The 2024 Stocktaking and What It Measures

Indicators, Self-Reporting and the Blind Spots of a Reform Cycle

Laura Kivi & Iñigo Arrieta
Laura Kivi, University of Tartu
Iñigo Arrieta, University of Deusto

The Bologna process measures itself through a stocktaking exercise that reports on the implementation of commitments every three years, and the report has become the process's principal instrument of accountability. This article examines what the exercise measures, how the indicators are constructed, and which commitments cannot be captured by an indicator at all. It argues that the stocktaking is accurate about formal adoption and silent about practice, and that the gap is a property of the method rather than of the systems it assesses.

Contents **Page**

1. An Exercise and Its Function **9**
2. What the Indicators Measure **10**
3. The Commitments That Resist Measurement **10**
4. The Questionnaire as an Instrument **11**
5. The Comparative Publication **11**
6. The Alternative Instruments **12**
7. What the Exercise Should Not Attempt **12**
8. The Cycle and the Institution **13**
9. Conclusions **14**
9. The Political Economy of a Score **14**

1. An Exercise and Its Function

The stocktaking began as a device for reporting progress to ministers and it has become the process's memory: a cycle of questionnaires, national reports, expert scores and a comparative publication that appears before each ministerial conference.

Its function is triple. It informs the ministers who commissioned it, it allows a country to compare itself with its neighbours, and it creates a body of evidence that researchers use. The three functions require different things of the same instrument, and the tension among them explains most of the criticism that the exercise attracts.

An indicator reports adoption and cannot report practice

An instrument that informs ministers must be timely and simple; an instrument that supports comparison must be consistent across cycles; an instrument that supports research must be transparent about its method. A single exercise that attempts all three will be simple where a researcher needs nuance, and nuanced where a minister needs a number.

2. What the Indicators Measure

The indicators of the stocktaking are constructed from national reports that governments submit, cross-checked with independent sources and scored by a panel of experts.

The scoring is the element that the exercise's critics discuss most, and the more consequential element is the source. A national report is written by a ministry that has an interest in a favourable score, and the expert panel corrects for some of that interest but not for the structure of the question: a question that asks whether a system has a qualifications framework will be answered affirmatively by every system that has published one, whatever its use.

The consequence is a body of indicators that measure adoption with reasonable accuracy and use with almost none. A system can score fully on a framework that exists in law and is unknown to every employer in the country, and the score is not wrong: it reports what the question asked.

3. The Commitments That Resist Measurement

The commitments of the process that matter most are the ones that an indicator cannot capture, and the list is short enough to state.

The first is the social dimension, which concerns the participation of groups that higher education has historically excluded. A system can report a policy, an office and a strategy, and the policy's effect on participation depends on a hundred local decisions that no national report describes.

The second is academic freedom, which is a condition of the institution rather than a policy of the state, and which a self-report from a ministry is the least reliable instrument for measuring. The third is the quality of teaching, which the process addresses through a vocabulary

of learning outcomes and which no indicator distinguishes from the vocabulary's adoption.

The fourth is the value of a qualification in the labour market, which is the outcome that students care most about and which the exercise approaches through the proxy of a graduate survey that few systems conduct consistently.

4. The Questionnaire as an Instrument

The questionnaire is the point at which the process reaches a ministry, and its structure determines what a system can say.

A closed question with three options produces a score and destroys the information about the cases that do not fit. An open question produces a paragraph that an expert must read and a score that two experts may disagree about. The exercise uses both, and the balance between them is the design decision that its authors revisit every cycle.

The redesign that would improve the instrument is not a longer questionnaire. It is a set of questions that ask for evidence rather than for a description: the number of qualifications recognised automatically in the last year, the number of appeals against a recognition decision and their outcome, the proportion of a cohort that received a maintenance grant. A number that a ministry must produce from an administrative system is harder to overstate than a paragraph that it composes.

Commitment	Indicator	Blind spot
Qualifications framework	Adoption	Use
Recognition	Legal basis	Practice
Social dimension	Policy	Participation
Academic freedom	Statement	Condition
Teaching quality	Vocabulary	Classroom

Table 1 What the exercise measures and what it cannot

5. The Comparative Publication

The comparative report is the instrument's most visible output and the one that a ministry reads.

Its tables rank systems by score, and the ranking is the element that produces the political effect. A ministry whose system scores below a neighbour's has a reason to legislate, and the legislation that follows is frequently a formal adoption of exactly what the indicator asked for, which is the mechanism by which an exercise designed to observe a reform becomes a cause of one.

The mechanism is not a defect and it requires care. A process that measures the publication of a framework will produce frameworks, and a process that wishes to produce their use must measure use, which is harder and which the exercise has attempted in its later cycles through indicators of practical implementation.

6. The Alternative Instruments

Three instruments would measure practice with more accuracy than a national self-report, and each of them is available to the process without a new legal basis.

The first is a register: a database of recognised qualifications, populated by the systems, from which the exercise could compute the share of a cohort that is covered by an automatic recognition arrangement. The second is an administrative statistic: the number of recognition decisions, their duration and their outcomes, collected from institutions rather than from ministries.

The third is a student-level survey conducted on a common instrument, which the process has piloted and which a minority of systems administer. The three instruments are more expensive than a questionnaire and they produce numbers that a ministry cannot compose, which is precisely their value.

7. What the Exercise Should Not Attempt

The temptation of every reporting cycle is to add a dimension, and the addition is what has made the exercise long without making it informative.

A dimension added to satisfy a constituency produces a question that the constituency can answer and that no one else uses. The process has accumulated dimensions of this kind, and each of them costs a ministry's time and produces a paragraph in a report that a minister does not read.

The discipline that the exercise needs is a retirement rule: for every new indicator, one is removed, and the removal is announced with the addition. The rule is unpopular with everyone who has built a dimension and it is the only mechanism that keeps a comparative instrument comparable across cycles.

A fourth instrument is the site visit, and the exercise has never used one. A panel that spends two days in an institution, reads three recognition decisions and interviews the officer who took them would learn more about the practice of a commitment than a national report can convey, and the cost of such a visit is the reason it has not been attempted. The process has preferred breadth to depth, and breadth is appropriate to an exercise that must cover forty systems in a single cycle.

8. The Cycle and the Institution

A stocktaking cycle is a national event and an institutional one, and the two are connected by a delay that the process has never measured.

A ministry that prepares a report in one year describes an arrangement that was made three years earlier, and an institution that changes its practice in the same year will be described in the next cycle. The delay is a property of any reporting instrument and it is shorter in systems that collect administrative data continuously and longer in systems that assemble a report from questionnaires sent to institutions. The difference is not a matter of diligence; it is a matter of whether the system has a register or a correspondence.

An institution can reduce the delay on its own side by maintaining the data that a report will require, rather than assembling them when a request arrives. A registrar who records every recognition decision with its date, its outcome and its duration has produced the evidence for a national indicator without knowing which indicator will be asked for, and she has also produced the evidence for her own management. The practice is the cheapest form of readiness and it is the one that institutions discover last.

The second connection between the cycle and the institution is the vocabulary. A ministry that must answer a question about student-centred learning will ask its institutions to demonstrate it, and the demonstration will be produced in the format that the questionnaire implies. The process therefore shapes institutional documentation twice: once through the national report and once through the request that the report generates. An institution that understands the mechan-

ism can prepare a single document that serves both, and one that does not will produce two documents that serve neither.

9. Conclusions

The stocktaking measures adoption with accuracy and practice with almost none, and the gap is a property of the method: a national self-report about a formal commitment cannot report what happens in an admissions office.

Three conclusions follow. The exercise should ask for administrative numbers rather than for descriptions, since a number drawn from a register is harder to overstate than a paragraph. It should compute its indicators of practice from registers and surveys rather than from ministry reports, and it should accept the cost. And it should adopt a retirement rule for its dimensions, because an instrument that grows every cycle becomes a document that nobody compares.



Figure 1 Instruments of measurement

9. The Political Economy of a Score

A score is a political object, and its effect on a system depends on who reads it and what they can do with it.

A ministry that receives a low score has three possible responses: to change the practice, to change the report or to dispute the method. The

third response is the most common in the cycles that follow a methodological change, and it is not always unreasonable, since an indicator that has been redefined produces a comparison that the system did not consent to. The exercise's authors have learned to announce a methodological change a cycle in advance, which is the minimum courtesy that a comparative instrument owes the systems it compares.

The second response is the one that damages the instrument most, because a report that has been written for a score is a report that a panel must discount, and the discounting is invisible in the published result. The exercise's defence against it is the cross-check with independent sources, and the cross-check is strongest where an administrative register exists and weakest where the commitment concerns a practice rather than a document.

A comparative instrument is only as good as the systems' incentive to be described accurately

References

[1] Kivi, L., & Arrieta, I. (2022). Measuring the Bologna process. Tartu: University of Tartu.

[2] European Commission (2020). The European higher education area in 2020: Bologna process implementation report. Brussels.

[3] Eurydice (2022). National student fee and support systems. Brussels.

[4] Gornitzka, Å. (2018). The Bologna process as a governance architecture. In European integration and the governance of higher education. Dordrecht: Springer.

[5] Ravinet, P. (2008). From voluntary participation to monitored coordination. European Journal of Education, 43(3), 353–367.

Biography:

Laura Kivi works on higher education policy and on the measurement of reform at the University of Tartu.

Iñigo Arrieta is a researcher at the University of Deusto working on comparative indicators and on governance of the European Higher Education Area.