ETF Working together Learning for life

European Training Foundation

THE EU AGENCY SUPPORTING COUNTRIES TO DEVELOP THROUGH LEARNING

DATA MISINTERPRETATION – A PROBLEM OR A TOOL?

An ETF article based on proceedings of the ETF monitoring forum 2025

This report has been prepared by the European Training Foundation.

Author(s): Mihaylo Milovanovitch (ETF).

When citing this report, please use the following wording:

European Training Foundation (2026), *Data misinterpretation – a problem or a tool?,* Turin: ETF.

The views reflected in this document do not necessarily reflect the corporate position of the ETF.

© European Training Foundation, 2026

Except otherwise noted, the reuse of this document is authorised under the Creative Commons Attribution 4.0 international (CC BY 4.0) licence (https://creativecommons.org/licenses/by/4.0/). This means that reuse is allowed provided appropriate credit is given and any changes are indicated. For any use or reproduction of photos or other material that is not owned by the European Training Foundation, permission must be sought directly from the copyright holders.

PREFACE

Mihaylo Milovanovitch (ETF), Stefano Lasagni (ETF), Marco Barreca (ETF), Abdelaziz Jaouani (ETF), Bujar Hajrizi (Acting Head of Division, Division of Social Statistics, Kosovo Agency of Statistics (KAS), Kosovo), Diana Hammoud (Central Focal Point and Coordinator, General Directorate of Vocational and Technical Education, Lebanon), Dr. Amr Abdelrahman Ibrahim Bosila (Head of the Central Administration for the Development of Technical Education, Ministry of Education and Technical Education, Egypt), Elona Sevrani (Head of Labour Statistics Sector, INSTAT, Albania), Güneş Inan Vicil (Expert, Turkish Statistical Institute TURKSTAT, Türkiye), Irma Gvilava (Head of Labour Statistics Division, National Statistics Office of Georgia, Georgia), Ivana Simic (Senior Advisor, Labour Market Department, Agency for Statistics of Bosnia and Herzegovina, Bosnia and Herzegovina), Ivana Tanjevic (Head of Labour Market Statistics Department, Statistical Office of Montenegro MONSTAT, Montenegro), Julka Ostojic (Head of the Division for Secondary General and VET, Ministry of Education, Science and Innovation, Montenegro), Lusine Kalantaryan (Head of Division, ARMSTAT, Armenia), Mayssa Daher (Responsible for Social Statistics, Central Administration for Statistics, Lebanon), Meryem El Habti (Statistician–Economist, Division of General Statistics, High Commission for Planning HCP, Morocco), Müberra Öztürk (Head of Department, Ministry of National Education, Türkiye), Pero Ilievski (Senior Associate, Labour and Living Standards Sector, State Statistical Office of the Republic of North Macedonia, North Macedonia), Said Soliman (Statistician, CAPMAS, Egypt), Valbona Fetiu Mjeku (Head of Division for VET Standards and Quality Assurance, Ministry of Education, Science, Technology and Innovation, Kosovo).

This article examines the interpretive stage in the use of education data in policy processes, where statistical results are first translated into claims about problems, progress, and policy priorities. It argues that the misinterpretation of data is common at this stage, and an ordinary feature of professional practice as actors routinely select, frame, and emphasise certain aspects of evidence when communicating results. Drawing on reflections from national stakeholders participating in the ETF Monitoring Forum "Evidence in Action", held in October 2025, the article explores how misinterpretation arises in everyday work, the forms it takes, and the reasons behind it. It also argues that data misinterpretation can be a legitimate and useful professional tool, and asks where its responsible use ends and malpractice begins.

Note on authorship

The authority of this article is grounded in the collective experience of practitioners from different countries who participated in the working session on data misinterpretation (Session B) at the ETF Monitoring Forum 2025, Evidence in Action. All listed contributors provided examples, reflections, and professional validation that shaped the interpretations presented here. The article therefore adopts a collective authorship approach.

The topic and overall conceptual framing were developed by the first author, in collaboration with the second and third authors. The remaining co-authors are listed in alphabetical order.

The manuscript was drafted and structured by the first author.

CONTENTS

1. INTRODUCTION 5

2. DATA INTERPRETATION AND MISINTERPRETATION 7
2.1 Conceptual framing 7
2.2 Why focus on data interpretation? 8

3. A SENSE OF PREVALENCE: EXPERIENCES FROM PROFESSIONAL PRACTICE 9

4. TECHNIQUES: HOW DOES MISINTERPRETATION HAPPEN? 10
4.1 Conceptual techniques of misinterpretation 10
4.2 Technical methods 11

5. REASONS: WHY DOES MISINTERPRETATION HAPPEN? 14
5.1 Strategic communication 14
5.2 Institutional and donor pressure 14
5.3 Policy and political pressure 15
5.4 Personal and professional self-interest 15
5.5 Making data more usable 15
5.6 Low statistical literacy and insufficient training 16
5.7 Loss of metadata 16
5.8 Weak verification culture 17

6. GUARDRAILS: IDEAS FOR RESPONSIBLE PRACTICE 18

7. CONCLUSION 20

ACRONYMS 21

REFERENCES 22

1. INTRODUCTION

Education systems are built on shared commitments: that everyone should have the opportunity to learn, that teaching should meet clear standards of quality, and that discrimination has no place in the system (Milovanovitch, 2013, 2019). These commitments guide professional practice and frame the expectations of learners and other stakeholders. They also set many of the objectives pursued through education policy (Milovanovitch, 2019; OECD, 2018). Multilateral frameworks such as the UN Sustainable Development Goals (SDGs) draw on these principles, and national policies and strategies are developed with reference to them (ETF, 2022).

The extensive datafication of education, which is understood as the transformation into quantifiable data of as many aspects of the educational environment as possible (Romero & Ventura, 2012), is increasingly seen as a principal strategy for achieving such aims (Williamson, 2017; Jarke & Breiter, 2019). Recent analysis of the content of more than 9,000 education reforms from 215 countries up to the year 2018 using the World Education Reform Database shows a clear rise in reforms related to data and information (Bromley, Nachtigal, & Kijima, 2024). The significance of datafication is such that the absence of data has become almost synonymous with limited progress, a symptom of a "data crisis" that contributes to learning poverty and impedes reform and improvement (Gorur, Dey, Faul, & Piattoeva, 2023).

Education data, however, belongs to one of the more vulnerable data ecosystems in the public sector (UNESCO Institute for Statistics, 2025). Some of the vulnerabilities are technical and impact the ways in which education the data is produced. The coverage of education indicators is uneven between and even within countries, and there are methodological inconsistencies and disagreements over definitions. Often, countries have multiple competing data sources and grapple with gaps in statistical capacity and availability. Coordination and standardisation across countries are also less common than in domains such as labour statistics (UNESCO Institute for Statistics, 2025).

Another set of vulnerabilities is political. Education indicators are not neutral. They are embedded in policy ideas and governance dynamics and often carry political implications (UNESCO Institute for Statistics, 2025). Therefore, education statistics can become politically powerful instruments that frame policy problems, shape reform agendas, and legitimise certain reforms (Verger, Fontdevila, & Zancajo, 2016). On the downside, this type of exposure can also bring tensions, disagreements and risks. Some researchers note that in the field of international education comparisons, for example, policy borrowing based on international indicators can be misleading as data is often used normatively rather than analytically (Carnoy, 2016).

There is an ongoing discussion regarding remedies and strengthening education data in response to such concerns. Much of the literature is about improving the technical foundations of education statistics and the systems through which data are generated and validated. This includes questions of indicator coverage, methodological soundness, comparability, and institutional arrangements for data collection and reporting.[1]

There is also an emerging body of work examining the political sensitivity of education data and the risks associated with its use in policy processes. Because this line of research is relatively recent, its availability and scope is more limited. So far, it has focused mainly on themes such as international benchmarking, the role of data in global education governance, and to some extent on policy analysis based on data (Carnoy, 2016; Verger, Fontdevila, & Zancajo, 2016; Council of Europe, 2025).

This article seeks to contribute to the debate about the better use of data in education by bridging the technical literature, which asks whether education data is methodologically sound and statistically reliable, and the broader political economy literature that addresses how education indicators become embedded in power relations, policy agendas, and processes of global governance. Our work focuses

[1] See for instance the UNESCO–World Bank Data Quality Assessment Framework for Education Statistics, World Bank & UNESCO Institute for Statistics, 2003).

on the interpretive stage between these perspectives, where the data is first read, selected, framed, and communicated, and numerical observations begin to be translated into a first round of claims about possible problems and policy priorities.

This is the moment the evidence starts to acquire meaning and begins to influence how policy problems may be formulated and understood. It is at this stage that narratives around the data begin to form, and divergent readings of the same evidence may become visible for the first time as different actors assign different meanings to the same data. Those divergences may deepen at later stages when translating data into insights, when the causes of results are analysed and policy arguments developed.

This interpretive stage is also the earliest point at which meaning can be handled wilfully and responsibly, before it hardens into analysis, recommendation, or policy justification. Although easily overlooked, it is a consequential stage. It is the point at which actors can still become aware of the choices they are making when they move from numbers to claims. That includes awareness of what is being selected, emphasised or downplayed, compared, or left unexplained. It is also the point at which the possible policy and real-life implications of a reading of the data can still be anticipated. Once an interpretation is repeated and taken forward into analysis, policy papers, or public debate, it becomes harder to correct.

Somewhat paradoxically, a degree of deliberate misinterpretation at this stage, if careful, bounded, and transparent, may sometimes help prevent cruder misuse later on and protect against more harmful distortions that arise when the data is treated as if it speaks for itself or is mobilised carelessly in policy arguments.

In this article we unpack this theme with the help of a guided discussion among national stakeholders from countries in Eastern and Southeast Europe, Central Asia, the South Caucasus, North Africa and the Middle East about three guiding questions:

- How does interpretation shape the meaning of data and how common is data misinterpretation?
- How and why does data misinterpretation occur in professional practice?
- Can data misinterpretation serve legitimate purposes and, if so, where should the guardrails lie?

The evidence and reflections we present in this article originate from a working session on data misinterpretation that explored these questions during the ETF Monitoring Forum “Evidence in Action” in Milan held in October 2025. The session brought together members of the ETF KIESE data network as well as national coordinators for the Torino Process from ETF partner countries in East and Southeast Europe, Central Asia, Northern Africa and the Middle East. The session was organised as a facilitated working group and combined short inputs, case-based exercises, and collective reflection on the experiences of participants with interpreting and misinterpreting education data. The perspectives emerging from that exchange form the backbone of the analysis developed in this article.

The working session generated evidence in two different ways. Some findings, for example the perceived prevalence of data misinterpretation in Chapter 3, emerged through relatively open soliciting of professional experiences by participants. Others, for example the discussion regarding techniques for data misinterpretation in Chapter 4, were produced through a more structured process in which moderators presented a provisional typology and participants reacted to it by confirming, refining, and illustrating it with examples.

The choice between more open or more structured approaches was guided by practical considerations of the workshop design, including the need to use the available time efficiently while still allowing for meaningful validation and enrichment of the analytical framework. In any case, the article treats the evidence from both approaches as equally valid.

2. DATA INTERPRETATION AND MISINTERPRETATION

2.1 Conceptual framing

The production of statistics is a sequence of activities rather than a single act. Widely used models of processes in official statistics and data analytics separate what is done to obtain the data, what the data contains, what audiences should understand from the data, and what factors might explain the results.

One example is the Generic Statistical Business Process Model (GSBPM), a framework that is widely used in official statistics. The model separates the Collect phase, which concerns capturing or obtaining data, from the Process phase, where data is edited, coded, and transformed into statistical aggregates. It then includes an Analyse phase, which covers the preparation of outputs, validation activities, and the interpretation and explanation of statistical results before publication (GSBPM, 2025). A similar logic appears in the CRISP-DM lifecycle used in data analytics. In that model, activities such as initial data collection and data description are placed in the phase of understanding the data, which precedes modelling and evaluation and treats the description of the dataset as a task which is separate from modelling relationships or the drawing of explanatory conclusions (Shearer, 2000).

Professional standards follow the same logic. Codes of statistical practice define statistical work to include the design of data collection, the summarisation and processing of data, its analysis, interpretation, and communication (Amstat, 2022). Each activity involves its own professional responsibilities and quality requirements.

These pragmatic frameworks suggest that meaning does not appear at once but is constructed step by step. The separation exists because each stage produces different outputs and involves different methodological tasks and documentation requirements. In this article, we extend this insight beyond the statistical treatment of data during its production and examine how data is translated into meaning within research and policy contexts.

This wider focus calls for a broader conceptual framework that can capture how the data is connected to real-world questions, policy concerns, and explanations. To describe this process, the article delineates three conceptual stages through which the data acquires meaning. These are the description of the data, the interpretation of its significance for a question or audience, and the analysis of the mechanisms that may explain the results. Together, these form what we call the *data-to-meaning continuum*. In the reality of working with data, these steps are likely to be iterative and overlap. Nevertheless, they can be distinguished conceptually as follows:

- The first stage in the continuum—**data collection and description**—produces observations and metadata that describe what is measured, how, when, and under which rules, together with neutral summaries such as counts, rates, distributions, and tables, accompanied by information on denominators, time windows, and breaks in series.
- The next stage—**data interpretation**—selects and emphasises (or withholds) elements of the data in ways that shape how a specific question is answered or what a specific audience sees in the data.
- Finally— **data analysis**—seeks to explain the results observed through the data by exploring the mechanisms, conditions, or factors that may lie behind them. This includes possible policy-related explanations.

This separation of steps also clarifies what the stages in the data-to-meaning continuum do not do. Data collection and description do not attempt to explain results; they report what the dataset contains. Data interpretation communicates meaning and relevance for a question or audience, but it does not

seek the real-world causes behind the results. That is the task of analysis, which in this context is understood as a step that must go beyond the communication of meaning, to consider what may explain the results.

2.2 Why focus on data interpretation?

This article is centred on the interpretive stage in the movement from data to meaning, that is data interpretation. This is the stage at which meaning is first extracted from the data through techniques such as selection, emphasis, and comparative framing, moving beyond raw description.

Because interpretation is a stage of choice, it also contains the possibility for divergence. As the stage at which meaning can first become socially and politically consequential, it is a high-risk boundary for misuse, where actors may select, highlight, downplay, compare, order, or leave unexplained different elements of the same data, often without attracting notice. Prior research shows that converting information into actionable instructional knowledge is demanding and easily derailed by misconceptions (Gummer & Mandinach, 2015). Choices in how statistics are framed can also create misleading impressions even when the descriptive numbers are correct (GAF, 2018).

If interpretation carries an inherent risk of misinterpretation, it becomes necessary to consider not only how it works, but also how it can go wrong. In this context, data misinterpretation refers to the *selective, incomplete, or skewed reading, framing, or presentation of data that can produce a false, distorted, incomplete, or misleading understanding of what the data shows*.

At the same time, the present discussion does not treat misinterpretation only as a failure to be avoided. The reflections in this article proceed from the assumption that, in real-world settings, intentional data misinterpretation is common and may at times be necessary. What matters, then, is whether a boundary can be drawn between more and less acceptable forms of misinterpretation, and what guardrails may help distinguish between them.

3. A SENSE OF PREVALENCE: EXPERIENCES FROM PROFESSIONAL PRACTICE

If the intentional misinterpretation of data is indeed common, how widespread is it in practice?

After reaching consensus on the conceptual boundaries of data misinterpretation, the working session moved to a facilitated sharing of professional experiences related to the phenomenon. The session focused on the extent to which participants recognised misinterpretation as a regular feature of their professional work, both in national contexts and in international reporting. The exchanges were captured in a transcript, which was subsequently denoised, anonymised, and analysed to extract factual findings.

The discussion brought up strong qualitative evidence that data misinterpretation is a routine, cross-sectoral, real-world phenomenon in multiple countries and not just a conceptual possibility. Examples from the own experience of participants in the room surfaced very quickly. They noted that misinterpretation occurs both nationally and among international organisations with whom they cooperate, that it can "happen everywhere", and that the topic is "not exotic" but close to their "everyday work" as data professionals and civil servants.

The tone of exchanges was one of the mundane. Data misinterpretation was discussed as though it was embedded in everyday normal workflows, such as reporting, media uptake, policy design, policy evaluation, and routine review processes where data is often taken on trust. As one participant observed, "in many review processes, the data is taken on trust, because triangulation and verification are time-consuming". This creates space for selective interpretation and other forms of misinterpretation to pass largely unnoticed.

Another feature of data misinterpretation is that it is not bound to a single sector. Although the workshop began with examples from education, the discussion quickly moved onto labour-market statistics, gender statistics, macroeconomic indicators, data on migration and the brain-drain, and even infrastructure reporting.

The spread of cases was telling. It included unemployment compared between countries without accounting for inactivity ("the comparison is misleading… often from ignorance, not deceit"); a gender-pay-gap shift driven by the arrival of low-paid male refugees ("it did not mean women suddenly earned more; the composition of the labour force changed"); employment and unemployment figures being misread because methodological changes were not carried through into secondary reporting ("journalists reported a 'sharp rise'… but in reality it was just a methodological change"); CPI being smoothed to make inflation look lower; brain-drain rates inflated by very small denominators; and an incident report omitting a cause in order to preserve donor financing for urgent repairs.

4. TECHNIQUES: HOW DOES MISINTERPRETATION HAPPEN?

It is noteworthy that among the wide range of examples of data misinterpretation shared by participants in the working session, none involved an outright fabrication of data. It seems that the practice most often involves the steering of meaning through interpretive and presentational choices around the framing, filtering, comparison, and contextualisation of data, without the falsification or manipulation of the raw evidence. As one participant observed, “the numbers themselves are rarely wrong, but… can tell very different stories”, depending on how they are presented, while another stressed that at this stage in the process of giving meaning to data, “you do not falsify, but you choose what to highlight” to guide audiences towards a preferred reading of the data. The same, technically sound dataset can therefore support several different, even opposing, statements, all of them technically true.

The discussion identified a range of techniques through which this happens, from framing, selective emphasis, and storytelling to graphical distortion, selective time windows, shifting denominators, omission of uncertainty, and the removal of contextual or methodological information. The participants agreed that these techniques can be organised into two broad groups: conceptual techniques and technical methods.

Conceptual techniques operate at the level of interpretation. They shape the angle from which the data is understood and answer the question: what does this mean, and how should it be seen? They can be described as *interpretive ways of organising and narrating correct data so that audiences are led towards a preferred meaning*.

Technical methods, on the other hand, operate at the level of presentation and statistical handling. They help shape the form in which the data is shown or summarised in view of influencing how they are read. This group of methods can be defined as *presentational or statistical devices that use the choice of scale, metric, comparison, or aggregation to make correct data appear more dramatic, more certain, more representative, or more causally meaningful than the evidence itself warrants*.

Below is a more detailed overview.

4.1 Conceptual techniques of misinterpretation

Strategic framing

Strategic framing is the deliberate organisation of facts so that audiences see an issue in a preferred way. The mechanism is simple and includes the choice of what to emphasise, what language to use, and what to leave out, without necessarily saying anything false.

The working session used the example of an imaginary project for girls in support of their participation in education in an imaginary country (“Catlandia”). If the aim is to suggest that the project is failing in order to direct resources elsewhere, one can focus on the stagnant female share visible in the data and ignore the rise in the absolute number of girls enrolled, visible in the same dataset. In that sense, the selective emphasis directs the attention towards one aspect of the evidence and away from another, making one reading of the data more likely than competing readings. This matters because many readers do not move beyond the first impression, and repeated framing can quickly harden into an accepted “truth”.

The same logic also appears in more practical forms of selective omission. During the discussion, examples were given of showing only one column of a table and presenting the rest as outside the scope of the report, or of formatting a table so that inconvenient rows disappeared when printed. Common to all these examples of strategic selectiveness is that the data remains the same, but the visibility is managed in ways that guide the reader towards a preferred conclusion.

Data storytelling

Data storytelling links selected data points into an intuitive story that carries a message. Misinterpretation here arises not from a single false figure, but from the way indicators are combined and sequenced so that they produce a coherent narrative effect.

One example provided during the session paired increased spending with a deterioration in the material base of schools in a country. Both indicators may be correct yet placing them side by side encourages the audience to infer that something is going wrong with the money invested in education. The force of the technique lies in the coherence and tempting simplicity of the story, which is often more persuasive and memorable than the fuller, more complex reality from which it is drawn.

This is why storytelling can mislead even when the underlying data are accurate. By selecting only those indicators that fit the intended message, it suppresses evidence that may weaken that message and presents a more unified and intuitive interpretation than the dataset as a whole would justify. Once images or other evocative elements are added to a presentation of the story, the narrative becomes even more immediate and strong.

Glass half-empty/half-full framing

Discussions called this conceptual technique “subtle” because it works through valuation rather than distortions. The same fact can be cast negatively or positively depending on the purpose of the speaker and the wider context.

The example used in the working session was of a country in difficult conditions spending around 2.8 to 3 per cent of GDP on education over several years. This fact that can be described either as a proof of under-spending relative to international benchmarks, or as maintaining stable levels of investment in education despite adverse economic circumstances. Both statements may be accurate, but the “spin” they give to the facts is obviously very different.

What makes this technique difficult to judge is precisely that it does not disturb or otherwise misrepresent the facts. The issue is the interpretive choice to cast the same reality in a pessimistic or optimistic light. It can become misleading when one valuation suppresses relevant context and/or implies a conclusion that is stronger than the evidence can reasonably bear.

4.2 Technical methods

Truncated or non-zero Y-axis on bar/line charts

Among the technical methods, truncated or non-zero y-axes were among the most familiar to participants in the working session. These methods are also visually the most effective. The same data can be shown twice, but if one version starts the y-axis at 15 rather than 0, the visual distance between values appears much larger and the change (a trend, for instance) looks far more dramatic than it is. Nothing in the numerical information has changed. What changes is the visual impression.

The method is persuasive because many readers respond first to the shape of a graph and only later, if at all, to its scale. A graph can therefore exaggerate differences through presentation alone. This can make even a modest variation appear substantial.

Withholding context when presenting percentages or absolute numbers

Different representations of the same phenomenon can support different messages. One may choose percentages or absolute numbers depending on which version serves the intended argument better. Turning again to the fictitious country of Catlandia, in a table on the participation of girls in education, the absolute number of girls enrolled in VET in Catlandia increases substantially over time, while their share of total enrolment remains largely unchanged. Both statements are correct, but each is supported by a different reading of the same data. The issue therefore lies in the interpretive force of selecting one representation over another, rather than in the formatting alone.

Further discussion sharpened this point by drawing attention to denominator effects and changes in population composition. A percentage may look impressive while referring to a very small base population, and a stable share may conceal substantial growth in actual numbers. In such cases, the same figure can appear either large or negligible depending on whether the reader sees the proportion alone or also the number of cases behind it.

Participants also noted that shifts in the composition of the underlying population can distort the interpretation of otherwise straightforward indicators. For example, a large influx of male refugees may change the ratio between boys and girls in the relevant age group. If one then looks only at share of girls in enrolment, the result may suggest deterioration even when the actual number of enrolled girls has not fallen or has even increased. A similar issue can arise in labour-market statistics. An apparent narrowing of the gender pay gap may not result from women earning more, but from changes in who is working, for instance if lower-paid men leave employment in greater numbers or if the composition of employed women changes. In all such cases, the statistics are relational: without a clear view of the reference population and how it has changed, its meaning can easily be overstated or misunderstood.

Cherry-picking time windows

A trend can look sharper, more favourable, or more alarming depending on the time window chosen to display it. In an unemployment example discussed during the session, a four-year series suggests dramatic improvement, while a longer view shows little more than a return to pre-Covid levels. The same data can therefore support very different interpretations depending on where the temporal frame begins and ends.

Further discussion showed that this problem extends beyond simply cropping a graph. It also includes failures of comparability over time. In two of the countries present, for example, changes in employment and unemployment levels were reported without adequately signalling breaks in the time series caused by methodological changes. This was done on the insistence of international partners. While in one case, the original data clearly noted the break, a secondary publication based on that data omitted that information; in another, figures produced under old and new standards continued to be mixed.

The resulting interpretation was misleading not because the numbers were invented, but because dissimilar periods were made to appear comparable. Here, omitted metadata and methodological breaks are best understood as selective reporting, because once the explanatory context is removed, a technically correct figure can become false in substance.

Small sample sizes and missing uncertainty

When the evidence is weak, one way of making it look solid is to leave out information that shows how precise and reliable the result actually is, such as the sample size, confidence intervals, or standard errors. A survey result such as 75 per cent versus 78 per cent may sound precise and persuasive, but without knowing how many people were surveyed and how uncertain the estimates are, the difference may be meaningless. Skipping information of this kind is not trivial, because it determines how much interpretive weight a result can bear.

This method works because percentages and point estimates are often received as self-explanatory facts, while the fragility of the estimate remains hidden. Without this “hidden” information, it is easier to overstate confidence and invite conclusions that may not be supported by the data.

Using averages without describing distribution

A single summary measure can hide substantial differences within a population. The average income is a straightforward example. A mean income of €10,000 may sound representative, but if most people earn much less and only a few earn much more, the mean does not show what is typical for the majority. Presented on its own, the average can therefore make a highly unequal situation appear more even than it actually is.

As participants in the working session noted during the discussion, the statistic itself is not wrong. The problem arises when it is presented without information about how values are distributed across the population. In such cases, a legitimate summary measure becomes misleading because one central value is made to stand in for a much more uneven reality.

Aggregation hiding subgroup effects (Simpson's paradox)

Aggregation can reverse the apparent meaning of a relationship. The example used in the session concerned school spending on digital technologies and test results. In the aggregate, schools that spent more appeared to perform worse. Once the data was disaggregated by municipality or region, however, the relationship reversed: within each unit, higher investment was associated with better outcomes.

The contradiction arises because aggregation hides the context, including possible confounding factors such as remedial spending by already underperforming municipalities. Aggregate figures often appear more authoritative and more digestible, yet they can conceal the relationships that matter most. In this sense, aggregation not only simplifies, but inverts the story altogether.

Correlation used as causation (spurious correlations)

This is one of the clearest and most consequential technical methods of misinterpretation. The session used the example that more job applications correlate with higher unemployment. A causal reading would absurdly suggest that applying for jobs causes unemployment, when in fact the causal direction is more likely the reverse. A third factor may also be driving both phenomena, or the correlation may be spurious altogether.

To make the point memorable, the discussion also drew on deliberately outlandish examples of spurious correlations, such as the association between ice-cream consumption and OECD PISA results (more ice-cream, better results). The underlying issue, however, is serious. Participants explicitly noted that confusing correlations with causation can lead to dangerous policy decisions. What begins as a mistaken statistical inference can thus become a false explanation and, from there, a basis for harmful action with long-term consequences.

5. REASONS: WHY DOES MISINTERPRETATION HAPPEN?

If data misinterpretation is fairly ordinary in practice, as previously before, does this mean that all the practices described above should be read in the same way, collapsed into a simplistic contrast between honest and dishonest conduct?

The techniques described above are not inherently good or bad. Their ethical status depends on why they are used, how they are disclosed, and how faithfully they remain tied to the evidence. A broad body of prior research on integrity of policy and practice in education suggests that the reasons people engage in problematic conduct in this sector are diverse and rarely fit into clear-cut categories such as good or bad (OECD, 2018; Milovanovitch, 2025). In the context of the working session, this supported the assumption that different forms of data misinterpretation may arise from different motives and professional situations.

The working session pointed to a broad range of possibilities in that respect, which differed in intent, context, and likely consequences. The examples suggest that the phenomenon occupies an ethically uneven terrain. Some practices were self-serving, while many others arose from routine professional pressure or from attempts to make data usable without abandoning responsibility for the interpretation.

With that diversity in mind, why does data misinterpretation happen?

5.1 Strategic communication

The discussion repeatedly returned to strategic communication as one of the clearest reasons for data misinterpretation in various institutional and country contexts. Participants described situations in which the numbers themselves were not altered, but the angle was chosen so that the message would “land”. Framing, storytelling, and the choice between a “half-full” and a “half-empty” readings appeared again and again as ordinary features of communicative practice. The same applied to graphs linking education spending and learning outcomes, where identical data could support either the claim that more spending improved results or the claim that spending alone guaranteed nothing.

The participants were explicit that the choice of indicator, denominator, time window, wording, and sequence of presentation is an especially convenient and effective way of addressing strategic communication needs, because most audiences absorb the first impression and rarely reconstruct the full context. More often than not, the wish to be clear, memorable, and persuasive took precedence over the commitment to be complete.

5.2 Institutional and donor pressure

The discussion also linked data misinterpretation to the need to secure support, justifying interventions, fitting pre-set agendas, and keeping policy processes active. In such contexts, professionals were often not choosing between truth and falsehood, but between several technically valid readings of the same evidence, one of which was more useful than the others for securing resources or preserving momentum.

One of the examples offered in the discussion made this tension very clear. A report about infrastructural damage in a school omitted the unintentional cause of the damage (electrical wiring) because including it would have jeopardised donor financing for urgently needed repairs. Leaving the cause out allowed the project to move forward. In the discussion, this was treated as a higher-purpose justification rather than as simple deceit. That did not make it unproblematic, but it showed how organisational objectives could make data misinterpretation appear reasonable, acceptable as the

“lesser evil”, even desirable. Misinterpretation, in that sense, emerged from practical incentives built into institutional life.

5.3 Policy and political pressure

The transcript also showed how easily the interpretation of data becomes tied to policy and politics. The participants spoke about strategies that had already been written, ministerial decisions that had already been taken, and situations in which experts were expected to select indicators that matched an existing direction. Selective (mis)interpretation was therefore described as a part of ordinary policy-related work, in which evidence was used to support pre-existing decisions, rather than to shape them.

Some participants stressed that such selectiveness can be a positive approach, for example when evaluation findings are framed in ways that encourage improvement or participation. Yet the discussion also showed how narrow the line is between legitimate use of data for policy purposes, and politicisation. Once evidence enters a political setting, the same selectivity could serve re-election prospects, the public image of decision-makers, the deflection of responsibility for negative results, or the need to justify pre-existing administrative decisions.

One scenario discussed in the session concerned consumer price data. In some countries, salaries were linked to the rate of inflation. Reporting higher inflation would therefore have required salary increases, which in turn would have had direct budgetary and political consequences. Participants described cases in which inflation data was therefore “smoothed” to make the rate appear lower in order to ease the pressure for higher wages.

That example captures the broader point well. Data is often interpreted inside pre-existing agendas, and once that happens, the line between providing policy guidance and serving political interests may become difficult to hold.

5.4 Personal and professional self-interest

Personal motives appeared in the discussion, but less often than institutional or political ones. The participants referred to the desire to impress clients, strengthen one’s reputation, or justify higher remuneration. One example given during the session concerned an expert embellishing results in order to appear more convincing and secure a higher fee.

The discussion also suggested that an element of personal motivation or interest may be present in many of the misinterpretation scenarios shared during the session. However, it was also noted that such elements are rarely the main or most visible driver. In most cases, the wider professional and political context remains a more convincing and likely explanation for why professionals resort to data misinterpretation. This finding aligns with a broader body of integrity research which shows that problematic conduct in education, for example, is usually linked to policies and structural conditions in the sector rather than to the individual disposition of education participants to engage in malpractice (OECD, 2018; Milovanovitch & Lapham, 2018).

5.5 Making data more usable

Another reason discussed in the session was the wish to make data usable and shield its recipients from avoidable misunderstandings. The participants suggested that professionals may sometimes simplify the presentation of data because a fuller or more literal presentation could be difficult to interpret correctly and could itself produce misleading readings. In such cases, the purpose is not to steer audiences towards a preferred narrative, but to reduce the risk that technically correct but unstable or marginal results would be read in exaggerated or distorted ways.

The clearest positive example discussed in the session concerned the publication of statistics on brain-drain. In one of the countries present, programme-level results were prepared to show the share

of graduates who had emigrated. One very small programme had only a few graduates, all of whom had left the country. In strictly descriptive terms, the programme therefore had a brain-drain rate of 100 per cent. Yet publishing this figure as a programme-level result would almost certainly have invited a misleading interpretation. It could have made a marginal case look like a major national finding and could even have sent a distorted signal to prospective learners.

The responsible choice was therefore to avoid displaying such very small programmes as separate programme-level results in the main graphs, while retaining coverage of almost the entire target population and explaining the choice in the metadata.[2] Although the brain-drain figures for these very small programmes were still included in overall totals and calculations, they were only suppressed in the final programme-level presentation. This was a selective presentation of data, and in that sense a controlled deviation from a fuller display of the evidence. However, the purpose was to prevent misunderstanding rather than to create it. It also helped protect statistical confidentiality in programmes with very small numbers of graduates.

This example is important because it shows that selective presentation can serve a legitimate professional and public purpose. The act would have been problematic if small programmes had been hidden to suppress inconvenient evidence, or if the exclusion had not been disclosed. In this case, however, the methodological choice was documented, the underlying data were not falsified, and the published results still covered the vast majority of the relevant population.

While the wish to help users interpret data more responsibly may itself be a valid reason for selective presentation, it does have similarities to some of the other, purposefully misleading techniques described in Section 4, however. Therefore, the participants noted that usability can be a legitimate motive only if certain conditions are met. These include the clear and diligent documentation of choices in the metadata, public disclosure of these choices, as well as retaining coverage of the vast majority of the relevant population.

5.6 Low statistical literacy and insufficient training

Participants in the working session repeatedly stressed that misleading claims often arose from ignorance rather than intent. People compared unemployment rates without taking inactivity into account, used percentages without denominators, ignored sample sizes, overlooked distributions, or failed to reconstruct the demographic structure behind a statistic. Confusion between correlation and causation was another recurring concern, because it could turn a superficial association into a false explanation and then into poor policy reasoning.

The discussion placed such cases in a zone where people did not know what they did not know. This formulation matters, because it shifts attention away from motive and towards capacity. In such cases, the misinterpretation of data is a symptom of a deeper problem that people have with capacity and with lack of training needed to recognise when their reading of the data was unstable.

The participants also added a newer version of the same problem: intuitive AI tools could generate outputs that looked reliable and polished, even when users had little understanding of the underlying process. This reinforces the importance of statistical literacy as a safeguard against both old and emerging, automated forms of data misinterpretation.

5.7 Loss of metadata

Several national examples showed that misinterpretation can also arise when the data loses its context as it “travels”. Figures rarely circulate with all the notes, definitions, and methodological

[2] *Although some programmes were thus hidden in programme-level tables, the published results still covered almost the entire target population in terms of total number of graduates.*

explanations needed for a sound interpretation. Once these elements are dropped, technically correct numbers could easily support false conclusions.

The discussions suggested that the loss of metadata can happen to both national and international participants in the data-to-meaning continuum. A country cited a case in which a World Bank presentation of ILOSTAT employment data omitted an important note about a methodological break. Journalists and policy actors then reported a rise in employment that was not real in the way they assumed. Another country had a similar example, with old and new labour market data continuing to circulate side by side even after a methodological change had made them non-comparable.

A different version of the same problem appeared in the example of a gender pay gap that seemed to favour women, when in fact the result stemmed from a shift in labour market composition after many low-paid male refugees had entered the workforce. In all of these cases, the figures themselves were not fabricated. What disappeared was the contextual information needed to keep the figures tied to their original meaning.

5.8 Weak verification culture

The discussion ended up pointing to a weak verification culture as a common condition that allows all of the other reasons to operate more easily. The participants spoke of data being accepted on trust, of triangulation being expensive and time-consuming and therefore avoided, and of limited incentives to check claims based on data carefully. That alone made data misinterpretation easier to get away with.

Yet the discussion went further by also noting uneven scrutiny and double standards. Some countries or indicators were questioned rigorously, while others passed with little challenge. Seeing questionable claims accepted in that way could itself be demotivating for those who invested time and careful work in the collection and interpretation of the data. Misinterpretation thus persisted not only because individuals had reasons to engage in it, but because the surrounding environment tolerated it and, in some cases, facilitated it thorough weak or inexistent verification.

6. GUARDRAILS: IDEAS FOR RESPONSIBLE PRACTICE

Our working session started from the assumption that data misinterpretation is not inherently wrongful. The discussion repeatedly returned to the idea that, depending on purpose, disclosure, and fidelity to the data, the same interpretive act may fall either on the side of legitimate practice or on the side of malpractice.

If all misinterpretation were bad by definition, the solution would theoretically be quite simple and would be dealt with via prohibition, sanctions, and enforcement. However, the working session insisted on the more difficult premise that some selective interpretation, simplification, and framing may be justified and may even serve the common good. We introduced the idea that, despite the negative connotation of the term, data misinterpretation can be a legitimate professional practice. At the same time, misinterpretation can also amount to malpractice. Accepting this dual nature creates an obligation to provide guidance and propose guardrails.

One could argue that once selective interpretation remains faithful to the data, serves a legitimate purpose, and is disclosed transparently, it no longer amounts to misinterpretation at all, but simply to interpretation. However, there are conceptual reasons to retain the term “misinterpretation”. The term signals that what is involved is a type of interpretive act with an ethically and epistemically risky structure. A legitimate misinterpretation is not just “interpretation”, because it remains a controlled deviation from a fuller reading of the data, one that reframes aspects of the evidence in order to produce a particular understanding. “Misinterpretation” therefore points to the risk of such acts being taken too far and underlines the need for safeguards and justification.

The starting point of discussions about guardrails was the distinction between intentional and unintentional misinterpretation. The two notions represent two very different problems. Unintentional misinterpretation belongs mainly to the domain of professional capacity and is usually a manifestation of weak statistical literacy, insufficient training, ignorance of context, or misplaced confidence in tools and outputs. Examples here include misleading country comparisons based on incomplete labour-market concepts, failure to recognise methodological breaks in time series, and cases where people misread data simply because they do not know what they do not know.

Intentional misinterpretation, on the other hand, means that people know what interpretative choices are being made (and why), and that these choices call for professional judgement, discretion, and integrity. The main question is: how far can we go in shaping the (intentional) interpretation of data before we cross the line? In other words, when does the framing of data become manipulation and with this, an integrity problem?

The quest for a dividing line between legitimate practice and malpractice is neither new nor unique to data work in education and employment. In medicine, for instance, where practitioners are often expected to exercise judgement under pressure, much has been debated and written about where the line lies between acceptable clinical discretion and conduct distorted by conflict of interest or disregard for professional duties. The same is true of public administration more broadly. Many integrity frameworks in government are built precisely to distinguish legitimate administrative judgement from favouritism, politicisation, or misuse of office.

In education, and especially in higher education, there are many initiatives on ethics, anti-corruption, academic integrity, and standards of conduct. One that focuses explicitly on intentionality is the Integrity of Education Systems (INTES) framework. Initially developed in the OECD anti-corruption context, it offers a pragmatic way of conceptualising the point at which professional conduct ceases to be a matter of acceptable judgement and becomes an integrity problem. The framework has been used for well over a decade in various countries, which suggests that its core distinctions are robust enough to travel across contexts whilst remaining relevant.

INTES suggests that the difference between borderline conduct and an integrity violation lies in the fact that someone knowingly departs from rules, standards, principles, commitments, and/or obligations in pursuit of some personal benefit (OECD, 2018). This provides a useful starting point for interpreting the examples discussed in the working session, where many instances of legitimate data misinterpretation were in fact choices compatible with professional obligations such as honesty, fairness, transparency, and fidelity to evidence. Conversely, most examples of problematic data misinterpretation involved the prospect of some undue benefit or advantage, often in disregard of prior commitments to act in accordance with values and principles.

Accordingly, one guardrail can be drawn around the purpose and beneficiary of data misinterpretation: whether interpretive choices serve a legitimate professional or public purpose, or whether they are shaped in pursuit of undue personal, political, reputational, or institutional advantage.

Another guardrail highlighted by participants was transparency as a methodological principle. The workshop accepted that data professionals may legitimately narrow the scope, foreground one message, or simplify a dataset, but for such choices to be acceptable, they must be documented. Users must be told what was left out, how the selection was made, and why.

The discussion also described fabrication of data as an “outer boundary” of legitimate data misinterpretation. The working session repeatedly accepted that people may choose what to highlight or suppress, what message to foreground, or what level of simplification is appropriate. However, it also drew a firm distinction between those choices and acts such as inventing numbers or quietly adjusting them. Once that line is crossed, one is no longer shaping the meaning of data but is committing malpractice by tampering with the evidentiary basis itself. This is why participants saw this last boundary as non-negotiable.

Although unintentional misinterpretation of data was not a central focus of the discussion, the working session did point to a number of practical measures that can help reduce it. These measures centred on strengthening the quality of data work by building the capacity of data professionals to preserve context, disclose uncertainty and relevant methodological choices, avoid misleading aggregation, refrain from causal claims without sufficient grounds, and strengthen verification through triangulation and double-checking.

7. CONCLUSION

This article has argued that data misinterpretation is an ordinary part of professional practice and, at times, a necessary one. It forms part of the routine task of translating data into meaning, which often takes place under conditions shaped by communication demands, institutional incentives, political expectations, and the need to make evidence usable.

Against that background, the main question explored in the working session concerned the integrity of such conduct: where does selective, strategic, or deliberately simplified interpretation cease to be compatible with professional obligations and enter the territory of malpractice?

The discussion provided some answers and proposed guardrails. At the same time, participants suggested that the line between acceptable misinterpretation and malpractice cannot always be drawn in a clear or stable way beforehand, because its placement depends heavily on context and purpose. This also means that responsibility for "doing the right thing" cannot be outsourced to fixed rules alone. Even in the presence of guardrails, professionals need to judge and bear continuing responsibility for how far they go in shaping the interpretation of data.

This also suggests that the responsible interpretation of data is not a one-off act. It is a continuing effort to make the right judgements under pressure. Thus, it is an exercise in integrity, understood as the capacity to remain true to professional commitments and obligations even under adverse conditions (Milovanovitch, 2019).

The working session itself is a proof that honest professional dialogue, peer reflection, and context-sensitive guidance have real potential in this area. The discussion showed that, where rigid formulas fall short, structured exchange can help professionals reflect on difficult boundary questions and test their judgement against that of others. In turn, this suggests that questions of data interpretation and misinterpretation could and should be more present in research, professional guidance, and integrity frameworks.

ACRONYMS

CRISP-DM	Cross-Industry Standard Process for Data Mining
ETF	European Training Foundation
GAF	Global Assessment of Functioning
GDP	Gross Domestic Product
GSBPM	Generic Statistical Business Process Model
INTES	Integrity of Education systems
KIESE	Key indicators on education, skills and employment
OECD	Organisation for Economic Co-operation and Development
PISA	Programme for International Student Assessment (PISA)
SDG	Sustainable Development Goals
UN	United Nations
UNESCO	United Nations Educational, Scientific and Cultural Organization
VET	Vocational Education and Training

REFERENCES

American Statistical Association. (2022). *Ethical guidelines for statistical practice*. Zenodo. https://doi.org/10.5281/zenodo.7092386

Bromley, P., Nachtigal, T., & Kijima, R. (2024). Data as the new panacea: Trends in global education reforms, 1970–2018. *Comparative Education, 60*(3), 401–422.

European Training Foundation. (2022). *Torino Process 2022–2024: Towards lifelong learning: A framework for monitoring system performance and reviewing policies for lifelong learning*. European Training Foundation.

Gorur, R., Dey, J., Faul, M. V., & Piattoeva, N. (2023, June 13). *Decolonising data in education: A discussion paper*. NORRAG. https://www.norrag.org/decolonising-data-in-education-a-discussion-paper/

Government Analysis Function. (2017). *Communicating quality, uncertainty and change*. Office for National Statistics. https://analysisfunction.civilservice.gov.uk/guidance/communicating-quality-uncertainty-and-change/

Gummer, E. S., & Mandinach, E. B. (2015). Building a conceptual framework for data literacy. *Teachers College Record, 117*(4), 1–22. https://doi.org/10.1177/016146811511700401

Heyneman, S. P. (1999). The sad story of UNESCO's education statistics. *International Journal of Educational Development, 19*(1), 65–74. https://doi.org/10.1016/S0738-0593(98)00068-6

Jarke, J., & Breiter, A. (2019). Editorial: The datafication of education. *Learning, Media and Technology, 44*(1), 1–6.

Milovanovitch, M. (2013). *Fighting corruption in education: A call for sector integrity standards* (Working Paper No. 31). Edmond J. Safra Research Lab, Harvard University.

Milovanovitch, M. (2019). Expectations, distrust and corruption in education: Findings on prevention through education improvement. *Current Issues in Comparative Education, 21*(1), 54–68.

Milovanovitch, M. (2025). An inclusive classification of illicit practices in education and higher education. In *Handbook on corruption in higher education* (pp. 77–97). Edward Elgar Publishing. https://doi.org/10.4337/9781035320240-8

Milovanovitch, M., & Lapham, K. (2018). Good intentions cast long shadows: Donors, governments, and education reforms in Armenia and Ukraine. In I. Silova & M. Chankseliani (Eds.), *Comparing post-socialist transformations: Education in Eastern Europe and the former Soviet Union*. Oxford University Press.

OECD. (2018). *Integrity of education systems: A methodology for sector assessment*. OECD Anti-Corruption Network for Eastern Europe and Central Asia.

Romero, C., & Ventura, S. (2012). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3*(1), 12–27.

Shearer, C. (2000). The CRISP-DM model: The new blueprint for data mining. *Journal of Data Warehousing, 5*, 13–22.

UNESCO Institute for Statistics. (2025). *Proceedings of the UNESCO Conference on Education Data and Statistics* (Statistical Commission background document, 56th session). UNESCO Institute for Statistics.

Verger, A., Fontdevila, C., & Zancajo, A. (2016). *The privatisation of education: A political economy of global education reform*. Teachers College Press.

Williamson, B. (2017). *Big data in education: The digital future of learning, policy and practice*. SAGE.

World Bank & UNESCO Institute for Statistics. (2003, January). *A framework for assessing the quality of education statistics*. World Bank Development Data Group & UNESCO Institute for Statistics.