ETF Working together Learning for life

European Training Foundation

THE EU AGENCY SUPPORTING COUNTRIES TO DEVELOP THROUGH LEARNING

PLENARY PROCEEDINGS OF THE ETF MONITORING FORUM 2025: EVIDENCE IN ACTION

Summary and analysis report

Disclaimer

This report was prepared by Elena Walls with contributions from Mihaylo Milovanovitch and Stefano Lasagni from ETF.

PREFACE

The review is part of the European Training Foundation's (ETF) core function of monitoring of education, skills, and employment developments in partner countries. Each year, the ETF collects and analyses evidence to track progress in these areas and to support countries in making informed policy decisions. This work is carried out primarily through two complementary initiatives: the KIESE (Key Indicators on Education, Skills, and Employment) data collection, and the monitoring strand of the Torino Process. The results of these activities culminate annually in country progress reports and a cross-country monitoring report. In October 2025, the ETF convened its Monitoring Forum in Milan to present the findings from the 2025 monitoring round, to gather feedback from partner countries, and to discuss priorities for the 2026 cycle. This review will support ETF's monitoring efforts by elucidating substantive exchanges on new developments in ETF's monitoring work that occurred during the convening and summarizing the resulting insights in a report that can be used to inform ETF's future work.

CONTENTS

INTRODUCTION

Following its mandate to support the improvement of the education and skills development sectors in its partner countries, ETF engages in continuous monitoring of accessibility, quality, relevance, and system management of partner country education systems, using data from international and national databases supplemented with expert reviews. These efforts result in a highly influential monitoring report, published annually, which describes how well partner countries provide accessible, equitable and high-quality learning opportunities and highlights cross-country trends[1].

The most recent annual publication was released in the fall of 2025 and was the subject of the 2025 Monitoring Forum in Milan, Italy, that brought together representatives from 19 ETF partner countries. During the day-long Forum, the ETF monitoring team presented the report findings and engaged participants in thematic discussions around the report's key points, the proposed new Skills Gap Index, the use of AI in monitoring, and the question of data misinterpretation.

This thematic report provides a reflective summary of the key themes that emerged from the Forum presentations and discussions and highlights areas of potential future directions for the monitoring efforts by the ETF team. The report starts with the Forum overview, followed by the summary of key themes. It concludes with a set of recommendations for the future monitoring reports and dissemination events.

[1] https://www.etf.europa.eu/en/what-we-do/transforming-vocational-education-and-training-etf-monitoring-initiatives-and

OBJECTIVES AND METHODOLOGY

The main objective of this review is to produce a reflective, analytical summary of the proceedings from the Monitoring Forum, based the transcript and the video recording, with a focus on how international evidence is received, contested, and reinterpreted by national actors. The review aims to identify common themes and ideas that emerged during the Forum and validate existing analytical findings. It also endeavours to capture how national stakeholders reacted to international data and interpretations. The report is not structured to mirror the agenda of the Forum but rather organise and present insights from the Forum proceedings according to the substantive ideas and cross-cutting issues visible in the transcript. The analysis remained grounded in the actual data, while also drawing out patterns, links, or questions that become apparent only when the proceedings are viewed as a whole.

Some themes in the proceedings were already pre-determined by the Forum's agenda; for these themes the analysis aimed to validate and legitimize the existing analysis conducted by the ETF team. Other themes emerged directly from the Forum transcript.

Methodology and Approach

The review used a qualitative data analysis approach to answer research questions, specifically utilizing the grounded theory methodology, which is particularly well-suited to uncovering implicit assumptions, conflicting views, and cross-cutting dynamics that may not be visible through agenda-driven or deductive analysis.

Grounded theory is a qualitative research and data analysis methodology that provides a systematic, inductive approach for analysing large bodies of textual data by allowing themes and patterns to emerge directly from the data rather than being imposed based on preexisting ideas. Through iterative cycles of close reading and coding, the review sought to identify recurring concepts, phrases, and points of tension across the text. These initial codes were refined and grouped through axial coding, which explored relationships among concepts and highlights patterns of convergence, divergence, and contradiction.

Due to the objectives of this assignment, the analytical process was limited to the review of the proceedings and did not involve any additional primary or secondary data collection and analysis. It also did not include any additional desk research outside of the documents used in or produced during the Forum.

FORUM OVERVIEW

The Forum consisted of three distinct parts. First, ETF monitoring team presented the highlights from the 2025 Monitoring Report on the key developments in education, skills, and employment from the ETF partner countries. The presentation included information about the focus and thematic coverage of the report as well as data sources; explained which indicators were used in the analysis; and summarized the results. The presentation also noted issues with data availability and comparability. Cross-country findings were contextualized with data on the gross domestic product (GDP) grown and inflation rate, and the size of youth population in included countries. Results highlighted included data on public spending on education, access and participation, conditions of teaching and learning, and quality and relevance of learning.

Second, the ETF team presented information on ETF's new initiative called "Skills Gap Index" (SGI) which aims to quantify the misalignment between the country education/training systems and labour market needs. The SGI will produce a composite measure that unites and quantifies several dimensions of the skills gaps impeding economic growth and social development. The team explained the data sources and the methodology for rescaling all included indicators to a common scale with a range of 0 to 100. The index is intended offer policy makers a succinct overview of how country education, training, and employment systems work together to close the skills gaps, and allows for tracking progress at a country and cross-country levels. The team solicited feedback from the participants and outlined their plans for the inclusion of the SGI in ETF's 2026 Monitoring Report.

The final part of the Forum was devoted to the discussions of the two novel areas of analysis and data use considerations:

- Using AI for policy and system monitoring, and
- Data misinterpretation: a problem or a tool?

The discussions took place in two breakout groups. ETF facilitators took notes during the group discussions and presented the summary during the close-out plenary session.

FINDINGS: KEY THEMES

The Forum followed an interactive format in which the ETF team alternated presentations with Q&A sessions and country-level updates by the delegates. Warm, welcoming style of interaction, adopted by the Forum leads, was conducive to participation in the discussions by the delegates, encouraging them to reflect on the presented content and contribute to the conversation.

Key themes pulled out below should not be viewed as an exhaustive summary of the Forum discussions but rather expositions of topical areas that were repeatedly discussed during the Forum and insighted a lot of interest from the delegates. They are framed as questions to suggest an inquiry-based approach as well as the contested nature of the highlighted areas.

What role does data play in policy and programmatic decision-making?

The Forum explored this question from different angles. Firstly, is data an objective description of reality or can its meaning be changed depending on the interpretation, context, and purpose? "*Data both matters and does not matter, because **what truly counts is what you do with it***", emphasized Mihaylo Milovanovitch, ETF's team leader for monitoring and systems performance and the moderator of the Forum, in his opening remarks. However, the Forum stops short of diving into the question of use by policymakers. Rather, it delves into the question of **juxtaposition of nationally produced versus internationally produced data**, both at the collection/aggregation stage, and at the interpretation/dissemination stage. This is both due to the increasing polarization and nationalization of politics as a global trend, and increased scepticism about the value of large-scale country comparisons which necessarily lose a significant amount of context-specific details. Such details are seen as valuable and even defining, as countries choose which data and evidence to use in their policy and decision-making, how to interpret them, and how to leverage them in national and international communication. As Milovanovitch noted, "*international harmonisation loses detail. Important nuances vanish*."

Speaking more broadly, **education systems are embedded in cultural contexts** making the efforts at comparisons especially problematic. As an ETF presenter Stefano Lasagni noted, "*We are not scientists counting molecules with fixed definitions. Especially in education, each country has its own concepts and structures*." This is one of the reasons why ETF is now using expert opinions in addition to quantitative data. The presentation on the key findings of ETF's 2025 Cross-Country Monitoring Report illustrates these tensions, when explaining the reasoning behind the data selected, in interpreting specific data points, and in soliciting feedback from attendees. The presenters emphasized this point to the country delegates: "*If you see something strange—say, your country missing or looking worse than it should—let us know. We can always "reinterpret" the data a little*." This approach does not signify fluidity in the data points, but rather a recognition that data are inherently not good or bad; their interpretation depends on national priorities.

Reflecting on the role data does play or can play in policymaking relating to vocational education and training (VET), the ETF team emphasized the potential role of fit-for-purpose composite indicators to bring together educators, employers, employer and trade associations, and other relevant stakeholders to facilitate a dialogue about skills shortages that impeded economic growth. **The new Skills Gap Index (SGI)** is designed to do just that, and the presentation generated a lot of interest as well as questions from national delegates, who raised the question of "who is at the table" deciding on what constitutes a "mismatch"? **The practical utility of the index will likely depend on the degree of a buy-in** into the inclusivity and robustness of the process as well as the success in the socialization of its intended purpose.

What does data tell us about education systems and how they support learning?

The question of education systems and monitoring their signals was a recurring theme during the Forum. The system-level indicators are designed to capture system inputs, process measures, and data on select outcome. Of the indicators discussed during the Forum, of particular interest was system performance indicators. In Milovanovitch words, *"**system performance means the extent to which a country delivers on its commitments in access, quality, and management**. It is not about what policies exist, but whether they work in favour of the people they target*", and what education-related investments, laws, and strategies actually deliver for the learners in those systems. System performance indicators presented at the Forum summarize skills, employment, and education in quantitative terms in a composite form. In his presentation, Lasagni stressed the particular advantage of ETF's approach of combining data from international data sources with qualitative expert assessments, to incorporate harder-to-measure aspects of system performance. The use of international data sources is necessary for the creation of cross-country comparisons, since national statistical offices frequently have differing definitions, making their data inherently incomparable.

Particularly noteworthy is the use of the **Labor Force Survey (LFS)** which nearly all countries conduct. While there are some methodological differences in the national implementation of this survey, it remains a critical data source for tracking changes in labour force participation for different populations, to provide up-to-date information as well as trends for policy and decisionmakers.

The core of ETF's monitoring approach is in converting a vast amount of education and employment-related data into composite indicators that help stakeholders track the "big picture" of system performance, to support policy-related dialogue, and identify which areas need particular attention. These indicators also provide clear and cohesive information about the results of national investments in education and skills development. Especially helpful is the disaggregation approach which allows users to view information for populations of interest, such as **all youth, disadvantaged youth, female and male learners, and vulnerable adults**. An important question was raised by the delegates about changes in outcomes at a cohort level. For example, **how do the foundational skills of youth today compare to the foundational skills of youth five, ten years ago?** Do these skills improve or do they remain at the roughly the same level, signalling the stagnation in the quality improvements of the system? The advantage of using consistent indicators and methodological approaches is that it allows us to answer such questions without pricey longitudinal studies.

What does the data tell us and what does it not?

The Forum discussed the depth and breadth of KIESE ("Key Indicators on Education, Skills, and Employment") collected and analysed by ETF which serves as the key quantitative data source for the annual Monitoring Reports. Both **data availability and data comparability** were shown to have limitations that are likely to impact the reliability of cross-country comparisons as well as conclusions that can be drawn. It was noted that a lot of effort goes into data collection, but there are differences in how the data are collected across countries that impacts data comparability and validity of cross-country comparisons. One Forum participant noted that "*for evaluations to be credible in our countries, we must be confident about data collection and methods. Since national statistical offices are responsible for official statistics, evaluations should reference NSO data. We see a similar issue with other evaluations (e.g., OECD): international institutions cite each other's data, but our statistical office is the official reference, and we follow standard methodological rules.*" The challenge highlighted was the national differences in how NSOs function and in terms of their willingness as well as capacity to adhere to international guidelines for data collection and to share information.

Forum participants noted similar issues with the ETF-proposed new Skills Gap Index, questioning how ETF might ensure cross-country comparability, given national differences in definitions and data collection processes and methodologies. International data would not solve this issue either since

international datasets tend to have a several years lag time and potentially have disagreements with national data, leading to uncertainties.

The presentations show a limited number of sources where the data, both quantitative and qualitative, originate, and noted the impact on data reliability. Of the three data categories highlighted – access, quality and relevance, and system management – data on system management are shown to have most data. As ETF presenter Marco Barreca noted, "*for access, coverage is especially limited. For system management, there is often a trade-off between the availability of data and the breadth of the outcome being measured*." PISA-derived data were highlighted as an example of a single source of data on quality and relevance of the education system, with data missing from the countries that do not participate in PISA.

Supporting National Capacity Strengthening

Delegates from several countries brought up an **issue of transparency relating to the methodological approaches behind data collection**. The matter was raised specifically within the context of helping countries establish and/or strengthen their own national data collection processes by improving their understanding of internationally used methodological approaches.

Delegates suggested producing a **methodological annex** to the report that would describe the methodological approaches used in collecting data included in the report. In addition to using such an annex to strengthen national data collection system, it could be used to inform data collection contracts issued by national statistical offices and Ministries that wish to outsource some of this work to private entities.

One of the critical points made was on the difference in the data that describes infrastructure and policies, and data that describes people. In his presentation on the findings relating to system-level indicators, Milovanovitch noted that ETF has "***far more information about systems—the "infrastructure" of education—than about the people in those systems***. *We have invested a lot in tracking big structural items: numbers of providers, funding levels, and so on. Yet when it comes to learners, the data are thinner; and for teachers, thinner still. In fact, when we tried to build indicators for system performance, we realised there is no internationally agreed indicator on school leaders—nothing that shows who they are, how well they do their jobs, or whether they are supported to do them well.* ***We know the systems; we do not know enough about the people inside them***." This fundamental limitation of the international data systems and approaches was also noted several times throughout the Forum, from different perspectives. For example, it was noted that there are not many tracer or longitudinal studies that would provide insight about the relationship between labour market experience and foundational skills in adults, as well as changes in labour force participation.

It is helpful to think about education and training for labour market as an ecosystem, with a variety of actors playing different roles. One participant highlighted the importance of **including employers as well as other key actors in the ecosystem**: "*When we say "ecosystem", it must include everyone. It cannot be made only of education actors. Even "mobilisation"—who is [involved]? Are we tired of mobilising VET schools, institutions, ministries—while those saying "you are not good" are not at the table*?" Another good example came from a delegate who discussed **tracking "brain drain**" in their country, and wondering how data systems might integrate such information in monitoring efforts.

Another substantial data issue discussed during the Forum was the use of some indicators or composite indicators as **proxies for the phenomena that are difficult or even impossible to measure** directly. Oliver Toskovic noted in his presentation on the SGI that "*the key methodological issue is composite scores—how to mix apples and oranges (or, as we say back home, grandmas and*

frogs). Initially we had eight dimensions, each with outcomes, each with multiple indicators. We had quantitative indicators, but due to missing data we added qualitative ones via expert self-reports/ judgements for some outcomes." These methodological complexities and potential solutions were at the centre of the discussion on the new SGI currently being developed by the ETF team.

An interesting point of **policy responses to data** was raised by a delegate from Kazakhstan who shared her country's solutions to the persistent skills gap challenge. While the Forum did not intend to cover policy landscape relating to the education and skills development in partner countries, **the monitoring data are an integral part of the Torino process**.

To conclude, issues with conceptual definitions behind indicators and indicator data accuracy and completeness are real, and it is important to be transparent about these issues and decisions made throughout data analysis and interpretation. As Milovanovitch summarized, "*If we reject all international data, we cannot do cross-country monitoring. If we rely only on national data, we hit other problems: some countries have no public data, or we cannot access them. The practical solution is* ***transparency plus context: cite the source, show the calculation trail, and interpret with the qualitative inputs***".

How well is the context reflected in the monitoring report discussions?

The Forum sessions emphasized the importance of understanding the context where the data originated from. **"Context" is understood as an intersection of the socio-economic situation and demographic factors**, such as countries' GDP, inflation rates, the size of the youth population, and the current employment rates by education attainment levels.

The ETF team returned to the role of the context in data interpretation throughout the Forum. As Milovanovitch noted in discussing one of the presented indicators, "*before concluding "high is good" or "low is bad," we need country context. […] A country may prioritise getting young people into the labour market quickly after compulsory education; another may aim to raise educational attainment for stronger employment prospects. In some places, nearly everyone stays on—perhaps because jobs are scarce for those with only compulsory education, or because families expect higher education regardless of motivation. Again,* ***interpretation requires context***".

ETF presenters noted that partner countries are diverse and highly idiosyncratic, making regional aggregations difficult. For example, ETF presenters noted that the way countries are organized into groups in cross-national data presentations is changing, with some groupings becoming less meaningful as the geopolitical context shifts. Countries traditionally grouped together, such as Eastern European countries, south-eastern European countries, and Central Asian countries, might not see their geographic neighbours as natural comparators, for a variety of reasons. While such familiar groups may be convenient, they are also analytically risky since they tend to group countries with diverse contexts and system metrics. Another point brought up by the ETF team was that some traditional groupings reflect a Eurocentric perspective which some stakeholders might object to and prompt analysts to use them less frequently in the future.

While recognizing the limitations inherent in cross-contextual analytics and data interpretations, there is an undeniable value in insights resulting from cross-country comparisons. In discussion, the Forum participants recognized how such insights might inform national as well as cross-national discussions aimed at **addressing persistent challenges with skill gaps and underemployment of graduates**, among others. In particular, this topic was widely discussed in relation to the new SGI, when delegates raised concerns about comparability of data points from different countries given known (and unknown) issues with variability of definitions, data collection protocols and methodologies, data quality, etc. In particular, the issue of errors being amplified during the rescaling process was raised as

a critical one, potentially jeopardizing the validity and reliability of the resulting estimates. The ETF team shared a **plan for testing reliability of existing indicators**, to determine which ones are more reliable and can be included in the SGI with greater certainty.

Future Trends

Two breakout sessions at the conclusion of the Forum explored themes that are likely to affect monitoring work in the future: using artificial intelligence for policy and system monitoring, and data misinterpretation. The breakout group focusing on AI in research identified use cases for AI throughout the stages of data lifecycle, from data collection to analysis, reporting, and dissemination, highlighting areas where AI can be particularly useful in improving the speed and quality of dataflows. The key takeaway was that AI can be particularly helpful in reporting and dissemination, while humans remain crucial for data collection and analysis.

The second breakout group focused on data misinterpretation which refers to the incorrect understanding, analysis, or presentation of data, leading to false or misleading conclusions. The group highlighted the distinction between data description, analysis, and interpretation, and discussed cases of intentional versus unintentional misinterpretation. Cases of legitimate misinterpretation were also noted.

The two topics converged around the issue of AI interpreting the data and the lack of transparency about the process, leading us to potentially question the results. How would the user know if the interpretation is legitimate or distorted? Human oversight coupled with transparency is critical for preventing intentional misuse or unintentional errors.

RECOMMENDATIONS

Recommendation 1. Include the description of the intended audiences, use cases, and policy research questions that ETF's monitoring report might help answer. Incorporate examples of policy responses, when available.

The monitoring report's rich contextual and comparative information provides great insights into the situation with education and skills attainment and relevant outcomes in the included countries. This information can be used in a variety of ways by target audiences, but the report stops short of describing specific use cases of this information by national policymakers or international organizations. A clear description of audiences and data use cases for each audience might be helpful in increasing the report influence and use. Specifically, an exposition of policy questions and situations that the report might be useful for, both at the national and international levels, might be very helpful in furthering the report's utility and use. Additionally, a natural segue from policy questions is policy responses employed by partner countries; allowing space for examples of such policy responses might elucidate the link between the monitoring report and the Torino process even better.

Recommendation 2. Include a glossary of the key terms used (such as access, quality, system, employment – among others) with a clear note of what is not included in the definition and potential data limitations.

The question of definitions of concepts is a critical one. The core concepts used in the report have multiple competing definitions. For example, access is traditionally defined as the extent to which individuals are able to enter and participate in formal education systems (per UNESCO definition), often measured through enrolment and participation rates, and uses indicators such as enrolment rates, admission eligibility, and availability of schools or seats to measure access. Increasingly, international actors are interpreting "access to education" with a broader lens, to incorporate physical and financial accessibility, inclusivity, cultural and contextual relevance, quality, and meaningful participation. The expanded definition is centred on the idea of education access as having the opportunity to obtain quality education for every individual, regardless of their background, socioeconomic status, or demographic characteristics.

Differences in interpretation of various terms are likely to be embedded in the data collection systems and might warrant an explanation or at the very least a side note. For example, some countries might not include migrant workers as part of their Labor Force Surveys, or migrant children when they count enrolment rates. Many official data systems do not include non-formal economic activity as part of employment statistics. Even if such exclusion is not intentional, it might occur in contexts where self-reporting of non-formal economic activity in a survey carries risks for the respondents. Such data (in)completeness may signal policy priorities relating to inequities in resource distribution. Highlighting such potential data gaps may help bring attention to the "missing" populations and reinforce the principle of "education for all". At the very least, the inclusion of definitions along with notes on potential data incompleteness will help the audience to position information within the policy questions more clearly.

Recommendation 3. Add an annex with methodological notes on the data collection on the monitored indicators in future monitoring reports

Delegates from multiple countries requested this information, to help them implement internal improvements for the national data collection efforts and, in the end, align them better with international data collection standards. Such an annex would be particularly useful for the smaller countries where national statistical capacity may be weaker. The added value of the annex would be to promote improvements in the whole value chain of data production, analysis, and use by ensuring the shared understanding among stakeholders of the critical methodological elements.

Recommendation 4. Consider nuancing discussions of education spending with actual information, where available, on allocations versus real expenditures ratios.

Both the report and the Forum presentations discussed at length the allocation of financial resources to the education and skills development sector, using “allocation” and “expenditure” interchangeably. However, there are practical differences between these terms. Many, if not all, countries spend significantly less on education than budgetary allocations, and even less than publicly committed. While regrettably international organizations do not require national authorities to report on the actual expenditures in the education sector, highlighting this information gap would help bring attention to this issue. For example, in the discussion of funding it was noted that “*Many—perhaps surprisingly many—of countries spend well above typical international reference points. Spending 6–7% of GDP, and in some cases more than 7%, is enormous*”. The reality may in fact be quite different and many countries that commit to high percentage in education allocations end up spending significantly less.

Recommendation 5. Implement and publish sensitivity analysis on the new SGI, both to ensure transparency and improve the long-term index’ reliability.

In statistical analysis, sensitivity analysis is a method for assessing how the data analysis results change when underlying assumptions, input data, or parameter choices are varied. In the context of composite variables such as SGI, sensitivity analysis examines how different decisions—such as indicator selection, weighting schemes, normalization methods, or aggregation rules—affect the resulting composite scores and rankings, thereby evaluating the robustness and credibility of the composite measure under uncertainty. Sensitivity analysis on the SGI is likely to achieve two important objectives. First, it will help users understand which data points or assumptions make most impact on the final composite score. Second, by demonstrating the impact individual indicators have on the index, the analysis will help shore up support for improving data systems.

ACRONYMS

AI	Artificial Intelligence
ETF	European Training Foundation
GDP	Gross Domestic Product
KIESE	Key Indicators on Education, Skills, and Employment
LFS	Labor Force Survey
NSO	National Statistical Office
OECD	Organisation for Economic Co-operation and Development
SGI	Skills Gap Index
UNESCO	United National Educational, Scientific and Cultural Organization
VET	Vocational Education and Training