

Working Paper F1

**ALL UNINTENDED EFFECTS WILL BE REMOVED BY THE
AUTHORITIES!
OR: NEW PARADIGMS FOR THE UNDERSTANDING OF
EVALUATIVE KNOWLEDGE?¹**

**Professor Peter Dahler-Larsen, PhD
Department of Political Science and Public Management
University of Southern Denmark**

DRAFT ONLY: NOT FOR QUOTATION WITHOUT PERMISSION

Abstract

Current trends in the sociology of knowledge suggests that conventional assumptions about rationality, instrumentality and linearity in the public use of evaluative knowledge should be fundamentally revised. If yes, this has serious implications for discussions of how and when to publish evaluative data on schools and "school performance". More specifically, my contribution will touch upon the following questions: What degree of consensus is there about what school quality is? What are some of the "constitutive effects" of school performance data? On top of public availability of data, what difference does the framing of data make? How stable/unstable are the images of schools produced by publicly available data? In the final section of the paper, a few normative implications are drawn concerning the publication of evaluative school data.

Introduction

The problem of publishing school data is an important one. According to its advocates and its opponents, it may have tremendous positive or negative effects on schools, respectively. I shall approach the problem from a specific perspective, largely defined by the fact that I am a researcher in evaluation and my affiliation is with a political science department.

Quite fundamentally, I do not view a so-called "code of ethics for the publication of school data" as related to an ethical problematique, but rather as a political project. Aristotle said that politics is the most architectonic of the sciences that concerns the human being. To

¹ Paper presented at the NERA conference Turku, Finland, March 14, 2007

follow the metaphor, if the house is badly built, we will have to spend our time doing insufficient fix-ups in the house (Castoriadis 1997: 111). In this sense, politics supersedes ethics. The question is how we as democratic human beings want to construct our institutions. When we decide that something should be said and something should not be said about schools in public, we are affecting the public sphere and in certain sense also the fate of society (Castoriadis 1997).

Before going into details with assumptions about the use of school data in the public sphere, I shall briefly put the discussion also in the context of the use of *evaluation* data.

The open-ended use of evaluation

People involved in evaluation often cling to assumptions about the use of evaluation which are normative and uncontested (Shadish et al 1991). This is partly due to the very definition and self-understanding of evaluation, which is that evaluation as an activity aimed at, among other things, improving the quality of services, if not in a larger perspective, social betterment (Mark, Henry, Julnes 2000). This built-in intention in the very definition of evaluation sometimes leads to a blind-spot towards the many types of consequences which evaluation may have under different circumstances. Evaluation research has developed a typology of forms of use, such as “accountability”, “learning”, “enlightenment”, “strategic”, “tactical”, “symbolic” etc. However, the typology is based on relatively unclear and partly overlapping categories, and they relate to different aspects of evaluation and different aspects of outcomes of evaluation and different analytical levels (Mark and Henry 2004). It also not clearly separates normative concerns (what is good use) with purposes of evaluation (what is intended to happen) and with empirical descriptions of use (what actually happens).

If there is a consensus in the field about utilization, it seems to be that use of evaluation can be short-term and long-term, it can related to evaluation processes as well as evaluation results, use can be more or less intended, and can take place on several levels. The use of evaluation is also highly dependent on complex public, political, historical, institutional and organizational factors. The frightening theoretical implication – of which few dare to think – is that the use of evaluation may be under very little control and may have no logical end point. We are far from a situation where “first we do the evaluation, then we use it, and that’s it” can be used as our guiding paradigm.

Nevertheless, the “opening up” and “flexibilization” of the utilization of evaluation makes very good sense in the light of findings within the more general and abstract recent theories of knowledge.

Theories of knowledge as socially productive²

A constitutive perspective on knowledge suggests that knowledge is not an essence, but a social accomplishment. The same is true for the effects of knowledge. Knowledge is “open-ended” in the sense that its “use” is not an inherent property, but depends on the articulation, representation and appropriation of knowledge in particular contexts (Woolgar 2004). Effects thus mean social accomplishments rather than causal outcomes.

Stehr (1994 p. 95) defines knowledge as a capacity for social action. He emphasizes the socially productive role of knowledge in contemporary society with regard to not only “the appropriation of appropriated nature” (p. 103) but also to the organization of the social order (rather than the material interaction with nature). Briefly stated, under modernity knowledge shapes social relations.

However, he immediately points out that knowledge as a capacity for action often assumes a bureaucratic or otherwise smoothened social order with linear structures which are “prepared for data processing” (p. 103). This assumption is, of course, often out of place. Therefore, knowledge rarely produces opportunities for perfect planning of a particular social intervention. Instead, the typical result is *an increasingly fragile social order* (Stehr 2001; Giddens 1990: 45). This means an order which is constantly ripe for change due to its ongoing integration and interaction with knowledge.

Much in the same spirit, Giddens (1990) explains how social practices are described and transformed in the light of incoming knowledge. Formal political and administrative systems, as well as a number of informal reflexive processes, operate with a mutual interplay between knowledge, data and social relations. Modernity is already fundamentally reflexive and sociological, to paraphrase Giddens (1990).

A number of factors help explain why the transmission of knowledge into changing social relations is far from linear and straightforward. The first is unequal distributions of power. Another factor is the vast influence from backfiring consequences of earlier applications of knowledge. Another is that changes in values follow other logics than changes in the systems which produce knowledge. Therefore streams of knowledge are confronted with value changes with which they are never fully synchronized. Finally, as a sort of synthesizing point, the reflexivity of modern social relations is itself a non-linear and thus de-stabilizing factor (Giddens 1990).

While systematically produced knowledge could earlier play a social role similar to that of traditional religion with respect to authority and certainty, uncertainties in knowledge are becoming clearer today. In today’s knowledge society it also becomes obvious that each new piece of knowledge does not always reduce ignorance and exclude alternative views.

² This section partly reproduced from my chapter Constitutive effects of performance indicator systems. To appear in S. Kushner (Ed.): Dilemmas of Engagement: Evaluation Development under New Public Management and the New Politics. New York: Elsevier.

Instead, there are complex social forces seeking control over different knowledge productions, and knowledge is not only accumulated to approximate truth. Some pieces of knowledge are partly contradictory and suggesting or recommending different social pathways. Different knowledge-producing systems on societal or institutional levels (such as governments and schools) interact in spiralling regulatory logics which paradoxically produce more risk and more need for control (Rothstein, Huber and Gaskell 2006).

Applying these general observations about knowledge and fragile social orders to evaluation processes, it becomes clear that classical notions of use of evaluative knowledge operate with very restricted assumptions. Instrumental use is a feed back from evaluation which relates selectively to only one particular aspect of social reality, ie. the one which has to do with decisions to improve the quality of the object under evaluation.

The reflexivity and social productivity of knowledge suggested by Giddens and Stehr suggest a number of other possible options. People under evaluation are not only objects, but subjects who construct their own knowledge about the evaluation, its process and results. In a similar vein, the public arena comprises a number of different groups with different agendas and different interests (Boyle, Breul and Dahler-Larsen 2008) related to evaluation data from schools. They respond to evaluation data in different ways. In other words, evaluation data may not be a truth with a given effect. It may be a construction which leads to other constructions.

With these ideas in mind, I shall now move to the discussion about a code of ethics concerning the publication of school data.³

A code of ethics? Or evaluation and politics in an open society?

In the following, I shall discuss some of the main points made concerning the normative prescriptions about the publication or non-publication of school indicators (SPIs, school performance indicators) which I read in Rowe (2000), Visscher et al (2000), Goldstein (1997), and Goldstein and Myers (1996).

I do apologize for not fully representing the views of each of these authors. Neither do I claim that any view in the following represents a consensus between these authors. My intention is merely to purify and distil a few central assumptions and proposition for the purpose of discussion. I shall rely a bit more on Rowe (2000) than on the other references, partly because I think he summarizes and builds on the work of Goldstein, and because I think Rowe expresses a more clear and less nuanced view than Visscher (who also refers to the work of Goldstein). Visscher seems to advocate careful and nuanced publication, where as Rowe leans a bit more towards non-publication as the first choice.

Perhaps it is fair to clarify the perspective from which I am reading Rowe's work. I am full of sympathy for the idea of taking publication of school data very seriously. But the good

³ I am grateful to Hannuh Simola for directing my attention to this specific discussion.

intentions do not guarantee a healthy underlying reasoning. I shall admit openly that there is an element of playing devil's advocate in my reading of the recommendations about how to handle school performance indicators. I do find a critical discussion necessary, because Rowe (2000: 85-86) suggests that "It may be necessary to establish formal regulatory mechanisms to ensure compliance" [with the code of ethics]. In other words, it is suggested to go far beyond personal ethical considerations; instead clearly legal and political means are proposed here, which should be carefully considered before enactment. My reservations have to do with the political legitimacy of the proposals and with the underlying assumptions about the use and consequences of evaluation, and whether these assumptions hold across cultural and national contexts. But now to the substance.

The main thrust of Rowe's position seems to be that the ability of educational performance indicators "to reflect objective reality may be extremely limited, and its publication may therefore case both misleading and incorrect inferences about schools and "school effectiveness" to be drawn. *In such circumstances, there is strong case for withholding publication*" (Rowe, 2000: 83). Next, "if publication cannot be prevented, the then information should have appropriate warnings about its interpretation." (Rowe, 2000: 83).

This general view can, according to my reading, be disintegrated into a number of central propositions.

1. Public disclosure is not an absolute principle (Rowe, 2000: 83).
2. There can be considerable individual and institutional harm of publishing school performance indicators (Rowe 2000: 77).
3. School performance indicators are not valid indicators of school quality.
 - a) this is especially true of raw test scores (without adjustments for student intake).
4. In addition, there will be a number of negative unintended consequences of publication, such as
 - a) over-emphasis on content that is tested
 - b) increased social segregation of schools (Rowe 2000:76)
5. There is uncertainty and instability in the ranking of schools based on school performance indicators.
6. "The mere fact of publishing information by an official body lends it credence" (Rowe 2000: 86).

I shall now critically discuss each of these six propositions and then turn to a few critical assumptions, which I think are more tacitly present in Rowe's work.

1. Public disclosure is not an absolute principle (Rowe 2000)

That is correct, but an open society (Popper 1969) seems to be our best bulwark against tendencies towards totalitarianism and abuse of power. In democracies, we usually tend to hold public disclosure of publicly relevant information in high regard, unless national security or the privacy rights of individuals are in jeopardy. With some exceptions for individualized information about teachers, none of these two principles can be invoked to legitimize non-disclosure of SPIs. According to a popperian view, one of the best ways to shed light on prejudices, ideologies, wrong theories and problematic data is to throw them into the public arena for public and critical scrutiny. The fact that some think that some information can be misused does in itself not justify withholding this information from the public arena.

Rowe (2000: 85) assumes that information should be accurate and relevant before the decision to publish is made, but this leaves the question of who makes this decision.

We are in democracies very attentive against people who say that for a number of reasons, information about public authorities cannot be made publicly available. We would usually prefer to regulate the publication of this information by legal and political means, not only by reference to ethical concerns among those who produce the data.

The idea that the public is generally not able to handle information properly and take acceptable action is democratically a very controversial idea and should only be relied on if we have very carefully considered alternatives.

Our democratic propensities would, all other things being equal, lead us to require publication of information, come what may, because the democratic ramifications of alternatives are often not acceptable once we think them through.

May I also remind us all, that there are actually some in society who thinks that public disclosure is an absolute principle, or pretty close to that, with very few exceptions. Some of these are auditors. Others are public information officers. Still others are ombudspersons. Others are political oppositions. Our concerns about the publication of school data should thus not be made in total isolation from other mechanisms we have to enhance democratic processes.

2. There can be considerable individual and institutional harm of publishing school performance indicators (Rowe 2000: 77)

The argument goes that league tables tend to produce winners and losers, and “the results for the recipients of the “failing/ineffective” label can be catastrophic” (Rowe, 2000: 76), including public shaming, loss of jobs, closing of institutions, and perhaps deterioration of the public status of the teaching profession. These things should obviously be taken into account. However, the fact that information may have negative consequences for a public institution and its employees does not generally constitute a reason for not making the

information available. Do we accept that the military does not provide information about military operations because of the negative consequences for military staff? No. If yes, we would institute a general insurance against political decision-making and priorities among public institutions.

Our argument is not that the consequences are negative for institutions and staff in themselves. We must argue either that the quality of services is threatened by publication of SPIs, or that unfair or unnecessary harm is done to individuals and institutions as a result of unfair or biased information.

3. School performance indicators are not valid indicators of school quality

a) this is especially true of raw test scores (ie. without adjustments for student intake).

It is said that school performance indicators are not valid indicators of school quality. One of the reasons is that quality is a multifaceted, multidimensional, relational and contextual phenomenon which is impossible to represent fully in a few indicators. (Rowe 2000: 80).

The argument is good, but its logical implication is not non-publication. We usually handle low validity through discussion and through comparison with other criteria and other methods; we rarely revert to prohibition of invalid information.

Next, in many areas of life, we handle information that gives only a partial or incomplete picture of some phenomena. For example, I monitor my bank credit but do not confuse it with the overall quality of my life, although perhaps some people would make this confusion.

A formalistic approach to the issue of validity suggests that the validity of an indicator is not a property of the indicator itself, but of the relation between the indicator at the operational level and a definition on the conceptual level of what one claims to measure (Hellevik 2002). This leads to two complexities:

First, two different persons with two different ideas of quality in mind will ascribe more or less validity to some indicator of school quality. Politically, it is fair to say that different ideas about school quality exist because they hold different values about schools, children, learning, and society. For those who may think that testable knowledge constitutes a central outcome of schooling, indicators based on have more validity than others ascribe to them.

Second, the social and political ideas of school quality are not given once and for all, and therefore the validity of an indicator may rise or fall depending on changing concepts of quality. In fact the evaluation wave in our “audit society” (Power 1997) may have as one of its main characteristics through measurement, quantification and evaluation to not only describe what is going on in schools, but to fundamentally *define* what is measured. In that sense, indicators may help operationally define the quality they are supposed to measure. Educational politics is exercised through the setting of quantified criteria.

Let me illustrate this point with reference to what van Thiel and Leeuw (2002) call “measure fixation”. Measure fixation happens when practitioners focus on exactly what is being measured as an indicator of quality, often at the expense of genuine quality. This may be termed a “trivial” form of measure fixation because it still operates with a sense of “genuine quality” from which the measure fixation deviates. In this sense, measure fixation is “unintended.” In contradistinction, under advanced measure fixation, the indicator provides a definition of quality along with an indicator of how to measure quality. For example: “intelligence is what we measure in an intelligence test.” Or: “Nobody knows what school quality is, but in our county we think that good test scores constitute an important aspect of school quality.”

With advanced measure fixation it is not possible to demonstrate a cleavage between genuine quality and quality measured by an indicator, since the latter helps define the former. This is a constitutive effect which should be taken seriously and which is not properly understood when categorized as “pathological” or “unintended.” The trick is that while the indicator is socially constructed, the phenomenon it is supposed to measure is also socially constructed. A critique of the measure, which holds the phenomenon constant as something we all know and agree on or rationally intend, is therefore insufficient.

What is here called “advanced measure fixation” absolutely deserves discussion on normative and political grounds, but it is too easy to dismiss it as merely a matter of non-valid measurement.

Let us apply the same line of reasoning to the discussion of raw scores versus scores controlled for socioeconomic differences and pupil intake. According to the proposed code of ethics, the recommendation is clear: “school rankings based solely on “raw” examination or test score results should not be published, says Rowe (2000: 85), whereas Visscher et al (2000: 262) recommend that they not be published alone, but only along with scores adjusted for student intake and relevant school features. This is debatable. Consider the following arguments:

1. Suppose that some parents think that a raw examination result is an important aspect of school quality, irregardless of whether this indicator measures school performance or the school’s catchment area. Statistical specialists argue that raw examination results measures the latter rather than the former. But some parents simply like the idea that their children go to school together with children who receive high grades. What is debatable is whether information should be made publicly available which accommodates this wish. It is also debatable, of course, whether the data are represented as indicators of some particular definition of school quality. However, once we accept that some parents may politically desire to know the raw examination results for what they are, then our problem is one of politics, not one of validity of indicators. Do we really stand up for the argument that parents should NOT know these data, because they draw such unfortunate consequences of their knowledge?

2. It is not evident across contexts what may happen as a result of the publication of raw examination results. Andersen and Dahler-Larsen (2008) have made a case study of the newspaper's handling of such data in Denmark, and although some ranking took place, journalists also gave space to school principals from low-performing schools, allowing them to explain why they think they did a good job in spite of difficult socioeconomic and ethnic tensions in their district. It would probably have been more difficult for low-performing schools to present their views, if a statistical expert had claimed that the results were already controlled for socioeconomic background factors. (See another section below for other potential consequences of raw data).

3. The ethical and political ramifications of publishing indicators controlled for socioeconomic factors are not evident. In Denmark, attempts have been made to do so, but parents have argued that they do not accept having data about socioeconomic factors published school by school. These factors include not only economic data, but also ethnic data, and social indicators such as the ratio of single mothers. Parents have found it ethically inappropriate to have these data about themselves published school by school, irregardless of whether this took place at a district level.

It is possible that these parents never understood the value of advanced statistical analysis. It is also possible that they simply have other preferences, and these other preferences should be taken into account, too, once we acknowledge that data production about schools is a democratic activity, not only a methodological-technical one.

There is a dilemma here: Visscher recommends both raw data and adjusted data. "All adjustments should be described carefully and displayed prominently" (Rowe 2000: 85). I assume, it would be necessary to publish socio-economic control factors, too (which is not without problems) as well as the raw data (which some may read in isolation, although they are not supposed to do so).

In conclusion, withholding information because of the validity argument often wrongly assumes that there is consensus about what the indicator is supposed to measure. Next, seeking to repair validity problems through statistical control leads to a number of other, very stimulating and challenging problems.

4. In addition, there will be a number of negative unintended consequences of publication

Such as

- a) over-emphasis on content that is tested and
- b) increased social segregation of schools (Rowe 2000:76)

This is perhaps one of Rowe's best arguments. A number of studies suggest that teachers and other professionals react to performance indicator systems in a number of highly problematic ways (Dahler-Larsen 2007; van Thiel and Leeuw 2002), such as teaching to the

test, narrowing down the material, losing motivation and morale, etc. When evaluation criteria are seen as meaningless, irrelevant, superficial and detrimental to the meaning of good teaching, a deterioration of teaching practices may follow (McNeil 2000).

Another set of negative consequences concerns the distribution of pupils among schools. If certain indicators are strictly read as an indicator of a school's quality, then schools will adopt a number of unfortunate strategies to attract the best students and send problematic students away. This is neither to the benefit of society nor to many pupils, but each school must protect its reputation as it is managed by means of indicator data.

The argument about negative, unintended consequences should be taken seriously. To an increasing extent, research on the use of evaluation seems to support this argument. However, the argument should be forwarded in a careful and considerate manner. First, it has to do with negative consequences for teaching, for pupils and for society, not only for schools and teachers in themselves. Second, care should be taken with regard to the two terms "negative" and "unintended."

There is limited consensus about what constitutes a "negative" change in teaching practices. Some politicians honestly want a stricter focus on testable knowledge, whereas others prefer the individual to become well-reflected, critical, and open for debate. These are real value-differences. We should not repeat the often-made mistake to conceal political value-conflicts under the technical veil of measurement problems and "unfortunate" effects of measurements. A similar caution pertains to the term unintended consequences, which, in fact, may be built on conceptual quicksand. Does unintended mean simply "not intended" or "counterintentional"? If unintended is the logical opposite of intended, which is exactly the set of intentions which an observer applies as an analytical standard?

Are some "original" intentions empirically mapped or simply assumed? If empirically mapped, how are unofficial intentions registered? Do we allow the architect of a performance indicator system to learn over time and develop new preferences as his/her experiences with the indicator system evolve? If we lift the restriction on who holds intentions, whose intentions count?

Perhaps the most fundamental problem with the idea of unintended effects, however, is that it is practically impossible to empirically determine a certain effect as "un-intended." This would require a knowledge about intentions which is out of reach. Again, suppose parents use data with poor validity to choose schools for their children. Suppose the indicators make it easier for parents to make this choice. Perhaps it is consistent with the preferences of socioeconomically privileged people to be able to identify schools where other socioeconomically privileged people send their children, too. Can an analyst guarantee that this effect was intended? Or unintended? Why should an analytical observer remove the tension, the controversy and the reality from this situation just to be placing it totally on either side of this distinction?

Next, even if we admit that there may under some circumstances be unintended, negative consequences of publication of school indicators, there may also be a large number of other consequences, such as:

- a public debate about what constitutes school quality;
- increased dialogue between teachers, principals, and parents;
- requests for more resources to underprivileged schools;
- alternative forms of evaluation are promoted, for example school self-evaluation and/or research to explain school differences;
- a public debate about how the public and the media handle school performance indicators (this happened in Denmark);
- some people become overwhelmed with data and do not show attention to such data anymore, partly because many find out that the effects are limited;
- sometimes political decisions are made to abolish, change or adjust evaluation systems.

To the extent that these reactions, some of which may be quire “nonlinear”, are triggered by the publication of SPIs, they should be taken into account when evaluating the effect of SPIs in toto. If some of these reactions are democratically sound, they should not be ignored by evaluation specialists or education researchers who claim that they *know* what the effects of publication of SPIs are. Perhaps it is more fair to say that in a complex society, some effects are relatively predictable, others are not, but the total picture is quite complex, and this is perhaps a more honest report than claiming that the unintended, negative effects complete the picture.

5. There is uncertainty and instability in league tables

It may be true that a particular league table exaggerates the real differences between schools, because many of these differences are really within insignificant statistical scopes. It may also be true that since cohorts provide the basis for each year's school data, any given school may jump up and down the league table from one year to another. These problems do lead to cautions about making conclusions from any given league table. But do they lead to a recommendation that league tables should not be published? Not necessarily. If parents move their children from one school to another based on league tables, some would have to move their children every year. At a certain point in time, parents may learn that this practice is unproductive, and they may make some conclusions about how much they will be dictated by league tables in the future. In Denmark, a league table was published by a private think tank which displayed the scores of schools over two years, and the fact that the placement of each school was very unstable actually made it quite obvious for many that the league table was not a stable and trustworthy indicator of inherent properties of each school.

6. The mere fact of publishing information by an official body lends it credence (Rowe 2000: 86)

Maybe so, at least in some political contexts. However, this fact can be used by the authorities themselves to carefully consider the interpretive “frame” in which they decide to publish data. For example, the Ministry of Education has made average grades publicly available, but no official league tables are made in the name of the ministry. Although journalists are able to produce their own league tables, each league table will look differently and be produced in the name of the newspaper, not the ministry. There are many different ways to present the meaning and purpose of data, and indicators can be presented along with a number of other forms of information, some of which may reflect more carefully the identity of each school (Andersen and Dahler-Larsen 2008). Rather than focussing on publication or not, authorities have a number of other options which lead to more or less thoughtful ways of communication about data.

Finally, let me discuss a question not asked explicitly in the texts discussing the code of ethics for publishing school performance data: To the extent that it is recommended not to publish SPIs, does that mean that public authorities do not publish them, or does it mean that it is generally prohibited in society to publish them? The question becomes increasingly relevant when consultants, journalists, think tanks, and info-providing webpages publish their own analysis of schools. This is possible in many countries based on official statistics and/or survey studies. If one ones to legislate against these types of information in general, it may become very difficult to draw the line between what is accepted and what is not. If it is possible to publish school data, but not to compare or rank schools, everybody can just make their own comparisons. If it is made illegal in general to publish systematic school indicators, one may ask: How unsystematic must data be in order for them to be publishable? Surely, one cannot prohibit the circulation of informal rumours about school quality. It would be somewhat paradoxical, then, if rumours are OK, but systematic data are not.

In other words, in a free and open society – one in which the information available in the public arena is messy, multifaceted, and of mixed quality – it is difficult to make sure that SPIs do not appear at all. The discretionary space left for schools and school authorities may be how to talk about SPIs, how to interpret them, how to draw distinctions about their relevance, how to perhaps use them, and whether or not SPIs should be organized in particular ways by schools and authorities themselves, but not whether SPIs should exist in society. Visscher et al (2000) mention practically the same point.

Conclusion

It may in some societal contexts be true that media are bashing schools which are presented as low-performing in league tables (as exemplified in Rowe 2000: 79). It is understandable that a reaction is sought in the direction of a code of ethics for publication of such data.

However, the patterns of behaviour of media and other factors in the public arena which exist in one socio-political context should not provide the unchallenged assumptions about how these factors are functioning in all contexts.

At second thoughts, a decision not to publish data about public matters in an open society relatively easily lends itself to a number of quite obvious criticisms, such as: how can we trust that those who make the decision do not serve special interests? Would not the decision not to publish also prohibit some acceptable or even good uses of the data? And how can we be sure that the public is and remains incompetent in its handling of such data?

In a very revealing passage, Rowe (2000: 87) says that "it is not legitimate to argue that ... league tables of international educational performance reflect the quality of national educational systems." I would argue that it, in fact, *is* legitimate. It is also wrong, at least in some respects. But wrong does not equal illegitimate. To confuse the two may be a classical mistake made with the best intentions in mind, but also one which leads to a problematic editing of which statements should be allowed in the public arena.

After carefully considering all aspects of the matter, it might be, at least for already-existing data, for which there are no other legal impediments, that there is no alternative to making them available for public discussion, and then doing the hard work of participating in public life to make the proper arguments *pro et contra* about the relevance, interpretation and potential use of such data.

Reference List

- Andersen, Vibeke Norman and Peter Dahler-Larsen (2008): "The Framing of Public Evaluation Data: Transparency and Openness in Danish Schools", in *Open to the Public*, Richard Boyle, Jonathan D. Breul and Peter Dahler-Larsen (eds.). New Brunswick: Transaction Publishers, 99-116.
- Boyle, Richard, Jonathan Breul and Peter Dahler-Larsen (eds.) (2008): *Open to the Public: Evaluation in the Public Arena*. New Brunswick: Transaction.
- Castoriadis, C. (1997): *World In Fragments. Writings on Politics, society, Psychoanalysis, and the Imagination*. Stanford: Stanford University Press.
- Dahler-Larsen, Peter (2007): "Constitutive Effects of Performance Indicator Systems", in *Dilemmas of Engagement: Evaluation and the New Public Management*, Saville Kushner and Nigel Norris (eds.). Oxford: Elsevier.
- Giddens, Anthony (1990): *The Consequences of Modernity*. Cambridge: Polity Press.
- Goldstein, Harvey (1997): "Value Added Tables: The less-than-holy grail", *Managing Schools Today*, 6, 18-19.
- Goldstein, Harvey and Kate Myers (1996): "Freedom of Information: Towards a Code of Ethics for Performance Indicators", *Research Intelligence*, 57, 12-6.

- Hellevik, Ottar (2002): *Forskningsmetode i Sosiologi og Statsvitenskap*. Oslo: Universitetsforlaget.
- Mark, Melvin M., Gary T. Henry and George Julnes (2000): *Evaluation: An Integrated Framework for Understanding, Guiding, and Improving Public and Nonprofit Policies and Programs*. San Fransisco: Jossey-Bass.
- Mark, Melvin M. and Gary T. Henry (2004): "The Mechanisms and Outcomes of Evaluation Influence", *Evaluation*, 10 (1) 35-57.
- McNeil, Linda (2000): *Contradictions of School Reform: Educational Costs of Standardized Testing*. New York: Routledge Falmer.
- Popper, Karl R. (1969): *The Open Society and its Enemies*. London: Routledge.
- Power, Michael (1997): *The Audit Society*. Oxford: Oxford University Press.
- Rothstein, Henry, Michael Huber and George Gaskell (2006): "A Theory of Risk Colonization: The Spiralling Regulatory Logics of Societal and Institutional Risk", *Economy and Society*, 35 (1) February, 91-112.
- Rowe, K. J. (2000): "Assesment, League Tabels and School Effectiveness: Consider the Issues and "Lets Get Real"", *Journal of Educational Inquiry*, 1 (1), 73-98.
- Shadish, William R., Thomas D. Cook and Laura C. Leviton (1991): *Foundations of Program Evaluation: Theories of practice*. Newbury Park: SAGE Publications.
- Stehr, Nico (1994): *Knowledge Societies*. London: Sage Publications.
- Stehr, Nico (2001): *The Fragility of Modern Societies - Knowledge and Risk in the Information Age*. London: Sage Publications.
- van Thiel, Sandra and Frans L. Leeuw (2002): "The Performance Paradox in the Public Sector", *Public Performance and Management Review*, 25 (3), 267-81.
- Visscher, A., S. Karsten, T. de Jong and R. Bosker (2000): "Evidence on the Intended and Unintended Effects of Publishing School Performance Indicators", *Evaluation and Research in Education*, 14 (3/4), 254-67.
- Woolgar, Steve (2004): "Marketing Ideas", *Economy and Society*, 33 (4), 448-62.