

# MEASURING THE SOCIAL WORLD<sup>1</sup>

Dietmar K. Pfeiffer<sup>2</sup>

## Abstract:

Quantification is no longer a practice of natural science only but has become part of human sciences and everyday life as well. As a direct measurement, which fits the axiomatic of the representational theory of measurement, is mostly infeasible in the social sciences, indicators are used and frequently aggregated to indexes. The index scores can be used to construct a ranking of the units (HDI, PISA). Measurement level and meaning of the data remain often unknown (pseudo-metrical scaling). Furthermore many variables in the field of education are not quantifiable. In the course of globalization and international competition, an indicator system for educational quality measurement (PISA, TIMSS) and goal achievement (Education for All, Millennium Development Goals, Sustainable Development Goals), was established and exerts an increasing impact on national educational systems. The main concerns of indicator-based steering are its methodological limitations and the transformation from a descriptive information base to a normative control system.

**Keywords:** index, indicator, measurement, PISA, educational management, globalization.

## The culture of quantification

From a historical perspective, scientific progress is closely linked to the measurement of phenomena, objects, and relations. This applies in particular but not exclusively to the natural sciences. In accordance with the representational measurement theory, measurement is “a process of assigning numbers to objects in such a way that interesting qualitative empirical relations among the objects are reflected in the numbers themselves as well as in important properties of the number system” (Townsend & Ashby, 1984). This means that real conditions, properties or structures are converted into numbers and formal relationships (Saint-Mont, 2011:13; Jarry & Pfeiffer, 2017:175). The process of quantification is carried out using special and often complex instruments. Not only sciences, but our everyday life is also pervaded by measurements and quantifying statements, often without us recognizing them as such: blood pressure, monthly income, leisure time, distance to work. To say it by Hand (2005): “We live in a world of measurement”.

The necessity and usefulness of measurements in the natural sciences, business or sports can hardly be denied. After all, we all want our smartphone to work, the refrigerator to

---

<sup>1</sup> The author is grateful to Mariama da Costa Ireland for the competent revision of the English version of the text.

<sup>2</sup> Possui mestrado em Sociologia - Westfälische Wilhelms-Universität Münster (1969) e doutorado em Sociologia - Westfälische Wilhelms-Universität Münster (1972). Membro do Centro de Estudos Brasileiros da Westfälische Wilhelms-Universität Münster. Consultor internacional (Brasil, Chile, Cuba, Moçambique, Malawi, Sri Lanka) nas áreas de economia educacional, planejamento e avaliação educacional, management de qualidade educacional, métodos empíricos e estatísticos de pesquisa educacional. dicape@uni-muenster.de

use as little energy as possible and the plane not to crash in turbulences. However, the basis for all these and many other achievements of our everyday life that are taken for granted today are based on scientific laws and their formalization in complex mathematical models.

In recent years, however, a growing trend towards quantification has also been observed in numerous social contexts. Indicators, indices, ratings, rankings, assessments, scorings, impact factors, coefficients, etc. are becoming central instruments not only in the social sciences but also in everyday social life. Nevertheless, the attempt to make measurable what is measurable only indirectly, at best, is not completely new. Already in the 18th century, there were attempts to rank painters and writers on the basis of certain indicators (Spoerhase, 2014). And even then, the result already depended on which categories - i.e. indicators - were considered as relevant. In the meantime, however, there are no limits anymore and all is measured whatever it takes: hospitals, universities, cities, countries, students, and teachers. And this, although it is, or should be known that measurements of social and mental phenomena rarely correspond to the axiomatic of representative measurement theory (Pasquali, 1996:23-25). It remains therefore often unclear whether the phenomena in question can be measured at all and, if so, at what level of scale. In the words of Cicourel (1964), it is mostly 'measurement *by fiat*'.

### **Measurement by indicators**

Felbinger (2000:17) names the following as the main problems of measurements of social phenomena:

- Lack of an international system of units (SI) as defined for physical quantities which allows comparability of measured values.
- Ambiguous concepts that can be defined in one way or another (poverty, intelligence).
- Complex concepts (constructs) which are not directly measurable and some of which are not measurable at all (educational quality, social justice).
- A great potential of systematic and random measurement errors, some of which cannot be determined at all.

These problems make measurement *stricto senso* practically impossible. In consequence, social sciences and policymakers in public administration, economics, and other fields operate with measurable indicators which are assumed to be empirical equivalents of the phenomena described by the construct. An important distinction refers to the difference between formative and reflective indicator measurement. In the first case, the indicators are the 'building units' of the construct, in the second case, the construct (latent) variable is the cause of the indicators (manifest variables).

In Brazil, for example, the *quality of learning processes* is measured using two indicators: the school performance rate (progression rate) and the average performance in the national exams. The two indicators are used to calculate the Index of Basic Education Development (Índice de Desenvolvimento da Educação Básica – IDEB).

On the other hand, *intelligence* is measured by the reaction (answers) to certain stimuli (questions), and the *(un-) satisfaction with a product* is measured by the evaluation that customers give on a multi-point satisfaction scale. Here the relation is inverse: the indicators reflect the constructs (answers, assessments). In both cases, however, numerous logical and mathematical problems arise.

In contrast to natural science measurements, where the value visible on an instrument is itself the measured value, in the social sciences there is a fundamental separation of measured variables and observable indicators. People produce a score when asked about their shoe size or age, but only because they know the score. If you ask them whether they are satisfied or whether they care about their family, you do not get a measured value, but an indicator. How to get from indicators to measured values, that is precisely the subject of measurement theories (Rost, 2005:6.).

Against this background, the choice of indicators that represent or reflect the concept is necessarily selective and must be based on theoretical considerations or normative guidelines. Indicators often enjoy the nimbus of objectivity, but they are in reality the result of scientific or administrative ideas about what is considered theoretically relevant, normatively desirable or valuable (Mau, 2019:205). For example, Rohwer & Hülsewig (2011:73) show that the results of studies on the "international competitiveness" of countries differ significantly depending on the theoretical concept of competitiveness on which they are based; an example that can be applied directly to other areas.

Indicators thus operate like a filter that focuses on certain aspects of reality and hides others, because the number of indicators is in principle infinite (Pfeiffer & Püttmann, 2018:44). Indicators can therefore always be the subject of criticism; moreover, they are historically and spatially relative. "O indicador útil em determinado lugar e época pode não ser proveitoso para outra região, ou até mesmo para própria região em outro momento" (Soligo, 2012:17). For example, even in the 19th century, the moral level of a population was measured by the number of children born out of wedlock (Pogliano, 1988: 274).

However, not only the selection of indicators is relevant for the measurement result, but also the measurement model by which they are processed. Usually, measurements using indicators are *multilevel* processes, since the targeted aspects of reality are too complex to be captured one-dimensionally. This means that the values of the individual indicators are aggregated into a *composite indicator* or **index**. This is a complex process of selecting, weighting and aggregating data. An internationally known example is the **Human Development Index (HDI)**, which is made up of the indicators: income, school years and life expectancy, which are first normalized on a scale of 1 to 100 and then aggregated by linear addition. In accordance to Siche *et al* (2007:139), "um índice é o valor agregado final de todo um procedimento de cálculo onde se utilizam, inclusive, indicadores como variáveis que o compõem". The result is a score, which is often used to create a ranking scale, a dubious trend, that nowadays nobody and nothing can escape. These rankings are a kind of **pseudometric scaling**, but most people misunderstand them as real metrics.

At this point, the question immediately arises, especially when it comes to international indicator systems, which are becoming increasingly important in the course of globalization, which organizations, experts or scientists actually have the power of definition, who develops and applies the indicator systems and the algorithms which are used to filter and measure the social reality. In the following, this question will be examined using educational indicators.

### **Measurement of Education**

While in economics, for example, formative indicators and measurements by index have a long tradition and are considered an important instrument for reducing a large amount of information and data (van Suntum, 2011), until not so long ago educational science

<i>Revista Dialectus</i>	Ano 9	n. 16	Janeiro – Abril 2020	p. 113 - 122
--------------------------	-------	-------	----------------------	--------------

resisted the trend towards standardised metrisation of competences, educational quality, and other pedagogically relevant parameters.

Key features of educational systems such as enrollment rate, school attainment, pupil-teacher ratio (Cardoso & Steiner-Khamsi, 2017:388) were already recorded in the 19th century within the framework of the international comparative educational science of Julien de Paris (1775-1848). Nevertheless, it was only in the course of globalization and increased international competition that the definition of educational goals (Education for All - EFA, Millennium Development Goals - MDG, Sustainable Development Goals - SDG) and comprehensive indicator systems was developed and implemented at the international level.

Clearly, two phenomena have achieved unprecedented prominence since the 1990s: international commitments (MDGs, EFA) monitored with data produced by the UNESCO Institute for Statistics and analysed by the Global Monitoring Report; and the growth of international large scale assessments (ILSAs). However, the two did not converge until the recent inclusion of learning outcomes in the SDGs (Cardoso & Steiner-Khamsi (2017:398).

The internationally best known of these learning assessment programmes is the Programme for International Student Assessment (PISA). Every three years 15-year-old students from all over the world are tested in reading, mathematics, and science. The tests claim to assess key competencies and to gauge how well the students master real-life problems and how knowledge is applied. The test items are reflective indicators because the latent variable "competences" to be measured determines the probability of a correct answer to the questions.

Thirty-two countries took part in the first test in 2000, the majority of them were OECD members. In 2018, over half a million 15-year-olds from 80 countries took the PISA, ten of them Latin American countries. The question here is not whether ILSAs are at all capable of measuring what they claim to measure (validity) or whether "competences" is a meaningful pedagogical concept, and what is its relationship to other educational concepts (knowledge, skills, abilities). What is at stake is the fact that an international regime of experts without democratic legitimacy has managed to define what constitutes a good education, in part against considerable national resistance. An international educational regime is emerging (Parrera do Amaral, 2011) that does not take into account national educational traditions, values, and cultural characteristics. The standards of the OECD will determine what is taught and how for a long time (Mau 2018: 203). Critics already fear the

danger of a narrowing of curricula and teaching methods (Meyer & Benavot, 2013) and an indirect preference for specific learning techniques. Internationalization, standardisation, measurability, and comparability of learning outcomes have top priority, whether or not they meet national needs.

An international curriculum based on an international large-scale assessment could further result in a misalignment between the education system and a country's needs in relation to its stage of economic development and labor market demands (Rose, 2015:488).

All measurement is based on comparison and in this view indicators and indices serve to make qualitatively different things comparable. The basic operation of comparison (Matthes, 2011), however, already presupposes quantities, so that the question must be asked why comparisons are made at all. It is the experience of differences, whereby these can be perceived quite differently by people. The results of such comparisons can become important for scientific knowledge and policy learning, providing information to the public, transparency, and guidance. However, it can be observed that in recent years an increasing shift from the descriptive recording of phenomena to normative statements and pre-critical statements has occurred (Pfeiffer, 2017: 151). Although the OECD repeatedly emphasises that its data do not serve to monitor but rather to provide political support for the country, it must be noted that the flood of data published annually in the report "Education at a Glance" (2019) as well as the PISA Reports (2019) are accompanied by instructions, praise, and criticism of countries whose theoretical and normative foundations are certainly not consensual. It is therefore not surprising that educational data provided by the OECD are interpreted very differently by political actors and the public.

It may be that in a globalising world, certain cross-cultural general basic competencies are indispensable for an individual's self-determination, and it is certainly important for every country to be informed about the knowledge, competences, and skills of its students. The problem with international measurement and indicator systems, however, is that they are limited to international competitiveness, rankings and what can be measured (Llavador & Aguilés, 2010). However, learning processes, educational experiences and biographical episodes of the people are not measurable and the attempt to measure them results in mathematical-statistical artifacts without reference to reality. Against this background, it is doubtful whether the measurement of learning outcomes, without

considering process-related and contextual conditions, can be a sufficient indicator for the quality of schools and entire educational systems. "Para comprender la realidad educativa de un contexto se ha de conocer sus condiciones de posibilidad, de no ser así, los resultados se naturalizan" (Llavador & Aguilés 2010: 140).

The general pressure to quantify has also affected the science system and its actors. Scientific success and reputation of researchers today are mostly measured by the H-index, based on the set of the scientist's most cited papers and the number of citations that they have received in other publications. Whether this simple performance metric reflects the quality of scientific work or only the resonance of the author achieves and his self-marketing competence is left open.

Even the universities themselves, once the epitome of a self-controlling professional system, have now subjected themselves to rating and ranking pressure to satisfy the information needs of their social environment. They have become brands with specific features that compete with each other for reputation, students and financial resources. Here, too, the question arises how the indicators used to determine the quality of a university are selected and weighted. These depend on which model of a higher education institution is used. The well-known Times Higher Education Ranking, for example, uses a total of 13 indicators from 5 areas (Teaching, Research, Citations, International Outlook and Industry Income), which are aggregated to an overall score using a rather complex model.

**Final considerations**

The triumph of measurements by means of indicators in all social areas, despite their often dubious validity, is based on their image of objectivity, transparency, and simplicity, which they convey. They satisfy the demand for quantitatively condensed information, which in turn can be used for a supposedly rational control of complex systems. Whether such rationality is always given is open to question and it's often overlooked that measurement has not only a technical but also an epistemological dimension.

This approach has a long tradition in business and administration while attempts to improve the management of education systems and the quality assurance of learning outcomes by means of indicator systems are more recent. They have arisen with the intensification of international competition, rising demands by various stakeholders, scarce

<i>Revista Dialectus</i>	Ano 9	n. 16	Janeiro – Abril 2020	p. 113 - 122
--------------------------	-------	-------	----------------------	--------------

public funds and public demands for the rational use of resources in an area that is largely beyond the control of market mechanisms. The availability of more and more data was a prerequisite for this process.

The question that arises is whether measurement, evaluation, and accountability can actually make a significant contribution to improving schools, the quality of learning processes, and learning outcomes (Klees, et al., 2019). The specific characteristics of *people processing organizations* in general and schools, in particular, were already analyzed thirty years ago by Meyer & Rowan (1977, 1978). They concluded that simple input-process-outcome models are not appropriate for learning processes, because the end means relation is complex and under-determined. People are not trivial machines and the pedagogical hope for outcomes is uncertain. Material and human resources are certainly important, bad management and untrained teachers hinder school development, but ultimately the curiosity of the students and their willingness to learn is decisive for learning success.

Consequently, the causality between pedagogical intervention and the outcome of learning processes is limited, and the possibilities of controlling the educational system by means of permanent measurements, interventions and better alignments of the actors (World Bank, 2018) are limited, especially since essential causes of learning success are to be found outside the educational system. The experience and results in numerous countries show the limits of a framing that relies primarily on indicators, measurements and management methods.

**Referências:**

Cardoso Manuel & Steiner-Khamsi, Gitta (2017). The making of comparability: education indicator research from Jullien de Paris to the 2030 sustainable development goals. **Compare, A Journal of Comparative and International Education**, 47 (3), 388-405. <http://dx.doi.org/10.1080/03057925.2017.1302318>.

Cicourel, Aaron V. (1964). **Method and measurement in sociology**. New York: Free Press of Glencoe.

Felbinger, Sabine (2010). **Messtheoretische Grundlagen**. LMU München: Institut für Statistik. <http://thomas.userweb.mwn.de/Lehre/wise1011/SeminarSozi/SabineFelbinger.pdf>.

Hand, David J. (1996). "Statistics and the theory of measurement". **Journal of the Royal Statistical Society, Series A**, 159 (3), 445-492.

Klees, Steven J.; Stromquist, Nelly, P.; Samoff, Joel; Vally, Dalim (2019). The 2018 World Development Report on Education: A Critical Analysis. **Development and Change**, 50 (7), 1-18. DOI: 10.1111/dech.12483.

Jarry R., Roberto & Pfeiffer, Dietmar K. (Col.). 2018. **Pesquisa Social**. Métodos e Técnicas. Editora Atlas: São Paulo.

Llavador, José Beltrán e Aguls, Alícia Villar (2010). Avaliação em Educação: algumas considerações a respeito dos indicadores" **Educação & Linguagem**, 13 (21), 132-149.

Matthes, Joachim. The Operation Called 'Vergleichen' (1992). **Soziale Welt, Göttingen, Vol. especial, n. 8**, p. 75-99.

Mau, Steffen (2018). **Das metrische Wir**. Über die Quantifizierung des Sozialen. Bpb: Bonn

Meyer, Heinz-D. & Bemavot, Aaron (eds.) (2013). **PISA, Power, and Policy**: the emergence of global governance. Oxford: Symposium.

Meyer, John W. & Rowan, Brian (1977). Institutionalized Organizations: Formal Structure as Myth and Ceremony. **American Journal of Sociology**, 83 (2), 340-363.

Meyer, John W. & Rowan, Brian (1978). The Structure of Educational Organizations. In *Marshall W. Meyer (ed.): Environments and Organizations*. San Francisco: Jossey-Bass.

Ministério de Educação. *Ideb - Apresentação*. <http://portal.mec.gov.br/secretaria-de-educacao-basica/programas-e-acoess?id=180>.

OECD (2019). **Education at a Glance 2019**. OECD Indicators. OECD Publishing: Paris. <https://doi.org/10.1787/f8d7880d-en>.

OECD (2019), **PISA 2018 Results (Volume I): What Students Know and Can Do**. OECD Publishing: Paris. <https://doi.org/10.1787/5f07c754-en>.

Parreira do Amaral, Marcelo (2011). **Emergenz eines Internationalen Bildungsregimes?** International Educational Governance und Regimetheorie. Münster-New York: Waxmann.

Pasquali, Luiz. (2009), *Psicometria*. **Revista da Escola de Enfermagem da USP**, 43, 992-999. <http://www.revistas.usp.br/reeusp/article/view/40416>.

Pfeiffer, Dietmar K. & Püttmann, Carsten (2018). **Einführung in die Methoden empirisch-pädagogischer Forschung**<sup>4</sup>. Münster-New York: Waxmann.

Pfeiffer, Dietmar K. (2017). As medidas da educação. Medição e gestão mediante indicadores. **Revista de la Asociación de la Sociología de la Educación**, 10 (2), 141-155.

Pogliano, Claudio (1988). La misura e l'ordine. L'aurora del pensiero quantitativo. **Belfagor XLIII (3)**, 255-276.

Rose, Pauline (2015). Is a global system of international large-scale assessments necessary for tracking progress of a post-2015 learning target? *Compare: A Journal of Comparative and International Education*, 45/3, 486-4890. <https://doi.org/10.1080/03057925.2015.1027514>.

Rost, Jürgen (2005). Messen wird immer einfacher. **ZA-Information**, 56, 6-7.

Rohwer, Anja & Hülsewig, Oliver (2011): Länderrankings: Zu welchen Ergebnissen kommt die neuere Forschung? In: Rankings und Indikatoren: Hilfreiche Kennzahlen oder Verkürzungen mit fatalen Folgen? **Wirtschaftsdienst**, 91 (11), 735-759. DOI: 10.1007/s10273-011-1295-6.

Saint-Mont, Uwe (2011). **Statistik im Forschungsprozess**. Berlin-Heidelberg: Springer Verlag.

Soligo, Valdecir (2012). Indicadores: Conceito e complexidade do mensurar em estudos de fenômenos sociais. **Estudos em Avaliação Educacional**, 23, n.52, 15-25.

Spoerhase, Carlos (2014). Das Maß der Potsdamer Garde. Die ästhetische Vorgeschichte des Rankings in der europäischen Literatur- und Kunstkritik. In: Wilfried Barner et al. (eds): **Jahrbuch der Deutschen Schillergesellschaft (58)**. Berlin: de Gruyter.

Siche, Raúl, et al. (2007). Índices versus Indicadores. Precisoões conceituais na discussão da sustentabilidade de países. **Ambiente e Sociedade**, Vol. X (2), 137-148.

Van Suntum, Ulrich (2011). Keine Scheindebatten führen. In: Rankings und Indikatoren: Hilfreiche Kennzahlen oder Verkürzungen mit fatalen Folgen? **Wirtschaftsdienst**, 91/11, 735-749. <http://dx.doi.org/10.1007/s10273-011-1295-6>.

Townsend, James D. & Ashby, F. Gregory (1984). Measurement Scales and Statistics: The Misconception Misconceived. **Psychological Bulletin**, 96, 394-401.

World Bank (2018). **World Development Report 2018**. Learning to realize education's promise. Washington, DC: World Bank.