

# Knowledge for Change

Thomas S. Popkewitz  
University of Wisconsin-Madison

## The promise of empirical evidence and benchmarks: The Loreli's whispers

Knowledge for Change lecture series #2

Malmö University October 17, 2017

Editors: Tapio Salonen & Hans Lindquist



# The Promise of Empirical Evidence and Benchmarks: The Lorelei's Whispers

There is a “commonsense” in the contemporary use of benchmarks and finding empirical evidence as a way of reasoning about change and quality. That common sense is that the correct mixture of research and policy will provide the pathways for effective change. This notion of change has produced prominent sets of connections between educational sectors, comparative research about the metrics of educational performance and policy in many countries, such as Sweden. The assessments are tied to a variety of models designed to change social welfare agencies, universities and national educational systems.

The ideas of benchmarks and having “empirical evidence”, when thought of historically and culturally, embody salvation themes of modernity embedded in the planning of social change. The international ranking lists of universities and school systems, for example, are coupled with models of change that in the language of assessment reports are to enable nations to have the world’s best-performing school systems. The promise of the reports is to provide national pathways to social equality, economic prosperity, and a participatory democracy. The promise of a better future is tied to standards expressed in the benchmarks. “Benchmarks” are the technologies to optimize the qualities and characteristics for the nation to function efficiently and achieve prosperity. The elixir to actualize the promise is numbers that provide the true, efficient and effective empirical evidence to chart national change. The salvation themes of the future, however, are a particular kind of utopic thought: they embody cultural principles about people and society as a collective space of belonging that the benchmarks and empirical evidence enable for the future.

This way of thinking and organizing national reforms is embedded in two prominent policy oriented efforts to assess and organize change in educational systems: The Programme for International Student Assessment (PISA), an international survey which assesses worldwide student skills and knowledge in science, mathematics, and literacy (<http://www.oecd.org/pisa/aboutpisa/>) and the McKinsey & Company educational reports, which draw on PISA results to “help educational systems and providers to improve outcomes for millions of students globally.” (<https://www.mckinsey.com/industries/social-sector/our-insights>)

The idea of benchmarks and “empirical evidence” are enticing, like the Sirens’ songs to the mariners along the Rhine River.<sup>1</sup> But like the Sirens, the enticements are dangerous and require caution when applied to institutions like schools and universities. This essay uses the OCED PISA and McKinsey reports on educational assessment and change to think about science, empirical evidence, and benchmarks as a way of telling the truth about people, social life, and institutions, such as universities and schools.

I explore how this science of change notion embodies a particular historical vision that is not merely descriptive but generates principles about to be actualized through *making kinds of people*; that is, as determinate categories about the qualities and characteristics of populations (see, e.g., Hacking, 1986; Popkewitz, 2008). Benchmarks and the notion of empirical evidence do not “merely” operate to describe the world for people to act on. They are an “actor” in social affairs. Benchmarks and “empirical evidence” are assembled and connected in a particular historical mode of visualizing problems, its notions of methods, and what counts as solutions to social issues. To speak of this a little differently, benchmarks and empirical evidence are like a cake. They are given intelligibility with a set of ingredients that when brought together creates the objects of seeing and acting on as important for change.

If I can play with words, I explore benchmarks and “empirical evidence” as words performed within a particular system of reason associated with international assessment. The system of reason is to think about change as practices that are to actualize kinds of people and “the social” for the future.

### **A Style of Reason: How the Recipe of Benchmarks and “Empirical Evidence” Becomes Possible**

I would like to discuss two historical dynamics in the making of the benchmarks and the ideas of “empirical evidence” before moving to the international assessments. One relates to the formation of social science in the long 19<sup>th</sup> century; that is, overlapping historical trajectories that come together and are institutionalized as the social and psychological sciences by the turn of the 20<sup>th</sup> c. The second are changes that occur in the social sciences after World War 2 through cybernetics. This is not meant as an evolutionary history but a history of the continual assembly and connections that entail continuities as well as discontinuities.

*Forming The Social Sciences, Making Kinds of People and Differences: Finding the commonsense of benchmarks and what counts as “empirical evidence” historically*

---

<sup>1</sup> I realize that historically such an analogy can be associated to gender; my intent is to point to the technological sublime, the seductions of modernity that are inscribed in discourses that seem like those of science and technology.

is in the emergence of what was called initially “the moral sciences”. This may sound odd as benchmarks and empirical evidence are thought of as neutral practices, descriptive practices outside of ideologies and social and moral value; they are thought of as only a descriptive knowledge about what works. Yet these phrases are not outside of human history but part of it. If we look to the beginning of the 1800s, the sciences about human conditions and people were called moral sciences. The concerns were with issues of deviancy and how to correct moral disorder by making kinds of people. This making of people embodied double gestures of the Enlightenment. The gesture of hope was that through the applications of reason and rationality, pathways to progress would bring liberty, prosperity, and happiness to humanity, if I can use these phrases. But moving with the gestures of hope were fears of the dangers and the dangerous populations. The populations embodied threats to the desired futures; talked about in the 19<sup>th</sup> century as barbarians, savages, backward and today spoken about with other notions to differentiate and distinguish cultural and moral differences, such as immigrants and ethnic groups as different from some unspoken normalcy and “the at-risk” child and “fragile” families.

Let me provide two examples of science and the making of kinds of people. One is the turn of the 20<sup>th</sup> century psychologies of child studies. One of the central figures of this movement was the American G. Stanley Hall. Hall argued that the science of psychology should replace moral philosophy as a way of interpreting Christian ethics and the arbiter of the moral good in social affairs, particularly in educational processes. Hall wrote that *psychology* should replace “out modeled philosophy that looks to the afterlife,” by making “new contact with life at as many points as possible.” In *Adolescence: Its Psychology And Its Relation To Physiology, Anthropology, Sociology, Sex, Crime, Religion, And Education* (1904/1928), Hall expressed this relation of science, moral order, and fears of deviancy. The idea of adolescence was not a new idea but it was applied in a new way to think about the transition between childhood and adulthood through scientific evidence. From the title of Hall’s book, the juxtaposition of science and moral issues and their link to education is evident.

The hope of adolescence, as expressed in this book, was the hope of psychology producing the future cosmopolitan child through a “more laborious method of observation, description, and induction”. But the gesture of hope of cosmopolitanism was engendered with fears of the poor, immigrants and racial groups of the new industrial cities, what Hall called the “*urban hothouse*”. The city was seen as a space of “perversion, ... and hoodlumism, juvenile crime, and secret vice ... increasing (what challenges) civilized lands.” Hall also worried about gender. His studies were of white males and “dangers ... of establishing normal periodicity in girls, to the needs of which everything else should for a few years be secondary.” Psychology, he said, should help develop men who were naturally “aggressive and prepare women for maternity.” Finally and also related to the city was the *unbridled capitalism of this period in American history where there was “the mad rush for sudden wealth and the reckless passions set by its gilded youth.”*

We no longer talk about the moral sciences and instead use a different language in which benchmarks and “scientific evidence” become a way of articulating moral questions of the present and the future. But to think about how science as making kinds of people is (re)visioned, reassembled and given the language that we now speak of as benchmarks and “scientific evidence”, the post-war years need to be brought into focus. This becomes the second part of the ingredients of the recipe that is assembled in the making of people.

A second example in the making of kinds of people is cybernetics. Benchmarks and scientific evidence are given expressions through cybernetics to think about human affairs as the relation of mind and machine. Initially tied to war efforts, cybernetics circulates as ways of thinking about cognitive psychology, “bounded rationalities”, political systems, sociological phenomena, anthropology. It entails a particular set of rules and standards to rationalize social life and issues of change. Cybernetics was developed during the war effort and brought into social analysis. It created a way to think about mind in relation to the machine – the machine as the computer and its analogy to the mind as artificial intelligence. The focus was processes and networks of communication that provided the method and strategy for change.

If I can summarize a recently emerging history of science, cybernetics provides concepts for mapping the processes and flows of information as stable objects for administration; the mode of reasoning whose principles give form to the current thinking of benchmarks and scientific evidence.

Cybernetics theories connect to systems thought. Systems as an abstraction to actualize future society and people; the abstraction embodies principles that are not empirically deduced but are *a priori* and *self-referential and self-authorizing*; that is, its mode of ordering and classifying inscribes internal boundaries in defining problems, contexts and the possibilities of change. This is not unique to cybernetics as a system of reason. What is given focus here, however, are the principles of systems thought as a strategy of change.

The idea of a system as an organism replaced earlier mechanical notions with more dynamic models of change. But the idea of a machine did not disappear. With language borrowed from biology, social institutions are conceptualized as a *social organism* having stages of growth and processes of development that change over time. Reasoning about social relations and change through systems is an abstraction for ordering what is seen, thought and acted on for producing change.

Cybernetics and systems thought move from the goal to obtain ideal types to thinking about standards concerned with optimizing utility of the system without striving for perfection. One of the debates in computer science during this time was whether the purpose of research was to create programs that eliminated all error, thus producing the modern philosopher’s stone. The other position was to try to

produce programs that would eliminate errors as best as possible, knowing that the perfect system was not possible. This later approach won!

The twin possibilities and the outcome in computer science when seen in the history of ideas was part of the larger epistemological debates in social science. To bring to the present, the international ranking systems of PISA and other social and economic indicators are not about finding the perfect system. The rankings draw on cybernetic modes of thinking to compare, order, and plan for efficiency in processes and communications patterns that optimize systems.

Another element in this new rationality was what constituted the rules and standards of empirical evidence. Historically, the idea of scientific and empirical evidence means simply systematically observing what happens in everyday life. A newspaper, a play, a sport game, as well as introspection in early psychology were ways of ordering and classifying empirical evidence. In post-war years social science was concerned with the administration of change incorporating the idea of algorithms to think through mathematics about empirical evidence. Algorithms, it needs to be noted, entail a particular kind of mathematical thinking about social life as having rigid rules that provide optimal solutions to given problems, or delineate the most efficient means toward certain given goals. The models of change offered by the OECD report on the Swedish school system (Pont, Donaldson, Elmore, & Kools, 2014), discussed later, inscribes the operation of algorithms as underlying principles for forming the model of change that is to lift Sweden from average to above average.

When cybernetics, systems theories, and “empirical evidence” are ordered as algorithmic rules, the numbers and benchmarks of international ranking become particular cultural practices about the making of society and people.

### **Making society/making people: The cultural practice of numbers**

By now, it should be clear that the benchmarks of international assessments of schools and international ranking of universities are not merely descriptions born of empirical data drawn from the present but historically embodied in trajectories of the social sciences that are about people to actualize a desired future. The OECD’s PISA and the McKinsey reports on education are ordered through cybernetics and systems analysis as a theory ordering assessments by focusing on processes and communication patterns of social life that, while, at the same time, it is about ordering the possibilities of change that anticipate what is the desired future of an imagined society and people. The school is studied as a system that has qualities of a biological organism, a metaphor to think about “*the educational needs*” in which social growth and development can be measured.

Numbers serve as the reference within the systems analysis, and benchmarks as the empirical evidence. Numbers connect as a further ingredient of this recipe

knowledge of assessment and change. The magnitudes of differences in the statistical correlations are placed into models of intervention that are to bring into existence kinds of people that can actualize the effectiveness of school viewed through an abstraction of systems to think about and administer social relations.

If I move to the present and again being synoptic, international assessments of the OECD are “merely” descriptive of some reality but “act” in making or fabricating what matters; what “acts” as a given to social problems and the strategies of change are to enact that “nature”. The statistics and numbers generated in the international assessments are taken as stable scientific facts for planning and interventions. Measures provide a comparative algorithm that “tells” of a continuum of values about people and the future that enables successful school systems.

The measures are to lead to a common world accessible as highways to rectify the dangers that are disruptive of the equilibrium of the system. That is what the models of change in the OECD Education Policy Review report of assessment and change are to produce. The models of change are not merely about systems. In the Swedish report, the universal characteristics and qualities of kinds of people are those that are actualized nationally, as the vision and rationality for thinking and acting as teachers, but also the social and psychological qualities of “well-being” of the abstractions that unity students, parents and communities! (See, e.g., Pont, Donaldson, Elmore, & Kools, 2014; OECD, 2017).

### **Benchmarks & variations: Desired people to be actualized**

The counting and numbers when comparing nations and educational systems perform as expectations about universal characteristics of society and people whose composition forms a common and harmonious world. The numbers embody an anticipatory reasoning about the future society and populations. McKinsey’s *How the world’s most improved school systems keep getting better* argue, for example, that benchmarks are an “*universal scale of calibration*” to create equivalences from, for example, several different international assessment scales of student outcomes discussed in education literature” (Mourshed, Chikioke, Barber, 2010:7). Benchmarks are standards placed in scales that order elements on a continuum from “*poor/fair to good*”, “*good to great*” and from “*great to excellent*”. In a different report on how school systems are improving, the scale is given as a clear and linear progression that is internal to each category and then correlated across categories (Barton, Farrel, & Mourshed, 2014), such as:

*Fair to good*: consolidating system foundations, high quality performance data, teacher and school accountability, appropriate financing, organization structure, pedagogical models;

*Good to great*: teaching and school leadership as a full-fledged profession, necessary practice and career paths as in medicine and law; and

*Great to excellent*: more locus of improvement from center to school, peer-based learning, support of system-sponsored innovation and experimentation.

The strategy is to address deviations from the norms in the development of country case studies. Variations are from the standardized norms that define differences and spaces of actions.

The benchmarks seem to be about *national development*. But the qualities and characteristics given attention through the benchmarks and the scaling are abstractions of kinds of people and differences. National student performances are linked to psychological qualities of the teacher and the child. Measures of achievement are correlated to who the teacher is, psychologies of the child, school organization, and norms about modes of living called “parent participation”; for example, “peer-led creativity and innovation” and “building technical skills of teachers and principals”. Measurement categories that focus on “creativity”, “innovation” and “participation skills”, embody principles about desired kinds of people and the kind of society that gives expression to the desires. The qualities and characteristics are normative, constituting values as well-being measures about the “enjoyment of life”, happiness, belonging, and self-realization.

The logic of change embedded in the scaling creates a continuum of value. The differences are standardized, codified and ordered into hierarchies of values for comparing. The hierarchy of values is created to differentiate nations and populations. The statistical analyses used to talk about school systems are said to “*examine why and what they have done have succeeded where so many others failed*” (see, e.g., Mourshed, Chijioke, & Barber, 2010).

The standardizing and codifying to find equivalences, ironically, erase difference by establishing difference. The reduction of complexities to those of rational management “systems” makes it seem that “all” national systems can anticipate equality through the application of categories that recognize difference that inscribes difference. Differences entail comparisons through creating sets of equivalences among disparate databases. The paradox of the international comparisons is its inscription of difference that “makes” differences so that some can never be at the “*top*”.

### **Double gestures: The hope and fears of kinds of people**

Benchmarks and their “empirical evidence” embody the universals that paradoxically compare and divide. Lists and rankings in the international assessments compare secondary statistical measures that create “*a universal calibration*” in which a spectrum of norms defines equivalencies among subsets of



data (Barton, Farrell & Mourshed (2013) *Education to employment: Designing a system that works*:7).

The comparison eliminates differences to produce distinctions that divide. If I draw on the OECD and McKinsey reports, effective education travels as the gesture of hope that forecasts the salvation themes of a good society, full employment, well-being, and the progress of the nation. The classifications and numbers connect to psychological categories of children's social and communicative patterns, such as family influence on children's achievement and the relation of education to employment. The social and psychological distinctions are about the hopes of future kinds of people. The hopes, however, simultaneously express the gesture of fear of the dangers and dangerous populations to that future. The fears are expressed as the kind of parent who does not enable the child's moral development for success in school and the kind of child who "lacks" motivation, well-being, and the proper modes of living. The delineating of stages of development are not only organizational factors but they also align with psychological qualities of youth that normalize what is functional and dysfunctional for employability, such as disengaged, disheartened, well-positioned or too poor to study (Barton Farrell, & Mourshed, 2013:32-33).

The gestures of hope and fear are double gestures. The statistical calibrations are about who people are and should be, as well as about who does not "fit" as part of the universal. The characteristics of people who succeed and don't succeed form a continuum of value about the hope to actualize a desired future with fears of populations inscribed as dangerous to the system's harmony and consensus. Codifying and standardizing are not merely about achievement. The ranking and classification engender differences in those "civilized" and those different in degree from that advanced stage of civilization – the school systems and nations at the top!

### **"Follow me!" Knowing the future as taming uncertainty**

The future is certain and the problem of measurement is to put nations and people on the highways to actualize the abstraction of the school system. McKinsey uses the highway metaphor, for example, to think about highways as not merely paths to the future. They embody the qualities and characteristics of the kinds of people who will inhabit that future. Not far away from the highways and pathways that are to "deliver better outcomes" for future harmony and consensus are fears of danger and dangerous people. To follow the models of change in reducing unemployment among ethnic, racial and poor populations is as "*to get rid of potholes, make educators and employers part of the solution by providing 'signs' and 'concentrate on the patch of pavement ahead'*" (Barton, Farrell, & Mourshed, 2013:54).

Benchmarks and “empirical evidence” are inscription devices that portray that the knowledge of the future is at hand for all nations to reach the top. The pathways posit social life as a mechanism or machine whose proper alignment (equilibrium) allows for it to administer system goals. The problem is how to tailor the highways individually so all can find the destination.

### **Some Concluding Thoughts**

I began with the Lorelei as an analogy of the Siren’s enticing the mariners’ ships into the rock. In some ways, benchmarks and “scientific evidence” provide the contemporary temptations to the issues of development and progress. The beckonings of today are expressed as benchmarks and “scientific evidence”. They embody salvation themes explored as having particular limits in thinking about change, and the making of people and society. An anticipatory future is a calculated rationality that shapes and fashions as ahistorical yet is located in a particular historical configuration. The international assessments are anticipatory, in the same manner as a Google, Amazon, or Netflix search anticipates who you want to be. The difference with the international assessments from the web searches is that our preferences have not been registered prior to the algorithm’s work on us. The preferences are prefigured in the abstraction of the school as a system. The irony and paradox of the system’s principles is that its harmony and consensus morph into cultural practices of normalcy and pathology. The comparing with the universal norms and distinctions provided differences and divisions. The divisions were pathologies of populations dangerous to the system’s models and highways and feared if not changed.

Benchmarks and “empirical evidence” embody the salvation theme of finding the future. They embody inscriptions that order and classify the present as a future that the research will actualize. That future entails a comparativeness that differentiates normalcy and pathology as gestures of hope and fear. At this point, if benchmarks and empirical evidence are to bring in the future – what future in education and other social institution is to be actualized?

**\* Note:** This essay brings together different research projects related to a history of present social science/educational reform-oriented research listed below. This includes a VR research project with Sverker Lindblad of the University of Gothenburg and Daniel Pettersson of the University of Gävle related to the sociology of science (International Comparisons and Re-modelling of Welfare State Education), and a book I am writing on while at Malmö University this fall, tentatively entitled “The Impracticality of Practice Research: Strategies of Change that Conserve”.

## References

- Hacking, I. (1986), "Making up people", in T. C. Heller, M. Sosna & D. E. Wellbery (eds.), *Reconstructing individualism: Autonomy, individuality, and the self in Western thought*, Stanford, CA: Stanford University Press, 222-236, 347-348.
- Lindblad, S, D. Pettersson and T.S.Popkewitz, (2015), *International Comparisons of School Results: A Systematic Review of Research on Large Scale Assessments in Education*. Delrapport från SKOLFORSK-projektet, Vetenskapsrådet 2015. Stockholm.  
<https://publikationer.vr.se/produkt/international-comparisons-of-school-results-a-systematic-review-of-research-on-large-scale-assessments-in-education/>
- Mourshed, M., C. Chijioke and M. Barber (2010), *How the world's most improved school systems keep getting better*. Chicago, IL: McKinsey & Company.
- Mourshed, M., D. Farrell and D. Barton (2013), *Education to employment: Designing a system that works*. Chicago: McKinsey & Company.
- OECD (2017). *PISA 2015 Results (Volume III): Students' Well-Being*, OECD Publishing, Paris. DOI: <http://dx.doi.org/10.1787/9789264273856-en>
- Pont, B., G. Donaldson, R. Elmore and M. Kools, M. (2014), *The OECD-Sweden education policy review. Main issues and next steps*. Paris: OECD.
- Popkewitz, T. (2008), *Cosmopolitanism and the age of school reform: Science, education, and making society by making the child*, New York: Routledge.
- Popkewitz, T. (in press). "Anticipating the Future Society: The Cultural Inscription of Numbers and International Large Scale Assessment." In S. Lindblad, S, D. Pettersson, and T. Popkewitz (eds.), *Numbers, Education, and the Making of Society: International Assessments and its Expertise*. New York: Routledge.

