Chapter 16
International Assessments of Student Performance: The Paradoxes of Benchmarks and Empirical Evidence for National Policy

Thomas S. Popkewitz

There is a "commonsense" in the contemporary policy that moves across Europe and North America. That commonsense is the use of benchmarks in welfare state reform to assure the proper articulation of goals that enable their measurement and attainment. The corollary of the benchmark statements is that research identifies the empirical evidence that testifies about what works to secure the desired changes. The putting together of benchmarks and the call for "scientific evident" entails the faith that the correct mixture of research and policy will provide the pathways for effective social and educational improvement.[1]

This chapter approaches the ideas of benchmarks and having "empirical evidence" as not merely policy instruments for educational improvement. They embody particular ways of reasoning about social planning and social sciences that have implications for thinking about the organization of society and the kinds of people desired for "progress" (see Popkewitz in press). These implications of the reasoning of benchmarks and empirical evidence are explored in The Organization for Economic Cooperation and Development's (OECD) Programme for International Student Assessment (PISA) surveys student skills and knowledge in science,

---

[1] The discussion brings together different research projects on the sociology of scientific knowledge in Lindblad et al. (2018) and Popkewitz (2018, in press).

This essay brings together different research projects related to a history of present social science/educational reform-oriented research. This includes a VR research project with Sverker Lindblad and Gun-Britt Wärvik, Gothenburg University, and Daniel Pettersson: (Lindblad, Pettersson & Popkewitz, 2015, Popkewitz, 2018). The University of Gävle and Uppsala, related to the sociology of science (International Comparisons and Re-modelling of Welfare State Education), and "THE IMPRACTICALITY OF PRACTICAL RESEARCH: A HISTORY OF SCIENCES OF CHANGE THAT CONSERVE (University of Michigan Press).

T. S. Popkewitz (✉)
Department of Curriculum and Instruction, The University of Wisconsin-Madison, Madison, WI, USA
e-mail: thomas.popkewitz@wisc.edu

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 chapter examines the PISA and McKinsey report models of educational change as expressing the salvation theme of modernity, expressing a particular kind of utopic thought about human betterment that combines political, social, and economic ideals. Explored are the principles in these assessment's statements of benchmarks and "empirical evidence"; principles about what matters, how problems are articulated, what notions of methods are reasonable, and what counts as solutions to problems identified. The first section explores historically two elements that underlie the assessment: a universalized conception of society and individuals that connects with systems and cybernetics theories to direct change. The second section focuses on how numbers enacted in PISA require categories and classifications about societies and people that the research is to actualize. The third section considers the notion of change implied, focusing on the social implications of the counting and numbers used in the international assessments. The fourth section argues that there is comparative reasoning about differences that is not only about nations. The measures generate principles about cultural differences among populations. The final section explores how social and cultural principles are erased through the system's focus on process, "highways" and "pathways" to follow for success.

The chapter is a study of these sciences as a historical phenomenon. The benchmarks are like the Sirens songs that drew the mariners into the rocky shores of the Rhine River. The salvation themes of the assessments are enticements that can be dangerous and require caution when applied in social policy to institutions like schools.

16.1 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 European and North American social sciences in the long nineteenth century; that is, overlapping historical trajectories that come together and are institutionalized as the social and psychological sciences between the late 1700s and early 1900s. The second are changes that occur in the social sciences after World War II through systems theory and cybernetics.

In what might seem as far removed from international assessments, the finding the commonsense of benchmarks and what counts as "empirical evidence" historically is in the emergence of what was called initially "the moral sciences" or moral philosophy. This may sound odd as benchmarks and empirical evidence are thought of as neutral practices in modern policy and reform-oriented research—descriptive practices that about what works.

Yet these phrases of contemporary sciences are not outside of human history but a particular part of it. If we look to the beginning of the 1800s, attention was directed to the sciences about human conditions and people were called moral sciences. At one level was the European Enlightenment commitment to reason and science in pursuing progress in "The City of Man" (sic). Attention was given by philosophers but also speculatively by social sciences[2] to the manners by which people live and work together and how to alter those people in light of some general moral qualities that were thought of as universal to all. The concerns were often directed to questions of deviancy and how to correct moral disorder that was associated with urban life and industrialization in Europe and North America. The domestic sciences that emerged later in the nineteenth century, for example, were to teach the poor and working classes hygiene, child-rearing, as well as how to organize a life determined by wage earning. These changes, however, were not only about the poor as they worked into the conduct of the middle classes.

The moral sciences designed to make kinds of people embodied double gestures. (see, e.g., Hacking 1986) There was the gesture of the enlightenment hope that through the applications of reason and rationality would identify pathways to bring liberty, prosperity, and happiness by producing particular kinds of characteristics and qualities to people. But moving with the gestures of hope were fears. The fears were of the dangers and the dangerous populations. The populations embodied threats to the desired futures, talked about in the nineteenth century as barbarians, savages, backward and today spoken about with other notions to differentiate and distinguish cultural and moral differences from some unspoken normalcy, such as the qualities of difference in Western societies among immigrants, ethnic groups, "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 twentieth 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* (Hall 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 was the hope of psychology producing the future cosmopolitan child through a "more laborious method of observation, description, and

---

[2]The people associated with the British and American Social Science Associations were not "trained" as such but came from different social fields to look at the effects of poverty in society and organize ways of mitigating its consequences.

induction." But the gesture of hope of cosmopolitanism was engendered with fears of the poor, immigrants, and racial groups of the new industrial cities, in 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 the "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 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. The changes in the language of science allow the discussion to move to the postwar years. This revisioning is the second part of the ingredients of the "recipe" of ideas and theories assembled in the making of people that connects to PISA.

With the making of people, the second part of this recipe of science is systems theory and cybernetics. Initially tied to war efforts, cybernetics joins with systems theories in multiple social and psychological sciences, such as cognitive psychology, sociology, and anthropology. Cybernetics brought into social analysis 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 on processes and networks of communication that provided the method and strategy for change.

Systems theory was not new. It appears in Adam Smith's Wealth of Nations in the 1800s, is placed with mathematics by John von Neumann in the 1920s, and is revisioned after World War II with the development of cybernetics. It is this later notion of systems that becomes important for thinking about the relation of research, policy, and change when drawing on the international assessments of student performance, notions of benchmarks, and the invoking of "empirical evidence." That is, systems analysis provides a "basic ingredient" to shape and fashion the spaces in the assessments as a salvation theme in which to order, classifying and act on what schools do.

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.

Systems thought, developed in the 1920s and assembled with cybernetics during the war, was taken in the human sciences as providing an "unprecedented synthesis" of the notions of human life. Biological metaphors of social life as an organism that grew, developed, and changed were incorporated in social theories to study and organize the objects of change. The openness of the system to change is expressed as correlations between functions (e.g., family life, child self-esteem, teacher professional development) and structures defined as the system (e.g., institutional units in school "system" such as classrooms and school leadership characteristics).

What was different was combining the biological analogy of system with cybernetics. Change entails the link between human behaviors with machines (e.g., computers, photocells, and radar) directed to systems goals. The language to describe change is processes expressed as inputs and outputs. The processes and communication (organism) function as networks, flows, and circuits within structures (the machine) as "feedback" loops to trigger systems development and growth, the operational definition of change. Information is not about meaning but choices between possibilities within a structured situation, structurally denoting a formally defined range of possibilities for communication. The purpose of social and educational research is goal attainment or what earlier was spoken about as knowledge utilization.

The object of change is the ordering of the constellation of components of the system that can achieve its optimal relations. Although not essential earlier in social thought, algorithms became important for thinking about the rigid rules that provide optimal solutions to the given problems or delineating the most efficient means toward certain given goals. Choice is between discrete units (Halpern 2014: 46–47). Cybernetics theories connected to systems thought bring into view a way to think about social life and change that entails determinacy and indeterminacy. When cybernetics and systems theories are examined as principles ordering the international assessments, the measurement procedures stabilize the components of the system as ontological objects (the professional teacher) in order to examine its processes that contribute to its optimization.

The principles of harmony and consensus in social and psychological research entail hypothesizing the state of equilibrium to express the optimum point to achieve. With equilibrium is what hinders or prevents the optimization of systems goals. This establishes a symbiotic relation between what otherwise appears as opposites: equilibrium and disequilibrium. Research is to minimize the points of disequilibrium to achieve stability and harmony.

When applied in the social and educational research about change, equilibrium and disequilibrium translate into social values that express normality and pathology. OECD's current measures of "well-being," for example, are to understand the psychological and social conditions that contribute to high student performance (i.e., the normal). The idea of well-being simultaneously brings into existence the qualities of students that limit, interfere, and restrain the functioning of the system, such as family and community experiences as well as personality traits that are lacking in the child, such as lacking motivation and engagement. The qualities of interference and restrain are the practical translations of system's theory of disequilibrium into cultural characteristics of pathology.[3]

---

[3] As Bürgi and Tröhler argue (2018), benchmarks emergence within OECD were educational "thermometers" that drew on a medical language about normalcy and pathology. The defining quality through benchmarks (re)visioned the language of social engineering. My concern is how this language is inscribed in systems theories as cultural principles of normalcy and pathology.

Harmony, consensus (equilibrium), and the disruptions (disequilibrium) of the system theoretically order the problem of change. The homogeneity and consensus make administration and prediction possible in strategies to change schools. To talk about the students' achievement gap to identity those children in need of educational remediations, for example, assumes the consensus of purpose and harmony necessary for the system components but which the gap disturbs.

But this harmony and consensus is predicated on potentialities where system performance actualizes what is desired. Benchmarks are the optimal goals to obtain (Halpern 2014: 45). 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 process and communication patterns that optimize systems. Optimization is where all girls equally learn mathematical knowledge and where there is no achievement gap, where *all* children read and where all are mathematically able, and the work of experts and professional teachers is engaged as full efficiency.

The complex epistemic framing of systems analysis was brought into multiple disciplinary projects that included education. The system's principles were connected and assembled with social and cultural notions about, for example, people as "naturally irrational" and managed through processes of decision-making (see Heyck 2015). The new mathematics curriculum of the 1960s, for example, focused on the processes and communication patterns that could be "theorized and its components identified through a particular set of behaviors and traits thought to make up that kind of person (and thereby a rational and democratic collective)" (Diaz 2017: 31). The professional organization for mathematics teachers, for example, argued that learning mathematics is "contributing to effective living, otherwise it does not have worth and usefulness" (National Council of Teachers of Mathematics 1945: 200). The "applications of mathematics to problems of industry, physical science, aviation and business should be used for purposes of motivation, illustration and transfer" (National Council of Teachers of Mathematics 1945: 201).

Systems as an abstraction 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 system theory. What is given focus here, however, are the principles of systems thought as a strategy of change in educational policy and research.

Another element in this new rationality was what constituted the rules and standards of empirical evidence. Historically, the idea of scientific, empirical evidence means simply systematically observing what happens in everyday life. A newspaper, a play, a sport game, and an introspection in early psychology were ways of ordering and classifying empirical evidence. In the postwar 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 et al. 2014), discussed later, inscribe the operation of algorithms as underlying principles for forming the model of change that is to lift Sweden from average to above average.

16.2   Numbers as Cultural Practices

While brief, the historical discussions directed attention to benchmarks of international assessments of schools and international ranking of universities are not merely descriptions born of empirical data drawn from the present. The numbers are brought into reports embody historically lines whose principles are about people that research is to actualize (see Lindblad et al. 2018). The OECD's PISA and the McKinsey reports on education are ordered through cybernetics and systems analysis as a theory ordering assessment 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 a desired 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 are parts of systems of communication whose technologies create distances from phenomena by appearing to summarize complex events and transactions (Porter 1995). As the mechanical objectivity of numbers appears to follow a priori rules that project fairness and impartiality, numbers are seen as excluding judgment and mitigating subjectivity. Numbers are a technology of distance and used as a claim of objectivity instantiated by moral and political discourses. They bring into existence kinds of people actualized within the boundaries of possibilities of the abstraction given as the school "system."

Numbers connect and are a further ingredient of this recipe of the reason organizing assessment and change. The domain of quantified knowledge is artificial through creating uniformity among different qualities of things that gives social authority to the interrelation of science and policy (Porter 1995: 6). The uniformity and quality of things in the statistical correlations of the international assessment are placed into models of intervention. The models of change identified by the OECD report on the Swedish school system (Pont et al. 2014) have qualities of algorithms. The problem solving and the "scientific evidence" expressed through numbers are to verify the benchmarks as algorithmic rules. The model appears as merely the application of statistical thinking which, as noted in the previous chapter, is a kind of mathematical thinking about social life that has rigid rules. The algorithmic rules provide optimal solutions to given problems or delineate the most efficient means toward certain given goals. The algorithms of the measurements are constructed to neutralize the indeterminate qualities of social life, culture, politics, and context (Barber and Mourshed 2007: 13).

The numbers and comparative listings of nations in PISA, for example, function as a GPS system for national school systems for people and governments to locate themselves and identify differences.

Embedded in the broad generalization are categorical constructions that are expressed to compare and rank nations are directed to the qualities of people—teachers, school leaders, children, and their family. The composites formed to classify school systems entail prior conceptions of the dispositions and sensitivities of what constitutes, for example, the classifications of school leaders and teachers who can "adapt" and implement the models of change. The taxonomies of the skills of an "expert" or professional teacher, for example, are qualities of "peer-led creativity and innovation" (Mourshed et al. 2010: 20), or "building technical skills of teachers and principals" (Mourshed et al. 2010: 28) that act comparatively. Creativity, innovation, and skills are words to differentiate particular kinds of people, their interactions, and sociality from those not creative, innovative, or skillful.

Mosaics of numbers are assembled as truth bearing statements about the effective functioning of schools that appear as a unified abstraction of "nation" and its potentialities (see, e.g., Popkewitz 2008). The complexities of the differences among nations and cultures disappear and reappear as standardized and comparable descriptions of numbers that represent singular, universal population of nations from which differences are calculated.

The visual techniques of OECD's graphs, statistics, and charts function as maps to organize the flow of information about stable objects that move among different social spaces to "tell" of the route to innovation (Halpern 2014). The graphs, statistics, and charts perform as "immutable mobiles" (Latour 1986). They are visualization technologies that collapse complexities into standardized categories and calculations in which phenomena seem well arranged, easily accessible, and can travel to different places for monitoring and steering what is seen and acted on.

The optical consistency translates statistical distinctions into information appearing as having a "communicative objectivity." The "optical consistency" entails a particular calculative rationality in which process and method are fabricated as material objects, with statistics a tactic for visual information. Numbers are given as the transcendent ordering of what nations need for development, growth, and equity. Cultural distinctions are erased to create a layer for comparison of differences through the superordinate qualities of the statistical equivalences. Numbers act like a communication practice through which statistical equivalency performs like the reasoning about comparability and differences.

The visualization technologies of numbers no longer appear as measuring personality and inner qualities, but are about nations "seen" through the standardization of those qualities and characteristics of people that need development (see, e.g., Borgonovi and Przemyslaw 2016: 132).

Change is given its directionality that signifies educational improvement. The processes of change are visualized as well known. The change models are given as orderly, linear processes that instantiate clear and logical procedures. The procedures are available to all if wise enough to follow the "highways"—a word used, for instance, by the OECD and the McKinney Reports (see, e.g., Mourshed et al. 2010).

Ignored in most policy studies and research is the paradox of inscribing equivalency and comparability through numbers. The technologies of numbers are embodied in a grid of cultural practices that "act" on teachers' and children's lives in classrooms. To talk about "achievement" and the "achievement gap," shorthand for numerical differences between children instantiates particular rules and standards of reason by which experiences are classified, problems located, and procedures given to order, classify, and divide. Exploring the "reason" through which numbers are made sensible and plausible puts focus on the processes of exclusion and abjection in the impulses to include.

If I move to the present, 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 et al. 2014; OECD 2017).

16.3   Benchmarks and Variations: Desired People to Be Actualized

The counting and numbers comparing nations and educational systems perform as expectations about universal characteristics of society and people. These universal characteristics form as images and narratives that express the common and harmonious world prescribed through its system's theory. While the graphs, charts, and magnitudes show differences that seem as only categories about the school systems of nations, the comparisons entail ranking extensive codifying and standardizing of characteristics of people and institutions that are elided in the visualizations. The 2015 PISA assessment is characteristics of children in relation to families that are about "kinds of people." The assessments are described as the student's "well-being" that contribute to successful school performance. The numbers embody "a comprehensive set of well-being indicators for adolescents that covers both negative outcomes (e.g., anxiety, low performance) and the positive impulses that promote healthy development (e.g., interest, engagement, motivation to achieve)" (http://

www.oecd.org/pisa/publications/pisa-2015-results-volume-iii-9789264273856-en.htm).

The comparison and ranking of nations are placed into models of change to actualize the desires generated as "the arrow of time." The OECD Education Policy Review for Sweden (Pont et al. 2014), for example, suggests a three-part process. Change is expressed as recommendations "tailored" to the specific education system's "needs." The tailored advice entails words like contextualization of "country's needs." The tailoring is, in fact, the generation of desires. The numbers appear as the "empirical" evidence of the future appearing innocuously in the optical consistency of the charts as "the needs" of nations.

The success and failure are visualized as scales that map about the development and changes of populations as the arrow of time. The scales appear initially as institutional trajectories that identify different characteristics of national and cities developmental patterns to achieve success. Variations are registered as a continuum of values about the normal and pathological. The lists and rankings in the international assessments produce a visual form of scaling that differentiates and divides (Hansen and Vestergaard 2018).

Scaling is produced through correlations of the data to project, for example, "integrated set of actions" within a hierarchy that forms "intervention clusters" for improving the performance levels of the system (Mourshed et al. 2010: 14). The scales combine institutional (organizational) with personal qualities in a seamless movement that give the system measures of "accountability, performance, and professionalism" (p. 14). The universalize standards are scaled and, in the case below, have no content and appear as a clear and linear progression discrete markers about "stage-dependent interventions" that produce school improvement.

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 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 et al. 2010).

The scaling entails an anticipatory reasoning about the future society and populations. McKinsey's *How the world's most improved school systems keep getting better* argues, 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 et al. 2010: 7). Benchmarks are standards placed in scales that order elements on a continuum from "poor/fair to good," "good to great," and "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 but directed to a philosophical ideal about what constitutes the desired school (Barton et al. 2013), 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. The numbers generated in the statistical measures are inscription devices that assemble and connect pedagogical, psychological, and social/cultural principles. The social/political outcomes are coupled with psychological outcomes to bring salvation themes into fruition: students' happiness, well-being, and life satisfaction.

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 indicators of national performance are cultural registers about people. "The evidence base … [of PISA] goes well beyond statistical benchmarking" to examine children's "enjoyment of life," asking

Are students basically happy? Do they feel that they belong to a community at school? Do they enjoy supportive relations with their peers, their teachers and their parents? Is there any association between the quality of students' relationships in and outside of school and their academic performance? … Together they can attend to students' psychological and social needs and help them develop a sense of control over their future and the resilience they need to be successful in life. (OECD 2017: 3)

Characteristics about people are re-visualized as macro-numerical consistencies and differences across nations. The statistical measures are based on equivalences that create universal categories from which difference is assessed and charted along continua of value. The visual ordering of numerical data creates variations of performance as they relate to measures of "endurance" and motivation as comparative qualities of collective and national differences. The skills and competences are connected to organizational qualities (e.g., teacher professional development, school leadership) and desired sociological and psychological characteristics of children.

Differences appear as comparisons created as sets of equivalences among disparate databases. The comparisons are formed through objectification about people embedded in universal calibrations. The microstudies entail classifications and numbers that connect to the 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 measures codify distinctions about the "needs" of better-performing and low-performing students, objectifications that elicit identifying processes of "feedback" loops talked through categories about autonomy, respect, parent involvement, and interactions with school and other parents, and as psychological characteristics of motivation versus disruptive behaviors (OECD 2017). The qualities as distinctions and differentiations are recalibrated into national tables in which the submeasures and statistical distinctions disappear as macro-statistical categories about society and nation.

The comparisons are formed through secondary statistical measures that form a spectrum that rests, in turn, on a universal scale of calibration that we developed by normalizing several different international assessment scales of student outcomes discussed in the education literature. Our findings are not, however, the result of an abstract, statistical exercise. In addition to assessment and other quantitative data, they are "based on interviews with more than 200 system leaders and their staff, supplemented by visits to view all 20 systems in action" (Mourshed et al. 2010: 12–13).

Yet 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."

16.4   Double Gestures: The Hope and Fears of Kinds of People

The mapping of the international assessment appears as about national development in a GPS whose ranking and lists seem about potentialities of what should be if only nations work hard and diligently through education. But the potentialities, as discussed above, are saturated with the potentialities of societies, people, and nations. There are hopes that simultaneously generate fears that are expressed as unless a nation makes "sufficient investments to develop capabilities in the present, students are unlikely to enjoy well-being as adults," writes the OECD report (2017: 62). The potentialities that nations are to achieve are double gestures. Benchmarks and their "empirical evidence" embody the universals that paradoxically compare and divide. Lists and rankings in the international assessments, for example, compare secondary statistical measures that create "a universal calibration" in which a spectrum of norms defines equivalencies among subsets of data (Barton et al. 2013: 7).

The gestures of hope and fear that are generated in 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 do not 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!

The paradox of the change to include is to normalize differences—differences as a comparative logic of nations that also has comparative notions of society and individual embodied in the macro statistics. The irony and paradox of the systems principles is that its harmony and consensus morph into cultural practices of normalcy and pathology. The preferences embody prefigured divisions that entail the pathologies of populations dangerous to the system's models and pathways that are feared if not changed.

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, described through categories of disengaged, disheartened, well-positioned or too poor to study (Barton et al. 2013: 32–33).

16.5 "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 et al. 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.

The mechanisms of change are universal. The proper alignment of these drives inaugurates the pathways to optimize systems goals. Change is the application of the universal "to navigate the challenges in their context and to use their context to their advantage" (Mourshed et al. 2010: 26). Innovation relates to how well the pathways are delineated to access the highways to success.

Finding the right highways also means recognizing that there are dangers and the dangerous people. The paradox of the pathways is the comparative reasoning of the system whose theoretical function achieves the optimum outcomes. For instance, a McKinney report expresses the dangers of not getting rid of potholes and the hope of "patching the pavement" for educators and employers to solve the future problem of unemployment (Barton et al. 2013: 54).

> In all nine of the countries we studied, the road from education to employment is under constant repair. Signs are missing and the traffic is heavy. Drivers tend to concentrate on the patch of pavement ahead, not on the long haul. The result, … only a small fraction of young people and employers reach their destination in a reasonably efficient manner. The situation is not hopeless. Not only do many educators and employers accept that they need to be part of the solution, but many also have proved distinctly ingenious in filling in some of the potholes. (Barton et al. 2013: 54)

The pathways and highways perform to achieve the optimum state of harmony and consensus. They are assembled and connected in the grid of principles that place the theoretical relation of equilibrium with disequilibrium as social and cultural distinctions in the assessments and numbers that rank, differentiate, and divide qualities and characteristics of children's home environments, positioned as double gestures.

16.5.1   Some Concluding Thoughts

I began with the Siren's songs as dangerous, 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 beckoning today is expressed as benchmarks and "scientific evidence." They embody salvation themes about national development and individual happiness that has particular limits in thinking about change and the making of people and society. The international assessments are anticipatory, a calculated rationality that has a utopic image but that image is within a particular historical configuration. The international assessments are anticipatory as 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.

The numbers are not merely describing and correlating. They are anticipatory. The future is calculated as desires that have algorithmic formats that are prefigured in the abstraction of the school as a system. That future entails a comparativeness that differentiates normalcy and pathology as gestures of hope and fear.

References

Barber, M., & Mourshed, M. (2007). *How the world's best-performing school systems come out on top*. Chicago, IL: McKinsey & Company.

Barton, D., Farrell, D., & Mourshed, M. (2013). *Education to employment: Designing a system that works*. Retrieved from http://www.mckinsey.com/industries/social-sector/our-insights/education-to-employment-designing-a-system-that-works.

Borgonovi, F., & Przemyslaw, B. (2016). An international comparison of students' ability to endure fatigue and maintain motivation during a low-stakes test. *Learning and Individual Differences, 49*, 128–137.

Bürgi, R., & Tröhler, D. (2018). Producing the 'right kind of people': The OECD education indicators in the 1960s. In S. Lindblad, D. Pettersson, & T. Popkewitz (Eds.), *Education by the numbers and the making of society: The expertise of international assessments* (pp. 75–91). New York: Routledge.

Diaz, J. (2017). *A cultural history of reforming math for all: The paradox of making in/equality*. New York: Routledge.

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 (pp. 222–236, 347–348)*. Stanford, CA: Stanford University Press.

Hall, G. S. (1904/1928). *Adolescence: Its psychology and its relation to physiology, anthropology, sociology, sex, crime, religion, and education* (Vol. 1). New York: Appleton.

Halpern, O. (2014). *Beautiful data: A history of vision and reason since 1945*. Durham, NC: Duke University Press.

Hansen, H., & Vestergaard, A. (2018). On the contest of lists and their governing capacities: How "tax havens" became "secrecy jurisdictions". In S. Lindblad, D. Pettersson, & T. S. Popkewitz (Eds.), *Education by the numbers and the making of society: The expertise of international assessments* (pp. 35–52). New York: Routledge.

Heyck, H. (2015). *The age of system: Understanding the development of modern social science*. Baltimore, MD: Johns Hopkins University Press.

Latour, B. (1986). Visualization and cognition: Thinking with eyes and hands. *Knowledge and Society: Studies in the Sociology of Culture Past and Present, 6*, 1–40.

Lindblad, S., Pettersson, D., & Popkewitz, T. (Eds.). (2018). *Education by the numbers and the making of society: The expertise of international assessments* (pp. 222–228). New York: Routledge.

Lindblad, S., Pettersson, D., & Popkewitz, T. S. (2015). *International comparisons of school results: A systematic review of research on large-scale assessments in education.* Delrapport från SKOLFORSK-projektet, Vetenskapsrådet. Stockholm. Retrieved from https://publikationer.vr.se/produkt/international-comparisons-of-school-results-a-systematic-review-of-research-on-large-scale-assessments-in-education/.

Mourshed, M., Chijioke, C., & Barber, M. (2010). *How the world's most improved school systems keep getting better*. Chicago, IL: McKinsey & Company.

National Council of Teachers of Mathematics. (1945). The second report of the commission on post-war plans: The improvement of mathematics in grades 1 to 14. *The Mathematics Teacher, 38*(5), 226–232.

OECD. (2017). *PISA 2015 Results (Vol. III): Students' well-being*. Paris, France: OECD Publishing. Retrieved from http://www.oecd.org/pisa/publications/pisa-2015-results-volume-iii-9789264273856-en.htm.

Pont, B., Donaldson, G., Elmore, R., & Kools, M. (2014). *The OECD-Sweden education policy review. Main issues and next steps*. Paris, France: OECD.

Popkewitz, T. (2018). Anticipating the future society: The cultural inscription of numbers and international large-scale assessment. In S. Lindblad, D. Pettersson, & T. Popkewitz (Eds.), *Education by the numbers and the making of society: The expertise of international assessments* (pp. 222–228). New York: Routledge.

Popkewitz, T. (in press). *The impracticality of practical research: A history of contemporary sciences of change that conserve*. Ann Arbor, MI: University of Michigan Press.

Popkewitz, T. S. (2008). *Cosmopolitanism and the age of school reform: Science, education, and making society by making the child*. New York: Routledge.

Porter, T. (1995). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.

Thomas S. Popkewitz is a professor in the Department of Curriculum and Instruction, The University of Wisconsin-Madison. His studies are concerned with the systems of reason that govern curriculum reforms, the sciences of education, and teacher education. The research works cross the fields of curriculum studies, the political sociology of education, and cultural history—to consider the politics of educational knowledge and paradox of exclusion and abjection in efforts to include. His current research focuses on international assessments of education. The research explores how numbers are cultural practices that express universal principles about the kinds of people and society that perform to register differences and a history of present educational research concerned with practical and useful knowledge, tracing historically the paradoxes of such research as defining change to stabilize and conserve. E-mail: thomas.popkewitz@wisc.edu.