CO-009

MEE: Manufacturing Execution Evaluation

Po-Hsun Chueh[1] Joseph K.H. Wang[2]

Abstract: Impact evaluation (IE) offers a robust methodology for evidence-based decision-making, crucial for assessing new technologies and management practices. This framework, vital for both industrial and SME sectors, relies on data from control and treatment groups, and pre- and post-intervention data. Previously, SMEs lacked such data, but advancements like the Industrial Internet of Things (IIoT) now provide field data. However, evaluating interventions in single factories remains challenging. We propose solutions: digital twins for simulation and regression models for hypothetical outcomes. Our research presents a framework for IE implementation in manufacturing, integrating case studies for analysis.

Keywords— Impact Evaluation, IIoT, Industrial 4.0

1. Introduction

Impact evaluation assesses the changes that can be attributed to a particular intervention, such as a project, program or policy, both the intended ones, as well as ideally the unintended ones.[1]

Over the past three decades, impact evaluation has emerged as a powerful tool to assess the effectiveness of programs and interventions. For instance, it has been instrumental in evaluating the impact of microfinance initiatives[2], [3], helping to determine whether these programs truly empower beneficiaries by increasing their income.

Impact evaluation, however, is not limited to international aid. In fact, it can be applied to any decision-making process. Provided we can clearly define the desired outcome and gather relevant data before and after implementing a change, impact evaluation allows us to assess the true impact of that change. This powerful approach is equally beneficial for industrial companies seeking to meticulously assess the development of new technologies and management techniques.

A successful impact evaluation hinges on collecting data from control and treatment groups, as well as pre- and post-intervention data. However, historically, manufacturing facilities, especially Small and Medium-Sized Enterprises (SMEs), often lacked adequate monitoring data. This presents a challenge for applying impact evaluation in these contexts.

For manufacturing companies, traditional evaluation systems often rely on metrics like product yield rate, energy consumption, and Overall Equipment Effectiveness (OEE).[4] While these metrics provide a clear and quantifiable picture, they can be lagging indicators. For instance, changes in production processes may not be reflected in yield rate data for weeks.[5] Additionally, concerns exist about the authenticity and reliability of data obtained through manual recording or manipulation.

A major challenge in applying impact evaluation to individual manufacturing plants lies in establishing counterfactuals. Unlike policy evaluations, where control groups are often feasible, isolating the impact of specific decisions within a complex factory environment can be difficult. Randomized controlled trials (RCTs) are often impractical at this level, and external factors can easily influence outcomes.

To address this challenge, we propose two potential solutions. The first leverages digital twins, virtual representations of physical factory systems. By simulating different scenarios within the digital twin, we can predict the potential impact of various decisions without disrupting actual production.

The second solution involves constructing a regression model using existing plant data. This model would identify key factors influencing production outcomes and allow us to estimate hypothetical scenarios based on the implemented changes.

Both of these solutions rely heavily on the availability of accurate and reliable field data. Fortunately, the emergence of Industrial Internet of Things (IIoT) offers a transformative solution. As the foundation of Industry 4.0 and data-driven manufacturing, IIoT enables real-time data collection directly from factory floors. This eliminates the delays inherent in traditional methods and provides a more accurate picture of production processes. Additionally, IIoT data is often tamper-proof and automated, enhancing its reliability.

This paper details how to evaluate production execution in manufacturing organizations. We begin by providing a foundational understanding of Impact Evaluation. Next, we introduce the Manufacturing Execution Evaluation (MEE) framework. Following that, we showcase the implementation of MEE through a practical case study. Finally, we discuss key takeaways and summarize the overall approach.

2. Background

2.1 Evidence-Based Decision

Manufacturing enterprises constantly strive to optimize production for higher profitability. This often involves a combination of adjustments, from routine maintenance to scheduling changes, aimed at boosting equipment utilization and production efficiency. However, a critical challenge lies in pinpointing which specific actions are driving these improvements. When implementing multiple adjustments simultaneously, it becomes difficult to isolate the impact of each decision and accurately measure the true gains in efficiency.

The growing trend of evidence-based decision-making is reshaping operation optimization in manufacturing. Impact evaluations play a crucial role in this shift, allowing us to move the focus from what we put in (inputs) to what we get out (outcomes and results).[6] This empowers us to not only analyze the effectiveness of individual decisions at the production site, but also to make informed trade-offs between multiple optimization measures for locally optimal solutions.

2.2 Impact Evaluation

Impact evaluation is one of many approaches that support evidence-based decision-making, including monitoring and other types of evaluation.[6]

In general, the following evaluation methods are most frequently used[7]:

- Before–After Comparison
- With–Without Comparison
- Difference-in-Differences Method
- Regression Discontinuity Designs
- Instrumental Variables

---

[1] Volapu Research, Kunshan, PR. China, 215332

[2] Volapu Research, Kunshan, PR. China, 215332, To whom correspondence should be addressed; Email: k.wang@acm.org

- Randomized Evaluation

Simple comparisons, like before-after or with-without designs, can be attractive due to lower planning and data collection needs. However, they struggle to isolate a program's true impact because they don't account for external factors influencing outcomes.[7]

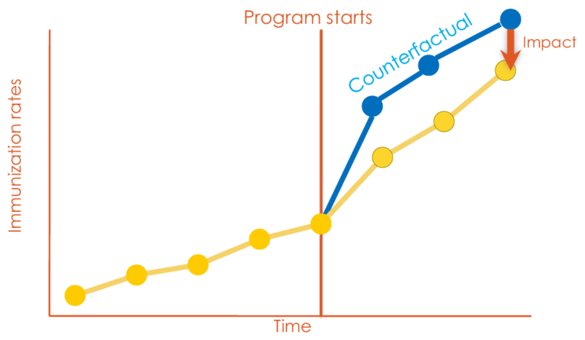
Advanced techniques like Randomized Discontinuity Design (RDD) and Instrumental Variables (IV) offer powerful tools for causal impact evaluation. However, their applicability is limited to specific scenarios with particular program assignment mechanisms. While program implementers can sometimes create such situations, these opportunities are rare.[7]

The Difference-in-Differences (DID) method is often recommended for its practicality. It addresses selection bias that simpler comparisons miss, while being more broadly applicable than RDD or IV. To ensure a quality DID evaluation, synchronize it with the program rollout and carefully select a comparison group.[7]

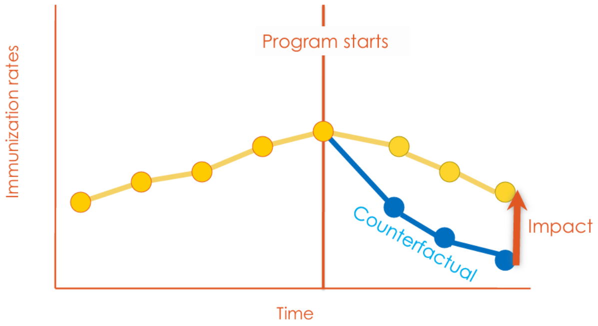
Randomized controlled trials (RCTs) hold particular promise for pilot projects, informing decisions about large-scale implementation. A well-designed RCT effectively combines policy interest and economic theory for a robust experiment.[7]

2.3 Counterfactual

A key aspect of the Impact Evaluation approach described above is the construction of counterfactuals. Counterfactuals represent what would have happened without the intervention being evaluated. By comparing the actual outcomes with these counterfactuals, we can assess the true impact of a decision.The figure 1 is from J-PAL Global[8].



(a) Scenario I



(b) Scenario II

Fig. 1 Program Impact and Counterfactual

Figure 1(a) illustrates a scenario where, despite a seemingly positive trend, the actual outcomes would have been even better without the implemented decision. This emphasizes the importance of counterfactuals in informing decision-making. We can avoid potentially counterproductive actions by comparing them to what might have happened otherwise.

Similar logic applies to Figure 1(b), which showcases a different counterfactual scenario. By analyzing counterfactuals, we gain valuable insights to guide effective decision-making in the context of production optimization.

2.4 Industrial IoT

The Industrial Internet of Things (IIoT) is revolutionizing manufacturing by enabling remote monitoring, intelligent analytics, and control of industrial processes.[9] This surge in connectivity has led to the generation of massive amounts of data, offering a wealth of insights into production.[10] IIoT data provides fine-grained details about individual events on the factory floor, allowing us to grasp the real-time operational status of equipment and model the production situation from multiple dimensions.

2.5 Single-Sample Analysis

A major challenge in evaluating decisions within manufacturing firms is the lack of opportunity for randomized controlled trials. Unlike other research settings, we typically only have a single production line or factory to analyze, making it difficult to isolate the true impact of a specific decision. Traditionally, this has required finding a homogeneous control group—a group comparable in every way except for the decision being evaluated. However, identifying such groups in real-world manufacturing settings is often impractical.

Fortunately, the emergence of the Industrial Internet of Things (IIoT) offers a powerful solution. By enabling real-time monitoring of production processes, IIoT generates rich data about equipment performance and product quality. This data can be leveraged to construct simulated control groups—models that represent what would have happened if the decision had not been implemented. This allows us to assess the true impact of the decision on production outcomes.

3. Methodology

3.1 Evaluation Framework

J-PAL Global's Figure 2 illustrates a framework for impact evaluation using Randomized Controlled Trials (RCTs). This framework outlines the key steps involved:

(1) Clearly define the issue being studied, often through a Theory of Change.
(2) Based on this understanding, select a specific outcome of the decision to be evaluated.
(3) Design and implement a randomized experiment to test the intervention.
(4) Analyze the data on the chosen outcome and summarize the conclusions.

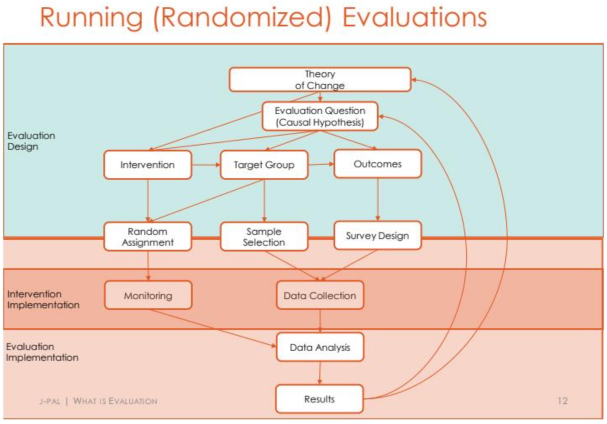


Fig. 2 Running Randomized Evaluations

Manufacturing Execution Evaluation (MEE) shares the core principles of impact evaluation outlined above. However, a key challenge arises in applying the experimental methodology of Randomized Controlled Trials (RCTs) within a manufacturing company. Unlike research settings, MEE focuses on evaluating decisions and optimizations implemented within the enterprise itself. Conducting pilot tests on a single production line or a few similar lines within the company is often the only practical option. Unfortunately, data collected from such internal pilots typically lacks the statistical power needed to ensure valid and generalizable results.

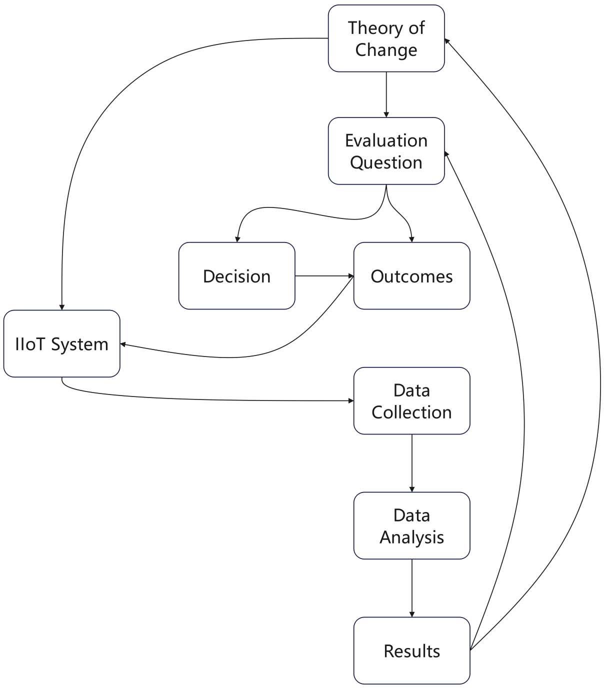


Fig. 3 Manufacturing Execution Evaluation

The limitations of traditional impact evaluation, particularly the reliance on limited data from internal pilots, hinder the ability to draw statistically significant conclusions. To address these challenges, we have enhanced the basic MEE framework in figure 3. A key improvement involves the introduction of an Industrial Internet of Things (IIoT) system. The IIoT system serves a dual purpose: first, it provides a robust infrastructure for collecting real-time field data from production processes. Second, it acts as a supervisory role, overseeing and guiding the entire MEE implementation process.

3.2 Industrial IoT as Intervention

A key consideration when implementing IIoT for MEE is the potential influence of the system itself on the data collected. Since the implementation of IIoT can often be part of the intervention being evaluated, it's crucial to account for its impact on worker behavior.

In practice, IIoT systems can introduce a ”chilling effect.”[11] By recording equipment data and user actions, they essentially function as a monitoring system. This newfound awareness might lead employees to subconsciously alter their behavior[12], operating more carefully than usual. This could lead to an overestimation of the positive impact of certain decisions.

To minimize the influence of the IIoT system itself on MEE results, we can leverage several strategies. One approach involves leveraging existing production data. If possible, it's beneficial to retain or deliberately collect a baseline of equipment operation and employee action data before introducing the IIoT system. This allows for a more accurate comparison when evaluating the impact of decisions.

Alternatively, if historical data is unavailable, we can use the production line's operation for a period after the IIoT system is implemented as a new baseline. This approach acknowledges that the 'chilling effect' might be present initially, but allows the data to eventually reflect the true impact of the decisions being evaluated.

3.3 Technical Formulation of the Manufacturing Execution Evaluation

The Rubin Causal Model (RCM) offers a valuable framework for evaluating cause and effect. However, it presents limitations in assessing the impact of our implemented optimization decision on the manufacturing firm's performance.

The key challenge lies in the single-sample nature of the analysis. RCM typically relies on comparing a treatment group (receiving the intervention) with a control group (not receiving the intervention). Since we cannot observe the hypothetical state of the firm without the optimization (the counterfactual), definitively isolating the causal effect using the traditional RCM approach becomes difficult.

Single-sample analysis limitations: Missing counterfactual. The notation clarifies the single-sample analysis challenge:

- Y represents a random variable for a key firm metric.
- T denotes the intervention status ($T = 1$ for firms implementing the optimization, $T = 0$ for those that don't).

However, since all firms are in the treatment group ($T = 1$), we can only observe $Y(T = 1)$, the outcome with the optimization implemented. We cannot observe the counterfactual outcome, $Y(T = 0)$, which is the firm's performance without the optimization. This missing control group makes it difficult to definitively isolate the impact of the optimization on the observed metric using traditional methods.

We model this indicator Y with a linear regression of covariates X, resulting in the following two expressions:

$$Y(0) = \beta_0 * X + u_0 \tag{1}$$
$$Y(1) = \beta_1 * X + u_1 \tag{2}$$

The residual term, denoted by u, represents the unobserved portion of the outcome variable Y. In a traditional setting, u_1 and u_0 (residuals for treatment and control groups) can be viewed as the idiosyncratic benefit specific to each firm. Additionally, the expected values of these residuals ($E(u_0|X)$ and $E(u_1|X)$) are typically assumed to be zero.

However, our approach using IIoT data to construct a counterfactual introduces a challenge. The implementation of IIoT itself can be considered an intervention that might influence the firm's performance. This invalidates the assumption that $E(u_0|X)$ and $E(u_1|X)$ are equal to zero, as the intervention (IIoT) potentially affects both groups (firms with and without optimization).

We hope that Treatment Effect (TE) should be:

$$\begin{aligned} TE &= E[Y(1) - Y(0)|X] \\ &= X(\beta_1 - \beta_0) + E(u_1 - u_0|X) \end{aligned} \tag{3}$$

Since $E(u_1 - u_0|X)$ is equal to zero, we need to do additional processing on it.

3.3.1 Regression Model

The first approach involves performing a linear regression on the residuals (u) alone. In this approach, we treat the residuals (u) as a random variable and estimate the following equation:

$$u(0) = \gamma_0 * Z + v_0 \tag{4}$$
$$u(1) = \gamma_1 * Z + v_1 \tag{5}$$

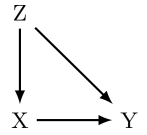
This time here we do have $E(v_0|Z) = E(v_1|Z) = 0$. So:

$$\begin{aligned} TE_u &= E[u(1) - u(0)|Z] \\ &= Z(\gamma_1 - \gamma_0) + E(v_1 - v_0|Z) \\ &= Z(\gamma_1 - \gamma_0) \end{aligned} \tag{6}$$

Here we plug equation 6 into 3, we get:

$$TE = X(\beta_1 - \beta_0) + Z(\gamma_1 - \gamma_0) \tag{7}$$

However, the challenge arises because Z is an endogenous variable, meaning it likely correlates with the unobserved factors captured in the residuals. This endogeneity issue can bias our estimates. To address this, we can introduce an instrumental variable that is correlated with Z but not directly correlated with the error term. It just like:



Our analysis faces a challenge because Z, the variable mediating between X and Y, is endogenous. This means it likely correlates with the unobserved factors captured in the residuals. To address this endogeneity and obtain more reliable estimates, we can consider two approaches:

Shifting the Evaluation Baseline: In a practical approach, we can acknowledge that the heterogeneity in the residuals likely stems from the introduction of the IIoT system itself. To address this, we could re-evaluate the impact of the optimization decision by setting the baseline after the IIoT system was implemented. This approach eliminates the confounding effect of the IIoT on the residuals.

Directly Measuring Z's Effect: Alternatively, we could attempt to directly measure the independent effects of Z on both X and Y. This would involve finding a valid instrumental variable that is correlated with Z but not directly correlated with the error term. Estimating the causal effect of Z on X and Y separately would allow us to potentially account for its influence on the relationship between X and Y.

In practice, we can potentially utilize data from various sensors within the IIoT system to estimate the independent effects of Z on both X and Y. For instance, data from a separate camera system could be used as an instrumental variable to estimate the direct impact of Z on Y. This camera data could potentially be correlated with the mediating variable Z (e.g., capturing aspects of production flow) but not directly correlated with the error term in our model (e.g., random fluctuations in Y unrelated to the intervention or camera system).

3.3.2 Digital Twins

A digital twin (DT) serves as a dynamic virtual counterpart of a physical entity (product, system, process, city, etc.). Continuously synchronized with data from its real-world counterpart and its surroundings, the digital twin bridges the gap between the physical and virtual worlds. This technology empowers data-driven decision-making and real-time optimization, positioning itself as a cornerstone of Industry 4.0 and a key driver of future innovation.[13]

Digital twin technology holds promise for exploring counterfactuals in manufacturing, building on its established applications in the consumer sector[14]. While directly modeling equipment or production processes allows for simulating scenarios and exploring "what-if" possibilities, it's important to acknowledge that obtaining perfect counterfactuals through digital twins can be challenging.

In theory, the digital twin could be used to simulate the production process without the intervention ($Y(0)$), allowing for a direct comparison with the actual post-intervention outcome ($Y(1)$). This comparison could serve two purposes:

- Model Validation:
  Discrepancies between the simulated Y0 and the actual pre-intervention data can help assess the accuracy of the digital twin model.
- Counterfactual Estimation:
  If the model is sufficiently accurate, the simulated Y0 value could be considered an estimate of the counterfactual outcome – the performance without the intervention.

However, it's important to acknowledge that directly obtaining a perfect counterfactual through a digital twin can be challenging. The effectiveness of this approach depends on several factors:

- Model Completeness:
  The digital twin must accurately capture all relevant aspects of the production equipment and process. Any gaps in the model could lead to unreliable simulations of $Y(0)$.
- Data Availability:
  High-quality data is crucial for training and validating the digital twin model. Limited or inaccurate data can result in misleading counterfactuals.

4. Case Study

4.1 Foundry in Shandong, China (2020-2021)

A lack of historical data is a common challenge in manufacturing plants. This case study explores how a foundry in Shandong, China, implemented a two-phase approach to production optimization using the Manufacturing Execution Evaluation (MEE) framework.

4.1.1 Phase 1: Establishing a Baseline with IIoT

The foundry, specializing in pipe castings, lacked any prior data on equipment performance or production processes. To address this challenge, the first phase focused on establishing a baseline for evaluation.

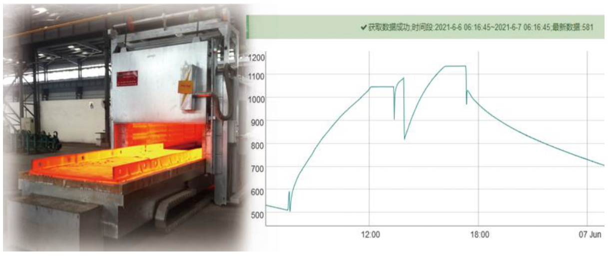
An Industrial Internet of Things (IIoT) system was implemented to collect real-time data from equipment and production lines. This data collection infrastructure formed the foundation for future analysis.

4.1.2 Phase 2: Optimizing Production and Evaluating Impact

With the IIoT system in place, the second phase concentrated on optimizing production processes. This likely involved adjustments to production scheduling, equipment maintenance, or other relevant factors.

Statistical data collected through the IIoT system was used to evaluate the effectiveness of the implemented optimizations. This allowed the foundry to quantify the impact of their decisions on production outcomes.

(a) Operation Data of the Roasting Furnace



(b) Operation Data of the Heat Treatment Furnace

Fig. 4 Operation Data Example of Main Production Equipments

4.1.3 Key Takeaways:

This case study demonstrates the value of the MEE framework, particularly for companies lacking historical production data.

By strategically leveraging IIoT, the foundry was able to establish a baseline and subsequently measure the impact of their optimization efforts.

Table 1 Optimizing Decision Comparison

Metric	Description	Before	After
Raw Material Cost	%, less is better	100	44
Repair Cost of Furnace	%, less is better	100	74
Power Consumption	%, less is better	100	91

Table 1 presents compelling quantitative evidence of the implemented optimization measure's effectiveness. It highlights significant cost savings, particularly a remarkable reduction of over 50% in raw materials. Additionally, the table showcases considerable improvements in production efficiency through reduced maintenance costs and power consumption of the equipment.

4.2 Precision Casting in Jiangsu, China (2019)

This case study demonstrates how a precision casting company in Jiangsu, China, addressed a low yield rate challenge through a data-driven approach using the MEE framework.

4.2.1 Limited Data and Low Yield

The company, facing a concerning product yield rate of only 33%, lacked comprehensive on-site equipment data. They did, however, have some historical production data, including production statistics, yield rates, and unit processing costs.

Improving the yield rate became the primary focus for the MEE project.

4.2.2 Phase 1: Utilizing IIoT to Identify Root Causes

In the absence of detailed equipment data, the first phase involved installing an IIoT system. This system captured real-time data on equipment operation and personnel actions.

After a month of collecting data, analysis revealed a critical issue: poor coordination in production rhythm between the melting and casting processes. This lack of synchronization was likely contributing to the low yield rate.

4.2.3 Phase 2: Optimizing Production Rhythm and Achieving Success

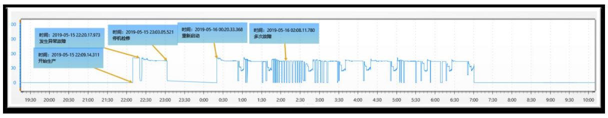
Based on the insights gained from the IIoT data, the consultant team provided specific optimization recommendations. These focused on adjusting the working times and synchronizing the production rhythm between the electric furnace and the molding line.

The second phase involved continuous improvement over six months. By implementing the recommended adjustments, the company achieved a remarkable improvement in its product yield rate, increasing from 33% to a significant 79%.

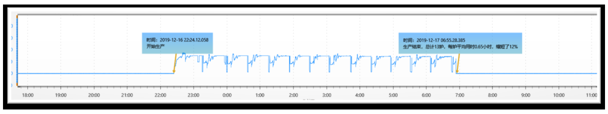
4.2.4 Figures 5: Contrasting Melting Furnace Operations

These figures illustrate the significant impact of production rhythm optimization on the electric furnace's melting process.

- Figure 5(a): Prior to Optimization (Chaotic Melting)
  Figure 5(a) showcases the electric furnace operation before optimization. We can observe a pattern of inconsistent melting times for the molten iron. This lack of consistency likely resulted in:
  1. Insufficient heating:
     Molten iron might be removed prematurely, leading to unstable product quality.
  2. Idle equipment:
     The furnace exhibits periods of inactivity, potentially limiting production output.
- Figure 5(b): Post-Optimization (Stable Melting)
  Figure 5(b) depicts the electric furnace operation after implementing the optimized production rhythm. The data reveals a marked improvement:
  1. Consistent melting times:
     The melting process exhibits greater stability, ensuring the iron reaches the desired temperature for consistent quality.
  2. Reduced idle time:
     The furnace operates more efficiently, minimizing downtime and potentially increasing production output.
- Overall Impact:
  By optimizing production rhythm based on IIoT data, the company achieved significant benefits. These include improved product quality due to consistent iron temperature and potentially higher production output due to reduced idle time in the furnace.



(a) Operation Data Before Intervention



(b) Operation Data After Intervention

Fig. 5 Operation Data Example of the Electric Furnace

4.2.5 Key Takeaways:

This case study exemplifies the power of MEE in identifying production bottlenecks, even with limited historical data.

The strategic use of IIoT data collection played a crucial role in pinpointing the root cause of the low yield rate.

By optimizing production rhythm through data-driven insights, the company achieved a substantial increase in product yield.

Table 2 Optimizing Decision Comparison

Metric	Description	Before	After
Production Capacity	batch, higher is better	10	13
Yield	%, higher is better	33	79
Unit Cost	%, less is better	100	50

Table 2 reinforces the remarkable results achieved through production rhythm optimization.As previously highlighted, the product yield rate experienced a significant increase, jumping from 33% to a noteworthy 79%.

This substantial improvement in yield directly translates to cost savings. The table likely shows a decrease in the cost per unit (e.g., cost per piece) by half.

5. Discussion and Conclusion

This paper proposes a basic framework for evaluating manufacturing execution, drawing on best practices from impact evaluation. Evaluating a specific decision or optimization solution within a single manufacturing firm presents a unique challenge due to the limited sample size (single-sample analysis).

However, the capabilities of the Industrial Internet of Things (IIoT) for data collection and field sensing offer potential to enhance the reliability of single-sample assessments. While the IIoT system's monitoring abilities can be seen as an intervention, we present solutions based on statistical models to address this challenge.

Furthermore, we explore the potential issues encountered in real-world Manufacturing Execution Evaluation (MEE) through two case studies. By examining the improvements and benefits gained by the involved organizations, we aim to validate the effectiveness of MEE as a valuable tool.

References

[1] Wikipedia. Impact evaluation — wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Impact_evaluation, 2024. [Online; accessed 11-June-2024].

[2] Abhijit Banerjee, Esther Duflo, Rachel Glennerster, and Cynthia Kinnan. The miracle of microfinance? evidence from a randomized evaluation. *American economic journal: Applied economics*, 7(1):22–53, 2015.

[3] Bruno Crépon, Florencia Devoto, Esther Duflo, and William Pariente. Impact of microcredit in rural areas of morocco: Evidence from a randomized evaluation. Technical report, Citeseer, 2011.

[4] Patrik Jonsson and Magnus Lesshammar. Evaluation and improvement of manufacturing performance measurement systems-the role of oee. *International Journal of Operations & Production Management*, 19(1):55–78, 1999.

[5] AJ De Ron and JE Rooda. Oee and equipment effectiveness: an evaluation. *International Journal of Production Research*, 44(23):4987–5003, 2006.

[6] Paul J Gertler, Sebastian Martinez, Patrick Premand, Laura B Rawlings, and Christel MJ Vermeersch. *Impact evaluation in practice*. World Bank Publications, 2016.

[7] Asian Development Bank. *A Review of Recent Developments in Impact Evaluation*. Asian Development Bank, 2011.

[8] Ben Morse. Why randomize? - lecture & case study. https://www.povertyactionlab.org/sites/default/files/Day2_Lecture_CaseStudy.pdf, 2020. [Online; accessed 11-June-2024].

[9] Jiangfeng Cheng, Weihai Chen, Fei Tao, and Chun-Liang Lin. Industrial iot in 5g environment towards smart manufacturing. *Journal of Industrial Information Integration*, 10:10–19, 2018.

[10] Akseer Ali Mirani, Gustavo Velasco-Hernandez, Anshul Awasthi, and Joseph Walsh. Key challenges and emerging technologies in industrial iot architectures: A review. *Sensors*, 22(15):5836, 2022.

[11] Alex Marthews and Catherine E Tucker. The impact of online surveillance on behavior. *Cambridge handbook of surveillance law*, 2017.

[12] Shane Dawson. The impact of institutional surveillance technologies on student behaviour. *Surveillance & Society*, 4(1/2):69–84, 2006.

[13] Yuchen Jiang, Shen Yin, Kuan Li, Hao Luo, and Okyay Kaynak. Industrial applications of digital twins. *Philosophical Transactions of the Royal Society A*, 379(2207):20200360, 2021.

[14] Samuel Levy. Digital twins: A generative approach for counterfactual customer analytics. 2023.