

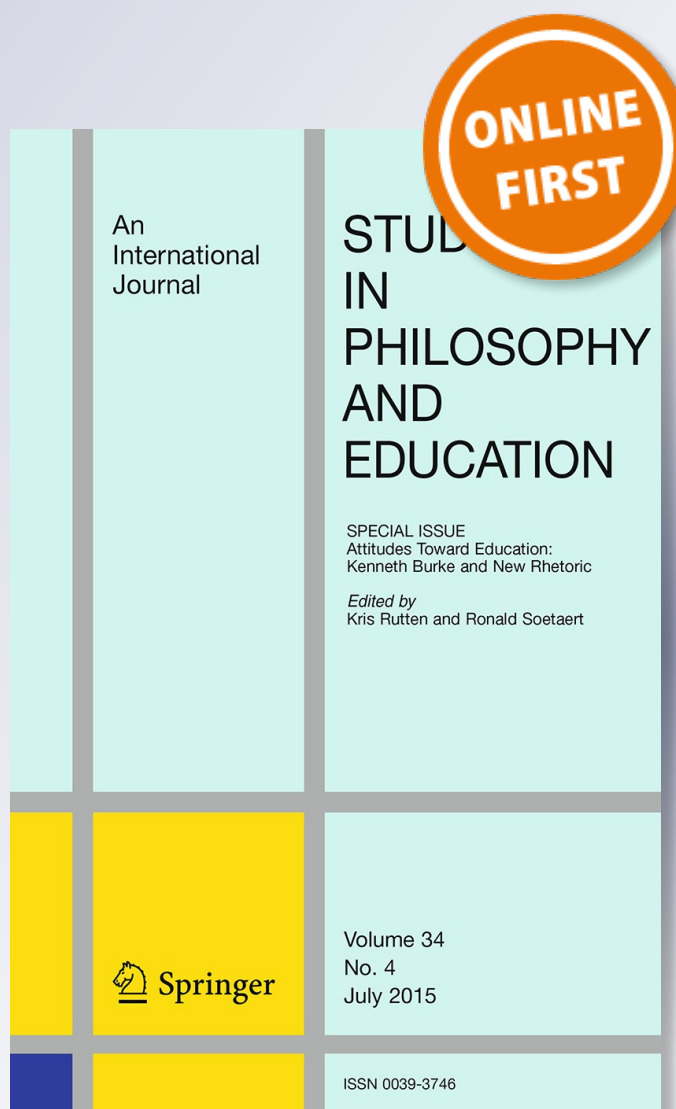
Campbell's Law and the Ethics of Immensurability

Alexander M. Sidorkin

Studies in Philosophy and Education
An International Journal

ISSN 0039-3746

Stud Philos Educ
DOI 10.1007/s11217-015-9482-3



Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media Dordrecht. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Campbell's Law and the Ethics of Immensurability

Alexander M. Sidorkin¹ 

© Springer Science+Business Media Dordrecht 2015

Abstract The paper examines “Campbell’s Law”: “The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor.” The examination of measurability leads to explaining the reason for existence of a class of unmeasurable phenomena. The author describes a kind of habitus in which a strong taboo against measuring must exist by necessity, not by choice. The taboo is, in effect, a result of degradation of a certain kind of habitus. And finally, the paper demonstrates under which conditions the Campbell’s law is in effect, and how we can mitigate its effects in social decision-making.

Keywords Campbell’s law · Bourdieu · Measurability · Immensurability

Campbell’s Law

One of the founding fathers of the program evaluation field Donald T. Campbell came up with the following jaded law:

The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor (Campbell 1976).

Campbell’s own examples are as numerous as they are compelling. Measuring effectiveness of police force by the proportion of crimes solved leads to “Failure to record all citizens’ complaints, or to postpone recording them unless solved.” It

✉ Alexander M. Sidorkin
asidorkin@hse.ru

¹ Institute of Education, National Research University – Higher School of Economics, 20 Myasnitskaya St, Moscow 101000, Russia

encourages police pressure on criminals to confess crimes they did not commit in exchange for plea bargains: police wants to count those crimes as solved, at a price of reduced sentences for the actual crime. The problem may have gotten worse since then, as the methods of recording and calculating statistics improved (This American Life 2010). Campbell also examined an early version of academic achievement testing: “when test scores become the goal of the teaching process, they both lose their value as indicators of educational status and distort the educational process in undesirable ways (Campbell 1976).” In a much darker example, the Vietnam era administration switched from *estimates* of enemy casualties (which were utterly unrealistic) to enemy *body counts*. The bureaucratic feat, he argues, contributed to the My Lai massacre of 1968.

In scholarship, the law has not been highly visible, although various concerns about the proper use of data abound. One notable exception is the book by Nichols and Berliner (2007), where the law was used as a polemical device against standardized testing. The authors offer a considerable insight into why the practice of measurement took root in American education. Unfortunately, they did not attempt to explain what makes it work. We all could take the law more seriously if Campbell himself explained the reason for its existence, but he did not. He implies it is the human propensity to game the system that is at work. However, it is more complicated than that. At least in some social systems, Campbell’s law is not a result of trivial cheating, but of a degradation of certain habitus which results in ethical failures.

Why may this be important? Campbell’s claim, if taken seriously, is striking. He does not say that quantitative measures are inaccurate. He says that the use of measurable indicators in decision-making corrupts the practice it intends to improve. If he is right, people involved in the same practices will demonstrate less ethical behaviour and/or less effective results, if evaluated with quantitative indicators. This is hard to swallow, for uses of data in decision-making are growing, as data technologies improve. Are we making such social practices as education worse off? It seems impossible to stop measuring social practices, or to stop using the data to make decisions. And yet, if making those decisions damages the practices, perhaps stopping is not entirely out of the realm of conceivable options. Or perhaps certain measures to mitigate the ill effects of Campbell’s law can be devised. It is very difficult to do without a theory of Campbell’s law, without an attempt to understand why it works.

Campbell himself was definitely not against measurements, but he doubted that the use of data in decision-making, especially in performance evaluation, can ever be safe from corrupting influences. People evaluated on the basis of data, know how they are being evaluated, and consciously or subconsciously change their activity or manipulate the data; both often with negative consequences. The problem lays not so much with decision-makers, as with the evaluated, and with their strategic response to being evaluated on certain measured indicators.

All will agree that things described by Campbell do *sometimes* happen; the important question is whether it happens most of the time, and it happens *by necessity* rather than by chance or by error. If yes, what is the force that drives it, what is the cause of its existence? Or is this a systematic error of execution that can be fixed with more appropriate instruments and their improved use? To present the dilemma as an analogy: if a drug has side effect, at which point does it become too strong to outweigh the benefit of the drug? To answer the question, we need to understand *why* the phenomena described by Campbell happen. But first, let us briefly consider other, non-Campbellian critiques of measurement.

The Improper Uses of Measurement

In the context of the education reform, the critique of standardized testing is common. One more recent version has been articulated, for example, by Daniel Koretz, who formulated the *sampling principle* of testing: “test scores reflect a small sample of behaviour and are valuable only insofar as they support conclusions about larger domains of interest (Koretz 1998, 21).” Drawing on his work, Richard Rothstein laid out the consequences: Standardized tests cannot be measuring the entire curriculum, because testing would take as much or more time as instruction. Sampling will inevitably narrow the curriculum. Teachers will react by teaching only the part of the curriculum that is being tested. The problem does not exist in such tests as National Assessment of Educational Progress (the NAEP), where different assignments are spread through a large sample of pupils—we don't need to know how each child is doing, and are interested only in populations. However, the narrowing of curriculum problem will always exist in high stake testing used with all pupils (Rothstein 2010).

This sounds like a version of Campbell's law, but it is not. The Koretz/Rothstein critique of the use of measurement does not have law-like properties. It may follow from the critique that we must apply more robust measurements with more rigorous validation process. It may also follow that if we switch to more NAEP-like sampling techniques, they will reduce and eliminate the curriculum-narrowing effect. It is quite feasible, for example, to use NAEP tests to evaluate performance of the entire states' educational systems. There are still incentives for states to manipulate samples, but it would be impossible to evaluate individual teachers' and even individual school's performance. Therefore, it won't affect teacher behaviour.

A similar critique is the extension of the sampling problem: the available instruments measure only narrow area of relatively simple cognitive skills. They cannot yet reliably measure creativity, social, emotional, and complex cognitive skills such as ability to solve problems, to analyse and synthesize information, etc. Such critique similarly does not mean that new measures cannot be invented and existing ones cannot be improved.

The kinds of criticism described above can be found among both the supporters and the opponents of the school accountability reform. While the proponents recognize problems with high stake testing, they reasonably call for fixing them, for coming up with better, more sophisticated testing techniques. In other words, the side effects of the drug called measurement are thought to be possible to control. If only we were able to have better standards and pull states' resources to produce better tests—the thinking goes—everything would be alright, and we will get better data and little corruption of teaching practices.

The critique above may have merit, but this paper is not considering it. The point of the next section is just to show that Campbell's concern is much more fundamental.

What is Measuring?

Measurability is not a natural quality of an object itself; it describes a certain relation between human beings and an object. What measuring and counting do is making—in the human mind—something comparable to other similar things in a certain way. Counting is not simply putting things in the same class by naming them—a nut tree, an antelope, etc. It is a kind of comparison that makes things interchangeable. Hunting and gathering societies rarely develop counting systems beyond four or five (Premack and Premack 2003, 29). A

herd of animals is just that—a heard, many animals; no need to know how many. There is no way to store and therefore no need to hunt more than the group can eat. Only when accumulation and exchange of surplus product begins in agricultural societies, people arrive at the idea that a sack of corn is essentially the same as another sack of corn, and they can be considered interchangeable. Economists call these “fungible” goods. The interchangeability is the underlying assumption of any counting system, expressed in the concept of the unit of measurement. Measuring is just another side of counting. One sack of wheat equals four or five buckets, which still implies that the unit of measurement—the bucket—is essentially equal and can be replaced by any other bucket. It also implies that the stuff mixed in the sack is homogeneous, mutually replaceable. It is fairly obvious that a sack of wheat may not be exactly identical to another sack: wheat can be of better quality, density, or appearance. The property of measurability reflects no specific physical phenomena, but the social convention that exchangers will ignore these details for the sake of convenience and to facilitate trade.

However, certain things in life remain to be valued only as unique ones, and cannot be exchanged or replaced. For example, the point of having more than one friend is that two or more friends would all offer qualitatively different kinds of connection. Friends are valued for their uniqueness, not for their sameness. Three acquaintances do not replace one close friend. The mathematics functions simply do not apply here, because the items are not fungible. In this sense, the applicability of mathematics to things is a function of economics.

For another example, romantic love in its contemporary Western context postulates not only uniqueness, but also exclusivity of the relationship. People are aware that a romantic partner can be replaced, and one can be romantically involved with several people at the same time. These kinds of things happen all the time, but they are not discussable in the context of a romantic relationship. Love is immensurable in the sense that measuring it threatens the existence of the relation. The value of certain phenomena is derived through uniqueness, not through quantity. And that value is negated, in fact destroyed when we attempt to compare and measure. Even though theoretically everything *can* be measured, it does not mean that everything *should* be measured, lest we agree to destroy the measured.

How is value destroyed by measurement? Following the language and the logic of Pierre Bourdieu, the fact of certain good's interchangeability is collectively suppressed. Bourdieu makes the case describing the collectively suppressed truth in his theory of the gift. He writes about the “structural truth,” that is, the common knowledge that any gift must be reciprocated (Bourdieu 1998, 94). However, special devices are erected to suppress this knowledge. For example, there should be a time interval, and prohibition to discuss the monetary value of the gift and the reciprocating counter-gift. He defines the “collective *self-deception*, a veritable *collective misrecognition* inscribed in objective structures (the logic of honour which governs all exchanges—of words, of women, of murders, etc.) and in mental structures, excluding the possibility of thinking otherwise (Bourdieu 1998, 95).” The notion of immensurability is the same kind of collective self-deception, and I mean it in a good way. It is just a description of the mechanism for the ethics of immensurability to function.

We do create certain “mental structures” that emphasize the uniqueness and incomparability of relations. They are designed to strengthen the institution of life-long monogamous marriage. Each romantic relationship is unique and may never be deemed one of many. The act of measuring reduces the measured phenomenon to its basic, common to other similar phenomena essence, and thus is an explicit threat of potential

replacement. The more marriage of that kind is under threat, the more active are the mental structures preventing us from comparing and measuring.

The taboo against measuring is not against measuring per se, but explicitly against knowledge resulted from such measuring. Campbell and other thinkers in the field of program evaluation seem to believe that it is possible to separate the act of measuring something from the act of making an evaluative decision about it. It is not a questionable assumption. Any measurement implies an identification of a certain feature of the measured object that is important for the measurer, and by extension, privileges that one feature, while suppressing all other features. Despite the appearance, decision-making does not occur after we measure and evaluate. To the contrary, the decision to measure and the decision what and how to measure, are really the most important ones; the decision of what to do with the data is secondary and often semi-automatic, which is not a decision at all. Decision-making precedes measurement, not the other way around. The data-driven decision-making is really a decision-driven data gathering. This fact explains why there exists the taboo is against *measuring* per se, not against the improper use of measurement data.

The taboo against measuring, like most taboos, is not applied consistently. Moreover, the existence of a taboo betrays a deep collective anxiety about the practice it intends to regulate. For example, measuring the value of human life is abhorrent; comparative measuring of different people's value is even more abhorrent. However, such calculations are routinely performed by insurance companies and judges. Under certain circumstances, the taboo against measuring is lifted, and the practice of measuring previously immeasurable things is allowed. The inconsistency confirms that the claim of immensurability has nothing to do with the physical impossibility of measurement, and everything to do with a prohibition against measuring. Moreover, the cultural impossibility of measuring is selective and context-dependent. Yet it also shows that the ethic of immensurability is not an accident, and that certain reasons exist to maintain it. In the next section, I will show that the decision allowing or disallowing measurement of certain things is not taken at will every time. Such a decision is governed by a set of pragmatic considerations deeply embedded into the social order.

The Roots of Immensurability

Bourdieu explains the duality of the taboo with the notion of different kinds of habitus—social and economic universes that employ different rules of the game. The implicit rules of social order become visible when it co-exists with a different kind of order. One of his more risqué examples is about affective relations: “Housewives who have no material utility or price (the taboo of calculation or credit), are excluded from market circulation (exclusivity) and are objects and subjects of feelings; in contrast, so-called venal women (prostitutes) have an explicit market price, based on money and calculation, are neither object nor subject of feeling and sell their body as an object (Bourdieu 1998, 106).” Consequently, there is a taboo against measuring labour provided by housewives, or material costs and benefits of marriage. Note that feelings are the means of enforcing the taboo, for the discourse of feelings makes the discourse of market value unthinkable.

Bourdieu makes a strong case that the denial of the truth about measurability is not *just* an act of deception or hypocrisy. The denial itself is *constitutive* (not only descriptive) of the relation. People who operate in non-measuring types of habitus are not deceived or

deceive others; they actually make this sort of habitus work. What may look like suppression of one kind of truth (everything can be measured), is really an act of creation of another kind of truth (there are things that should not be measured). Why do so many people choose the second truth over the first one? Bourdieu convincingly rejects the simplistic notion that the non-measuring habitus is *necessarily* backward or is fading away.

Bourdieu acknowledges that often a conflict exists between the two kinds of habitus, where they intersect in the same institution. He considers, among others, the institution of family.

We see that, contrary to economic reductionism à la Garry Becker, who reduces to economic calculation that which by definition denies and defies calculation, the domestic unit manages to perpetuate in its core a quite particular economic logic. The family as an integrated unit, is threatened by the logic of the economy. A monopolistic grouping defined by the exclusive appropriation of determinate kind of goods (land, the family name, etc.), is at the same time united and divided by property. The logic of the prevailing economic universe introduces, within the family, the rot of calculation, which undermines sentiment (Bourdieu 1998, 106).

The immensurability of relations exists to support one kind of habitus. It is mostly associated with familial relations. I will now argue that schools constitute a similar kind of habitus. They strongly depend on the collective self-deception about measurability of relations, especially relations between pupils and teachers. In other words, I will show that no matter how inconsistent and illogical, the prohibition against measurability of relations is constitutive of the school life. Within the schools' habitus, it makes much sense, even though *explaining* it from within the habitus is very difficult, precisely because of the denial of structural truth.

School as Habitus

In schools, relations are tied to work, and measuring relations would inevitably entail measuring performance. And vice versa, measuring performance directly reflects on relations. This is because school relationships are instrumental, not immediately affective. In other words, people enter relations in school context because they have to be there, and because they may share an interest, not because they chose each other's company. And for that reason, the relations are mediated primarily by the activity of teaching and learning, by doing things together. If we say that relations between a teacher and her pupils are wonderful, we will be inevitably and reasonably asked if this results in better academic achievement. School is a purposeful institution, public schools are supported by taxpayers, and all schools are regulated by public authorities. Even though we may have trouble agreeing on the purposes of public schooling, and even though we may want different things from it, no one is suggesting the entire purpose of schooling is to have pleasant relationships. Schools are not social clubs.

School is ultimately about work—work by pupils, teachers, teacher aids, and principals. All but pupils are paid for their work, and all depend on pupil compliance for their livelihood. The labour pupils perform is a response to the teacher's authority. However, both pretend that the labour is performed only for the sake of the pupil's own future. The deal is very similar to a common tribal practice of bringing tribute to the chief, while pretending that the tribute is meant for gods. Both parties are interested in denying the

knowledge of the real recipient of the tribute. The giver saves face and avoids the direct recognition of his subjugation by the chief. It is easier to maintain honour by pretending to submit to a supernatural being. The chief clearly benefits from the denial, because it adds an extra kick to his authority without the need to resort to violence.

The contemporary mass school is a twist on the common place practice of extracting tribute from people without hurting their feelings. Instead of God, we have the idealized vision of the pupil, and her imagined interests. However, in reality, the tribute goes to teachers, whose job is to extract certain amount of labour from pupils. This is how teachers earn their living. Teachers are paid, and pupils are not, although both parties contribute labour to the enterprise. Teachers immediately benefit from the steady supply of pupil labour, although in the long run, pupils also benefit by building their own human capital. Note that in tribute-based societies, similarly, all people eventually benefit from paying tribute. The tribute creates reserve wealth that can be used in case of emergencies, as social safety network, and to concentrate resources for public projects. Power, as we all have heard, is productive. However, a tribal man has no more intrinsic desire to part with his yams than a typical inhabitant of the industrialized world enjoys paying her taxes. Yes, in the long run, pupils benefit from their labour, but this fact by no means removes the need to compel them to work now. Labour of any kind has to be extracted (by force, by magic, by inspiration, by pay check) no matter how much it is in the future self-interest of the labourer.

Labour of pupils also needs to be organized, managed, assisted, and evaluated. Teachers definitely do all of these; they also often produce the supply of information to be processed. The extraction of labour from pupils is by far the most important and the most problematic of their functions. The early experiments with technology-assisted learning show that while almost everything can be automated, providing motivation for learning so far cannot. Teachers are needed to make pupils work; it is what Hardt (1999) aptly named the “affective labour” of teaching. We do not see it because we have too much at stake by ignoring it. The affective labour of teachers needs to stay invisible, otherwise we ruin the myth of self-motivated learner.

Openly presenting what Bourdieu calls the “structural truth” makes the arrangement untenable. What is the alternative? Let us just imagine a situation when the structural truth is revealed, and the myth of self-interested learner is abolished. Pupils acknowledge they are merely peons working for free so that teachers can earn their salaries. Teachers acknowledge that they hold no power whatsoever over pupils, can be replaced by MOOC's and are only needed as glorified prison guards. In the next step, we can try imagining the situation of reckoning, where pupils and teachers come to understand the quality and quantity of labour they both contribute, and fair levels of compensation, and accountability for the results, and the need for shared governance. While it is possible in theory, I have a hard time even vaguely picturing a school with fair labour practices in the modern, wage labour market sense of the word. No, we better avoid that kind of a conversation for now. In a very real sense, human societies are always based on pretence and on denial of truth. Or rather, they are based on manufacturing other kinds of truth.

The only way we know how to raise the young is by extracting deeply discounted, semi-feudal labour of children and their mothers. Until relatively recently—in historical terms anyway—that was also how we obtained most goods and services. And when there were no economic alternatives, slavery, bondage, indenture, peonage and all kinds of other bound labour arrangements seemed just as reasonable as they now seem outrageous. I don't want to speculate about the next industrial revolution that will make learning so incredibly cheap and efficient that we won't have to struggle compelling children to work for 17–20 years

for free. It may or may not happen soon; the fact is—it has not happened yet. In the contemporary society, making a new-born into an informed citizen, eager consumer, and a productive worker involves huge quantities of tedious manual labour, mainly the child's own. Nothing in the most recent technological developments suggests a radical improvement in pupil labour's efficiency.

The coexistence of two very different economic systems (the market-driven habitus and the school habitus) requires an elaborated cultural mechanism of dual truths, suppression of the truth, and other fog and mirror tricks. The valuable affective relationships between pupils and teachers are, in effect, pulleys and levers of the mechanism. Let us return to the analogy with tribute giving. The relationships between the tribal people and the chief are described in terms of honour, respect, and most importantly, affection. It would be absolutely horrendous for the chief to be subjected to an objective evaluation of his performance. Let's imagine a 360 evaluation, with numeric performance indicators related to the quality of leadership, and to redistribution of the tribute back to the tribal members, benchmarked to the most advanced neighbouring tribes. This would immediately destroy the cultural foundation of the chief's authority and in the worst case scenario render the entire political system ineffective. Similarly, the exposure of the structural truth in schools will destroy the educator's authority and may damage the entire political structure of schooling.

The Habitus of Immensurability

To examine how specifically measuring runs against the school habitus, let us consider another Bourdieu's device, the notion of the honour society. The honour economies are not necessarily opposed to market economies; in fact, the two can be strongly intertwined and complement each other. In reality, even the most effectively measurable monetary transactions involve a great deal of trust, which is one of the main goods of the honour economy. However, under certain circumstances, the two may clash and replace one another.

In honour economies, measurement is perceived as immoral, as an act of debasement. What does it mean, exactly? A gift is an attempt to create a social bond, an obligation; it is not to furnish someone with a useful item at a fair price. If I give you a gift, this makes us friends, which means that in case I need your help in the future, you must provide it to me. It is critical to keep the amount of future help as vague as possible: I may just drop by for a chat, or I may die and you will have to raise my child. That may be the range of possibilities. The vagueness has its definite economic purpose. It deals with probabilistic risks and needs, not with definite ones. Similarly, relationships of kin and marriage provide the broad-spectrum social support that is difficult or impossible to provide with insurance policies or retirement schemes.

In schools, too, vagueness of mutual obligations is equally important. It goes something like this: As a pupil, you owe me some work. And yes, it is for your own benefit. And yes, it is also for my benefit. I have authority over you, and it is sanctioned by the State (and sometimes by God), but it is really in your own interest, so it is sanctioned by you. Can I guarantee you the results if you apply yourself?—Yes, no, maybe; there is also the talent, and the ability. But you still have to give me your work.

Schooling belongs to the same broad class of honour societies described by Bourdieu. Teachers have little ability to enforce their authority directly: through violence or monetary

compensation. This is why they must act as nobility, and create and maintain the anti-measurement habitus. In another paper, Bourdieu explained that symbolic capital...

...becomes symbolically efficient, like a veritable *magical power*: a property which, because it responds to socially constituted 'collective expectations' and beliefs, exercises a sort of action from a distance, without physical contact. An order is given and obeyed: it is a quasi-magical act (Bourdieu 1998, 102).

Teachers struggle to control crowds of children and adolescents without law enforcement, monetary incentives, and the power to expel. Like feudal lords, they are keenly aware of their own vulnerability and paucity of power. What they must do is accumulate and project a certain kind of symbolic capital. Their work must look like a noble activity, where calculations of relative effectiveness is suppressed and explicitly forbidden. They are evaluators of pupil behaviour and performance, and therefore should be above evaluation. They compete with each other—not on efficiency, but on disinterestedness in efficiency. As a corporate body, teachers have a tremendous stake in preventing anyone from comparing their efficiency as instructors. Teachers have met the first clumsy attempts to evaluate them on objective criteria like the end of the world—because it *is* the end of a particular social world. But it would be a mistake to interpret teacher resistance as pure self-interest. It is also motivated by their desire to protect the habitus of schooling, and in a sense, protect children from the possible crumbling of the fragile social arrangement called “school.”

What people think and do depends on what kind of habitus they occupy. Why do many voters support educational accountability and teachers overwhelmingly reject it? Teacher resistance to accountability may be explained by their conservatism, self-interest, or the selfish influence of their unions. However it is very unlikely to be the case, for teachers come from the general population, mainly the middle class, whose sensibilities they in theory should inherit. But their habitus is different than that of the majority of middle class inhabiting offices and factories. Teachers resist because they engage in certain social practices where denial of measurability is constitutive of the key relation between pupils and teachers. Similarly, when a pupil performs poorly on standardized tests, it cannot reflect poorly on teacher performance. The teacher authority is based on the monopoly to evaluate pupil performance. Challenging that monopoly calls to question the very foundation of the school habitus.

Measuring pupil performance for evaluating teacher performance ruins the school habitus, because it devalues all relationships in favour of one measurable indicator. Making one aspect of school life visible renders the rest of them invisible, and therefore unimportant. But those aspects—the power relations between pupils and teachers—are the main engine for generating the academic achievements. By focusing on one particular set of ends, we undermine the means.

While actual gains in knowledge and skills may be eventually measurable, the social organization of schooling does not include a simple production of learning. There are two groups of labourers, pupils and teachers, who enter into complicated labour arrangements that exclude clear roles and obligations. On one hand, learning motivation is not clearly established. In effect, it largely depends on relational structures put in place. While teachers ostensibly act as managers of learning activities, and as knowledge experts, they are in fact provides of the affective infrastructure of communities. The relational structures they rely on to motivate pupils are within the realms of the honour economy. Therefore even if for an outsider, measuring the effectiveness of teacher's labour makes a lot of sense, the insiders all implicitly understand that that would violate the ethics of

immeasurability that govern the entire architecture of the school's social structure. Why cannot they explicitly articulate what I have just written?—because articulation is also against the rules of the habitus.

Now What?

Is Campbell's law really a law? Yes, it is—within the context of a habitus such as school. Systems based on complex human interactions, with relational foundations do require a degree of imprecision to function. This is why the taboo against measuring exists. It is not a whim, not a sign of backwardness or conservatism of their members; the ethics of immensurability is a perfectly pragmatic device. If we go back to the set of original Campbell's examples, we will find them situated in non-market settings, such as government offices, the army, the Soviet industry, schools, etc. Even if you doubt the usefulness of the taboo, it will do you good to at least take it seriously.

The school habitus *when it is damaged* by measurements produces the gaming behaviour. The hidden variable in Campbell's law is exactly that—the habitus under stress that allows previously negotiated ethical norms to collapse. If people are always corrupt to begin with, the *lack* of measurements and accountability in theory should produce much wider cheating and gaming behaviours. Yet we do not have any evidence that such a thing happened. It is not the case that all teachers behaved ethically, far from it. However, within the school habitus, measurements produce both the incentive to game the system and the weakening of the old honour code regulators. It is the violation of taboos against measuring that causes the gaming behaviours.

This does not mean there cannot be a well-functioning habitus without the taboo against measuring. The taboos against explicit measurement are inactive within the market economy proper. Why it is the case can be a subject of another paper. I will only point to the field of economic anthropology. Polanyi (1957) critiques Adam Smith for the assumption of man's "propensity to barter, truck and exchange one thing for another." "In retrospect—writes Polanyi—it can be said that no misreading of the past ever proved more prophetic of the future (43)." Polanyi argues that both historical and anthropological data show market economies to be an exception rather than the rule. For tens of thousands of years, men and women acted not in pursuit of their individual material interest, but on other considerations:

The outstanding discovery of recent historical and anthropological research is that man's economy, as a rule, is submerged in his social relationships. He does not act as to safeguard his individual interest in the possession of material goods; he acts so as to safeguard his social standing, his social claims, his social assets. He values material goods only in so far as they serve his end. Neither the process of production nor that of distribution is linked to specific economic interests attached to the possession of goods; but every single step in that process is geared to a number of social interests which eventually ensure that that the required step be taken (Polanyi 1957, 46).

The historical cause of Campbell law is an attempt to borrow the way of using measurement in market economies and apply them to non-market kinds of habitus. The transposition simply does not work, and we have little collective understanding why.

One practical solution for the Campbell's law is to erect an organizational wall between data and decision-making. In social sciences, researchers have been traditionally playing the role of intermediaries, interpreting data for other researchers and policymakers. Two conventions distinguish the practice. First, there is always a time delay between availability of the data and whatever decision may be influenced by findings. So it is difficult to use the results for personnel decisions. Second, it is the contentious nature of scholarly interpretation of data. The contention is internal, when a scholar tends to second-guess her own conclusion, and defend against possible criticism with numerous caveats. The most common caveat, distinguishing causality from correlation, is little understood by the majority of public, most journalists, and, alas, many policymakers. It is also external, coming from the peer review and scholarly debate processes. This convoluted and mediated use of measurements appears to be frustratingly slow and inefficient to policymakers, who have to rely on common sense and anecdotal evidence in addition to the always inconclusive findings of social scientists. Yet such a use often manages to avoid the pitfalls of Campbell's law. It does not always do that, and in education policy, we can see many decisions based on uncritical interpretation of research findings. As a system though, the scholarship does seem to work much better than the direct use of data.

The problem appears when we try to use the direct, unmediated data in decision-making; from spreadsheets to directives. The decisions appear to follow the objective data, hence the appeal of evidence-based, or even evidence-driven policymaking. However, as I mentioned, the decision what and how to measure actually precedes the measuring, and in effect, determines the decision that comes after the measure is being taken. As a result, the evidence-based decision-making is often less rigorous and more biased than the conventional comprehensive and intuitive way of making decisions. There is something poisonous about using measurements without a proper interpretation of data. And it looks like people directly involved in the social setting are incapable of careful interpretations. One needs an eye of a disinterested stranger to do that.

It appears we must restore the monopoly on measurements to researchers. I understand how selfish and self-serving this recommendation may look like, coming out of the mouth of a scholar. However, that is the conclusion brought by my investigation of Campbell's law. (I don't think we should deny our findings when they happen to benefit us; to disclose is all we owe to the reader). We must make sure no measurement data is used directly to make important social decisions. This may sound contrary to the current trends and dreams of "big data", and perhaps a tad anti-democratic. I fail to see any legal ways of doing that. However, a complex society must be sophisticated enough to understand the value of non-measuring habitus and protect its fragile relational ecosystem, until an alternative is found. The ethics of immensurability needs to be intentionally re-created not only within each school, but also within the broader educational systems.

Can we improve efficiency of non-market habitus without the direct use of data for evaluation purposes? Yes, we definitely can do that. The non-market habitus must be improved within its own rules and assumptions. In a seminal work, Wilson (1989) describes how certain government agencies are efficient, while others are not, even though they may operate under similar constraints. He attributes the differences to particular culture, and sometimes the history of different agencies, which is not much of an explanation. Regardless, his detailed descriptions of government agencies show how a non-market habitus operates, and how things can go right and wrong within it.

Can we also create a new school habitus not averse to measuring? I am quite confident we can. However, it is a sophisticated task that needs to involve reconsidering the entire

institution of schooling, including its authority structure. The data-driven decision-making cannot be simply superimposed on the existing school habitus without corrupting the latter.

References

- Bourdieu, P. 1998. *Practical reason: On the theory of action*. Palo Alto, CA: Stanford University Press.
- Campbell, D.T. 1976. *Assessing the impact of planned social change*. Hanover, NH: The Public Affairs Center, Dartmouth College. <http://goo.gl/AAaPx>.
- Hardt, M. 1999. Affective labour. *Boundary 2* 26(2):89–100.
- Koretz, D. 1998. *Measuring up: What educational testing really tells us*. Cambridge, MA: Harvard University Press.
- Nichols, S.L., and D.C. Berliner. 2007. *Collateral damage: How high-stakes testing corrupts America's schools*. Cambridge: Harvard Education Press.
- Polanyi, K. 1957. *The great transformation: The political and economic origins of our time*. Boston: Beacon Press.
- Premack, D., and A. Premack. 2003. *Original intelligence*. New York: McGraw-Hill.
- Rothstein, M. 2010. *What makes a great teacher?*. Providence, RI: Public forum by the Rhode Island Foundation and WRNI.
- This American Life*. 2010. *Right to remain silent*. <http://www.thisamericanlife.org/radio-archives/episode/414/right-to-remain-silent>.
- Wilson, J.Q. 1989. *Bureaucracy: What government agencies do and why they do it*. New York: Basic Books.