραδόσεις, την «αποκεντρωτική» βρετανική και τη «συγκεντρωτική» σκανδιναβική.

Η ανάλυση καταλήγει σε μια σειρά συμπερασμάτων: (i) Οι επίσημοι στόχοι της αξιολόγησης δε λειτουργούν πάντα συμπληρωματικά μεταξύ τους. Απεναντίας, συχνά αντιφάσκουν, όπως συμβαίνει με την παράλληλη επιδίωξη **ελέγχου** από τη μία και **βελτίωσης** της ποιότητας από την άλλη. (ii) Ο (ίδιος) στόχος μπορεί να επιδιωχθεί με διαφορετικές μεθόδους. Κάτι τέτοιο παρατηρείται στον κοινό στόχο για ευαισθητοποίηση των εκπαιδευτικών σε θέματα ποιότητας, ο οποίος στην Αγγλία προσεγγίζεται μέσω εξωτερικών επιθεωρήσεων, ενώ στη Σουηδία μέσω της αυτοαξιολόγησης των σχολείων. (iii) Οι μεταρρυθμίσεις στις δύο χώρες συνοδεύονται συχνά τόσο από αδυναμία επίτευξης των στόχων τους όσο και από μη προβλεφθείσες ή/και ανεπιθύμητες παρενέργειες. Παράδειγμα μη επιτευχθέντος στόχου αποτελεί η —μη ερευνητικά τεκμηριωμένη— ποιοτική βελτίωση. Μη προβλεφθείσα συνέπεια συνιστά το αυξημένο εργασιακό βάρος και άγχος του εκπαιδευτικού προσωπικού και αντιβαίνουσα την αρχική στοχοθεσία κατάληξη συνιστά η περαιτέρω υποβάθμιση των μέτριας/κακής φήμης σχολείων.

Στηριζόμενο σε τέτοια ερευνητικά συμπεράσματα το άρθρο καταλήγει σε τρεις κατη-
γορίες συστάσεων/προτάσεων. Η πρώτη απ' αυτές τονίζει τη σημασία της αποδοχής των
διαδικασιών αξιολόγησης από τους εκπαιδευτικούς. Επισημαίνεται ότι το μέγεθος της
αποδοχής εξαρτάται από το βαθμό επαγγελματικής ελευθερίας των εκπαιδευτικών και
αυτονομίας των σχολείων. Όσον αφορά τις εξωτερικές αξιολογήσεις, η αξιοπιστία τους
(μεταξύ των εκπαιδευτικών) είναι πολύ μεγαλύτερη όταν οι αξιολογητές επιδεικνύουν
σημαντική σχολική εμπειρία, ενώ, αντίθετα, οι αξιολογήσεις από ιδιωτικές εταιρείες αντι-
μετωπίζονται με δυσπιστία. Η αποδοχή της αξιολόγησης ενισχύεται όταν τα συμπεράσμα-
τά της παραπέμπουν σε συγκεκριμένες, προσαρμοσμένες στις ιδιαιτερότητες του κάθε
σχολείου, πρακτικά εφαρμόσιμες λύσεις και όχι σε στείρα θεωρητικολογία. Η επιζητούμε-
νη σε πολλά ευρωπαϊκά εκπαιδευτικά συστήματα «αξιολογική κουλτούρα» είναι περισ-
σότερο εφικτή στην περίπτωση που τα αποτελέσματα των αξιολογήσεων επιδρούν βελ-
τιωτικά **και** στη δουλειά των εκπαιδευτικών. Τέλος, τα επίπεδα αποδοχής είναι υψηλότε-
ρα όταν οι σχολικές ηγεσίες είναι και οι ίδιες αντικείμενα αξιολόγησης.

Για το στόχο βελτίωσης της ποιότητας απαραίτητη θεωρείται η επίτευξη μιας ιδανικής
ισορροπίας μεταξύ εσωτερικής και εξωτερικής αξιολόγησης. Αν η πρώτη φαντάζει απει-
λητική για το σχολείο και το προσωπικό του, η δεύτερη τείνει να λαμβάνει διακοσμητικό
χαρακτήρα ή να λειτουργεί ως άλλοθι. Το ίδιο ισχύει και όσον αφορά τις διαφορετικές
εννοιολογήσεις της «ποιότητας» όπως αυτές διαμορφώνονται, από τη μια, από τις εμπει-
ρίες των εκπαιδευτικών και, από την άλλη, από τους —βάσει σταθμισμένων μετρήσεων—
ορισμούς των εξωτερικών φορέων αξιολόγησης. Η γεφύρωση τέτοιων χασμάτων μπορεί
να επιτευχθεί μέσω μορφών «μετα-αξιολόγησης».

Όσον αφορά τις ακολουθητέες πρακτικές, υπογραμμίζεται ότι η αμφίβολη επιτυχία των
σχετικών με την εκπαιδευτική αξιολόγηση μεταρρυθμίσεων θα πρέπει να καταστήσει τους
διαμορφωτές των εκπαιδευτικών πολιτικών ανοιχτούς σε αναθεωρήσεις και έτοιμους να
λάβουν υπόψη τους τις εμπειρίες που συγκεντρώνονται από τις εφαρμογές της διεθνώς.
Αναγκαίος είναι ο κατανοητός τρόπος πληροφόρησης για τα χρησιμοποιούμενα αξιολο-
γικά κριτήρια και η παροχή βοήθειας προς τα σχολεία για το πώς θα μεταφράζουν τα
αποτελέσματα των αξιολογήσεων σε σχολική πράξη. Σε κάθε περίπτωση, καθώς η εκπαι-
δευτική αξιολόγηση αποτελεί εκτός των άλλων και μια (γενικότερη) κοινωνική εξέλιξη, η
εισαγωγή της είναι αναγκαίο να γίνεται μέσω και μιας διαδικασίας προσεκτικών και συ-
νετών βημάτων.