



OPEN

A new perspective on Workload Control by measuring operating performances through an economic valorization

Davide Mezzogori¹, Giovanni Romagnoli²✉ & Francesco Zammori¹

Workload Control (WLC) is a production planning and control system conceived to reduce queuing times of job-shop systems, and to offer a solution to the lead time syndrome; a critical issue that often bewilders make-to-order manufacturers. Nowadays, advantages of WLC are unanimously acknowledged, but real successful stories are still limited. This paper starts from the lack of a consistent way to assess performance of WLC, an important burden for its acceptance in the industry. As researchers often put more focus on the performance measures that better confirm their hypotheses, many measures, related to different WLC features, have emerged over years. However, this excess of measures may even mislead practitioners, in the evaluation of alternative production planning and control systems. To close this gap, we propose quantifying the main benefit of WLC in economic terms, as this is the easiest, and probably only way, to compare different and even conflicting performance measures. Costs and incomes are identified and used to develop an overall economic measure that can be used to evaluate, or even to fine tune, the operating features of WLC. The quality of our approach is finally demonstrated via simulation, considering the 6-machines job-shop scenario typically adopted as benchmark in technical literature.

Nowadays, more than 40 years after its original conception, Workload Control (WLC) has been deeply investigated¹ and its manifold benefits are unanimously acknowledged by the scientific community². Nonetheless, although they have increased over the last decade, successful industrial implementations of WLC are still limited³⁻⁵. A possible explanation of such a limited industrial consideration could be the complexity and the heavy informational requirements of WLC, which are both higher than those of other ‘paper and pencil’ lean-techniques, such as dual Kanban or Constant Work In Progress-ConWIP⁶. Another hurdle to WLC acceptance can be found in^{7,8}, where the large gap between theoretical contributions and industrial applications is clearly discussed and exemplified. Both authors noted that most of the studies are carried on using over-simplified event-based simulation models, which do not account for real-life constraints. Perfectly balanced systems with a rather low utilization level (typically set at 90%) are generally considered, while additional operative issues such as handling times, limited buffer areas and parts-feeding requirements are neglected⁹. Although the spectrum of the business cases is much wider than that¹⁰, only a few researchers focused on unbalanced manufacturing systems (see for example¹¹) and tried to investigate the applicability of WLC to these system too^{8,12}.

A last and more subtle problem has been identified by⁶, who noted that in scientific literature WLC performance is generally measured using time-related measures, such as the percentage of tardy jobs or the like. There is no doubt that the capability to set and respect short delivery dates in response to customers’ enquiries is a primary market requirement for Make-To-Order (MTO) manufacturers^{13,14}. Yet, focusing exclusively on this measure is somehow myopic because, if WLC is regulated by simple operating rules, the percentage of on time deliveries does not deviate much from that of a purely push operating system, with optimized dispatching rules^{15,16}. This is a strong disincentive, as manufacturers do not see the point of adopting a complex production planning and control (PPC) system, if similar results could be achieved with a much simpler MRP based system, provided of adequate capacity control and scheduling mechanisms^{17,18}. To get rid of this misconception, the research community should extend the analysis beyond ‘time related’ performance measures, and should try to quantify other WLC related benefits, such as WIP reduction and containment. WIP, in fact, is the worst cause of

¹Department of Engineering “Enzo Ferrari”, University of Modena and Reggio Emilia, Via P. Vivarelli, 10, 41125 Modena, MO, Italy. ²Department of Engineering and Architecture, University of Parma, Viale delle Scienze, 181/A, 43124 Parma, PR, Italy. ✉email: giovanni.romagnoli@unipr.it

waste of a lean-oriented manufacturing systems and its reduction, beside containing the shop floor throughput time, would also cut holding costs, free up space, improve order and tidiness, and favor visual management and control¹⁹. As noted by⁴, performance measures related to these elements of competitive edge exist, but they are seldom used to assess performance of WLC systems. Also, these are unstructured approaches that focus on a single and specific feature of a WLC systems, so that the overall view, and the overall benefits are lost anyway.

For this reason, an interesting branch of research has developed over the last decade, with the aim of evaluating WLC performances by means of economic values, that is costs and/or revenues^{20–22}. Although this research topic is also reflected in the wider literature on weighted earliness and tardiness (see for example²³), we believe that an economic evaluation in monetary terms is the main way to merge and combine different benefits and costs. To this aim, all the relevant costs (of variable nature) and incomes of a specific WLC configuration might be quantified, confronted, and discussed, to assess and compare the expected economic benefits of the WLC implementation.

However, the mentioned studies on economic assessment of WLC performances present some important limitations from our point of view. First, the cost system they propose are specific of the simulated case^{22,24,25}, and often industry-specific and related to semiconductor manufacturing^{20,26–28}. Second, although the cost factors, or clearing functions, have been reported in previous research, we note that these methods were seldom adopted to the so-called research field of rule-based WLC (see^{29,30} for an overview on rule-based and optimization based WLC).

Indeed, we believe that such a *system*, as well as the cost function, should be better generalized and detailed, especially in the rule-based WLC field. We also note that previous research mostly focus on costs, whereas we believe that incomes should also be considered to provide a better performance indicator. Finally, in the authors' opinion, previous research works do not clearly report how an economic indicator could be used to identify and fine-tune a specific WLC configuration, and it could also provide to companies, and especially SMEs, a method to assess economic benefits to be expected from a WLC implementation.

To overcome these limitations, the present paper incorporates a general method for economic evaluation in WLC research, by proposing a Revenue per job function that could be used to select and fine-tune the best performing WLC configuration in given circumstances. This function is also applied to a simulated 6-machines job-shop scenario, which is typically used as a benchmark in technical literature, and supported by a sensitivity analysis, so as to show the benefits of our approach in a common scenario, and to assess how these benefits might change under different ratios of cost parameters. By doing this, we suggest a possible solution to the problem identified by⁶, i.e. we provide a general method, as well as a sample implementation and a sensitivity analysis to identify a trade-off between different performances and metrics.

The remainder of the paper is organized as follows. In “[An overview on WLC research](#)” and “[Performance assessment of a WLC system](#)” sections we provide a brief overview of WLC research, also including studies on performance assessment of WLC systems. Moving from these premises, the economic assessment model is developed in “[A proposal for an economic evaluation of WLC systems](#)”, where all the cost and revenues voices are discussed in detail. The approach is finally tested via discrete event simulation in “[Performance assessment of a WLC system](#)” and “[Performance assessment in monetary terms](#)” sections that describe, respectively, the simulated environment and the obtained outcomes. Comments, managerial insights, and conclusions are finally provided in “[Conclusions, managerial implications, and future works](#)”.

An overview on WLC research

A major guiding principle of WLC is input/output control³¹, asserting that the input and the output rate of a shop should equal. Scientific literature, however, presents several WLC methods, whose unifying theme is use of a pre-shop pool and order release mechanism¹. In this paper, we refer to the so-called LUMS WLC (see for more details³²), a rule-based approach to WLC consisting of three (or four) control levels, namely the (i) customer enquiry and order entry, which are sometimes separated in two different stages; (ii) order release and (iii) priority dispatching. At each level, two methods of control are available: input and output control³³. The former regulates the input flow of orders to the next stage, and the latter controls workload by regulating the outward flow, e.g. by setting due dates and adjusting capacity. In this regard, we note that WLC is partially aligned with the stream of research on Value-Chain Flexibility^{34–36}. According to this framework, in fact, to answer to the environmental uncertainty that is typical of job shops oriented manufacturers, and which has significantly increased in the post-pandemic era, organizations should exploit four different types of flexibility, namely: (i) product development; (ii) manufacturing; (iii) logistics and (iv) spanning flexibility³⁷.

The alignment between these two concepts lies in the fact that WLC could significantly impact on manufacturing flexibility, and especially on machine and routing flexibility (see³⁷), and it could also provide some positive outcomes on other types of flexibility.

To improve flexibility and stabilize lead times in a context of environmental uncertainty, WLC uses workload caps or norms, generally expressed as number of expected working hours, to regulate the release of new production orders. More precisely, work in process is constantly monitored and new jobs are not released in the shop floor, unless all norms are fully met. If one or more norms are violated, jobs are stopped into a Pre-Shop-Pool (PSP) of pending orders, waiting to be reconsidered for possible future release. Due to this pattern, the time a job spends in the PSP, namely the PSP_{time} , may not be negligible relatively to its Shop Floor Throughput Time (SFTT), i.e., total time from order release to dispatching. Hence, in a WLC system, these two quantities are always considered together as they determine the Gross Throughput Time of a job $GTT = (PSP_{time} + SFTT)$.

When a customer places an order (or makes an order enquiry), the manufacturer must decide whether to accept the order or not. Only profitable and technically feasible jobs should enter the PSP. Since this evaluation step is standard and typical of many manufacturing system, it is rarely considered by the WLC research

community. A more interesting decision, which has sparked much more research interest, concerns the customer enquiry stage and the assignment of a due date to the accepted jobs³⁸. Such choice, which often requires a bargain process with the customer, may also impact on the acceptance decision because, if the required due date is too tight, the manufacturer could incur into penalties, and could even jeopardize the respect of the overall production plan, by trying to preempt a critical job, to speed up its delivery. Hence, customer enquiry stage and order acceptance are generally treated in parallel.

The job release phase is the crucial part of a WLC system. The adopted technique for selecting and releasing orders from the PSP to the shop floor can make the difference between a successful and unsuccessful WLC implementation³. Specifically, all jobs pending in the PSP are periodically considered for possible release to the shop floor. This task is performed in sequence, by considering one job at a time, from the first to the last one in the PSP. In the PSP, indeed, jobs are sorted in order of priority, typically by means of a dispatching rule, which is usually the same used to sort jobs in the shop floor. Release depends both on the workload of the considered job and on the current system's state (expressed as the level of the norms), as detailed below.

Let $w_{j,m}$ be the workload of job j on machine m and let \tilde{j} be one of the job in the PSP. When, at time t , \tilde{j} is selected for possible release, its workload $w_{\tilde{j},m}$ is used to update the cumulative workload released to the system (W_t) and/or that released to a specific machine ($W_{m,t}$).

The update is done as in Eqs. (1) and (2), whereas the cumulative workloads are computed as in Eqs. (3) and (4), respectively.

$$W_{m,t^+} = W_{m,t} + w_{\tilde{j},m} \quad (1)$$

$$W_{t^+} = W_t + \sum_m w_{\tilde{j},m} \quad (2)$$

$$W_{m,t} = \sum_{j \in J_{(S,m)}} w_{j,m} \quad (3)$$

$$W_t = \sum_m \sum_{j \in J_{(S,m)}} w_{j,m} = \sum_m W_m \quad (4)$$

where t^+ is the time immediately after job \tilde{j} has been selected for possible release, $J_{(S,m)}$ is the subset of jobs released in the shop floor, which still need to visit machine m , $w_{j,m}$ equals zero for all machines m that does not lay on the routing of job \tilde{j} .

It is worth noting that these equations are used at different times. Whereas Eqs. (1) and (2) are related to the selection process, Eqs. (3) and (4) are triggered by the manufacturing processes in the shop floor. Equations (1) and (2), in fact, are evaluated any time a job of the PSP is select for possible release, whereas Eqs. (3) and (4) are applied every time a job in the shop floor moves from a machine to the next one.

We also note that, if a single norm is used to constrain the cumulative workload of the system W_t , the release phase is named 'total shop load' control. Conversely, if multiple norms are used to limit the cumulative workload $W_{m,t}$ of individual machines, the release phase is labelled as 'bottleneck load' or 'load at each machine', if control is limited to the supposed bottlenecks or extended to all machines, respectively (for more details, see³⁹⁻⁴¹).

Anyhow, the updated values W_{t^+} or W_{m,t^+} are compared to the system's norms and the job \tilde{j} is released to the shop floor, provided that none of the norms is violated. If not so, \tilde{j} remains in the PSP and the process is iterated until the whole list of jobs in the PSP has been completed.

Relatively to the release phase, another crucial decision regards the quantification of the individual workloads $w_{j,m}$ of the entering jobs. Generally speaking, the workload released to a certain machine m can be split into a direct and an indirect part. The first part is related to the jobs that are in the queue of machine m ; this part is considered 'direct' because these jobs are currently using the productive capacity of machine m . The second part, instead, is due to the jobs that are still upstream of machine m and that will visit this machine later on. This part of the workload is considered 'indirect' because these jobs will use the productive capacity of machine m only in a later moment. Hence, the workload $w_{j,m}$ should be fully included in W_m if job j is queuing at machine m and should be partially included in W_m if job j is still upstream of m . In other words, $w_{j,m}$ should be rescaled (or reduced) to consider the time lag that will elapse before j reaches m : the more m is downstream, the fewer load should be attributed to it⁴². How to properly rescale the workload $w_{j,m}$ is a highly researched topic. A first attempt, known as the Load Oriented Order Release (LOOR) approach, was proposed by⁴³, who dynamically rescales $w_{j,m}$ using a depreciation factor obtained through a statistical (mainly regression based) analysis. Other probabilistic approaches have been proposed next as in^{44,45}. Also, a specific methodology was developed at Lancaster University Management School, also referred to as LUMS methodology, designed for the specific needs of MTO companies and concentrated upon the Aggregate Load field of WLC. However, despite the research interest that they roused, these approaches have seldom found their way into practice and nowadays they are increasingly less used in the literature⁴⁶. As⁴⁷ pointed out, in fact, sophisticated methods are too complex for a straightforward application in the industry and, consequently, they have been misused for lack of understanding and neglected over time. To solve this operational problem, other authors (see for example⁴⁸) simplified the problem by considering $w_{j,m}$ as a fixed value equal to the sum of the processing and set up time of job j on machine m . This approach is known as the 'aggregate approach', since the direct and indirect part of the workload are not differentiated and $w_{j,m}$ is never rescaled, not even if job j is queuing at a machine upstream of m . A more refined alternative, known as the 'corrected aggregate approach', is frequently applied too.

Category	Measures	Sample references
1. Time-related	Lead time	51
	Shop floor throughput time	52
	Supply chain time	53
2. Dependability	Lateness	54
	Percentage of tardy jobs	15
	Service level	25
	Tardiness	32
3. Shop load-related	Number of jobs in queue	55
	Utilization rate	56
	WIP level	8
	Workload balance	57
4. Market-related	Percentage of accepted orders or 'strike rate'	58
	Volume of accepted orders	59
5. Cost-related	Inventory level (WIP and Finished goods)	30
	Production yield	58
	Comprehensive cost measure	22,30

Table 1. Performance measures used in recent scientific works.

According to this approach, introduced by³³, the workload is rescaled as in Eq. (5):

$$\tilde{w}_{j,m} = \frac{w_{j,m}}{n_{j,m}} \quad (5)$$

As Eq. (5) shows, the rescaling rule is very simple, as the corrected load $\tilde{w}_{j,m}$ is obtained using as scaling factor the position $n_{j,m}$ of machine m in the routing of job j . Like the 'aggregated approach', the load is rescaled as soon as job j is released and, next, the value of $\tilde{w}_{j,m}$ will no longer be changed. The use of the corrected aggregate approach is generally preferable, as it assures high performance, and it also simplifies the fine-tuning of the norms, as described in greater details by^{15,41}.

Finally, jobs in the shop floor are managed as in a standard batch and queue system, i.e. they are pushed from a machine to the next one. For this reason WLC is generally considered a hybrid push–pull technique. Hence, when more jobs compete for the same machine, WLC uses some predefined dispatching rule. However, if norms have been properly fine-tuned, queues tend to be short and stable and so jobs dispatching become a less important issue that can be simply treated using one of the manyfold dispatching rules proposed in the literature⁴⁹. For a comprehensive discussion of the most common dispatching rules we refer the interested reader to⁵⁰.

Performance assessment of a WLC system

As already mentioned, WLC assessment is frequently carried on using tardiness and/or percentage of tardy jobs as performance indicator. However, some examples of alternative measures can be found in the literature, as exemplified in Table 1, which shows the most common measures grouped in the broad categories suggested by⁴. We note that, with respect to⁴, the internal co-ordination category is not reported. To the best of the authors knowledge, in fact, such measures (e.g. co-ordination between production and marketing) have not been used in the last decade.

Almost all the metrics reported in Table 1 are self-explaining; only some of them deserve some further comments. *Supply chain time* was introduced by⁵³ to measure lead time of a MTO supply chain of highly customizable goods. Also⁵⁸, uses two interesting measures, namely the *Strike rate*, that is the proportion of successful quotations among total enquiries, and the *Production yield*, referred to as the successful rate of the production of a product, or the percentage of jobs that do not generate quality problems, such as reworks or scraps.

It is important to observe that most of the studies makes use of 'time related', 'dependability' and 'shop load-related' measures, with the first two categories being the most widely used⁶. 'Market-related' measures are, indeed, less common: the table reports⁵⁹, for instance, which compared alternative production planning and control systems in terms of the generated volumes (and monetary value) of the accepted orders; and⁵⁸, where the information architecture for WLC is detailed, reporting a case study of a small precision-engineering MTO manufacturer.

Although the set of available measures is rather ample and variegated, a common and standard way to assess real-word performance of WLC and, more in general, of a PPC system, is still missing. Researchers, in fact, are often tempted to report the measures that better confirm their hypotheses, often arguing that those performance measures are the most pertinent in the industrial case under analysis. Consequently⁴, suggested that researchers should "ask themselves whether the same set of measures should be used to assess WLC implementation in practice, in order to determine appropriate measures for their project". Unfortunately, despite this suggestion, many scholars still continue to use very specific measures that might fail to capture the overall benefits of a PPC system, also due to the fact that some of these measures are conflicting (e.g. short lead time versus high service level, or high volume of orders versus small percentages of tardy jobs) a fact that makes extremely hard the

definition of adequate benchmarks. The need, and the relevant practical implications, of a standardized measure is evident, but little has been done so far, toward this goal.

A recent stream of research has indeed started to investigate WLC performances from a more comprehensive point of view. A preliminary single-machine cost function was proposed in 1997 by⁶⁰, by calculating a weighted sum of the WIP and tardiness costs under a periodic order release policy. The first cost function that could be applied to a WLC system, however, can be found in the influential work of Asmundsson et al.⁶¹, proposing the clearing function models, which are deemed as advanced and widely used models to date in terms of optimization-based order release models³⁰. In 2011⁶², jointly considered different and potentially conflicting performance measures, by converting them in monetary terms. The same approach was followed by⁶³, who compared alternative production plans considering three different kind of costs: (i) WIP holding costs, (ii) finished goods holding cost, and (iii) backlog cost. Within this stream of research²⁸, reported an interesting experimental comparison of different iterative approaches and a clearing function, estimated by means of two different approaches, in a simplified model of a wafer fabrication facility. One year after²⁷, implemented two linear programming models for cycle time estimate and presented the computational testing of a production planning model based on clearing functions on a large-scale data set that represents a wafer fabrication facility with two products. The performances of different plans are then compared based on the profit realized by the facility when the execution of the plans is simulated. Reference²⁶ assessed and compared, under a wide range of operating conditions, the performances of a production planning model with fixed lead times and a second model with workload-dependent lead times in a rolling horizon setting.

Another emerging stream of research has made use of adaptive planning system, such as machine learning algorithms and artificial neural networks, to dynamically set lead times, plan capacities and decide on whether to release orders or not^{15,24,25,41}. Finally, starting from the seminal work²⁹, interesting studies have been carried out in the very last years to bring together and compare the rule-based WLC, i.e. order release mechanisms determining the release times of the orders by means of a set of rules, and optimization-based WLC, i.e. multi-period order release mechanisms that use optimization models^{20,22,30}.

While we praise the effort of these authors, we also note that:

- To the best of our knowledge, an economic valorization has seldom been used in rule-based WLC literature;
- Relative to such a WLC scenario, the cost structures proposed are not complete, as they do not account either for production costs or for sales prices, thus missing the huge difference between accepted and non-accepted customer's orders;
- A general method for economic evaluation of WLC systems is still missing in rule-based WLC literature.

The approach to properly evaluate WLC based upon economic considerations is the topic of the following sections.

A proposal for an economic evaluation of WLC systems

Our proposal is to evaluate performances of different WLC systems, by considering incomes and costs items that closely match the classification structure provided in Table 1. All proposed costs and incomes are traditionally used in inventory management theory and can be easily estimated and measured in production systems⁶⁴. If a Manufacturing Execution System is in place, all the inputs of the income/cost equations (provided next), can be easily collected on the field. Should this not be the case, the standard costs and the standard times reported, respectively, in the bill of materials and in the production-cycle can be used as a viable alternative. Similarly, if a simulation model is needed, these costs can be easily embedded in the simulation logic and used to fine tune the norms and to operate the system at the point of maximum economic viability.

Time and shop load related costs. The first category contains two costs items related to the time a job has spent in the shop floor, either being processed or waiting in queue. These are the Transformation (*TR*) and WIP Holding costs (*WH*), respectively. Concerning *TR*, it includes direct materials costs, all direct production costs (sustained on the shop floor) and, eventually, the setup costs, as shown in Eq. (6).

$$TR_j = \text{Direct Materials} + \text{Processing Cost} + \text{Setup Cost} = \mathcal{M}_j + \sum_m \left(t_{j,m} p_m + \frac{\zeta_{j,m} s_m}{b_j} \right) \quad (6)$$

where \mathcal{M}_j is the total cost of the direct materials (or parts, or components) used to manufacture job j , $t_{j,m}$ is the actual processing time of job j on machine m , p_m is the cost of use per unit of time of machine m , which should include depreciation, energy, direct labor and, eventually, auxiliary operating costs, $\zeta_{j,m}$ is the actual setup time of job j on machine m , b_j is the batch size of job j , s_m is the setup cost per unit of time.

If needed, the handling costs related to the distance covered by a job along its route could be included in Eq. (6) too. We decided not to do so, because handling costs (as well as other cost items of logistic nature) depend on the layout of the system but not on the way in which a WLC system has been dimensioned and fine-tuned. The same would not be true if the aim were to compare alternative PPC systems that leverage on specific handling and/or feeding systems such, for example, as a dual Kanban system with milk-run feeding strategy⁶⁵. In this case, indeed, handling costs would change with the considered alternative. Please note that also the transformation costs herein considered do not depend on the adopted WLC system; yet a precise quantification of these costs is essential for a correct valorization of WIP and Stocks that, instead, are highly influenced by the WLC system in use.

Concerning WH , this item accounts for the cost sustained to hold work in process in the shop floor and it is obtained by multiplying the total time in the system, or Shop Floor Throughput Time ($SFTT$), for the unitary holding cost per unit of time h . That is:

$$WH_j = \text{Time in the sistem} \cdot \text{Unit Holding Cost per unit of time} = (SFTT_j \cdot h_j) \quad (7)$$

The holding cost h_j depends on the value of job j and, therefore, it increases as job j moves downstream its manufacturing process. Also, assuming that materials are evenly added throughout the manufacturing process (as for an assembly line) and that the cost of use p_m does not vary much from one machine to another, the increase of h_j can be considered linear. So, the holding cost can be expressed as in Eq. (8):

$$h_j = k_h \cdot (0.5 \cdot TR_j) \quad (8)$$

where k_h is a fixed coefficient lower than one, and $0.5TR_j$ is the average value of job j .

If all materials enter at the beginning of the production cycle, the computation of the holding cost changes as in Eq. (9):

$$h_j = k_h \cdot \left(\mathcal{M}_j + 0.5 \cdot \sum_m \left(t_{j,m} p_m + \frac{c_{j,m} s_m}{b_j} \right) \right) \quad (9)$$

Dependability related costs. This category contains the costs incurred by the company to meet customers' needs in terms of delivery times. Two items are considered, namely the Storage costs (ST) of end goods and the Penalty costs (PN).

The first item, related to earliness, accounts for the fact that orders completed on time must remain in stock for some time before being finally aggregated with other orders and dispatched to the end customer. So, expenses for storage of the manufactured items are incurred. The second item, instead, is related to tardiness and is incurred anytime orders are shipped late. Note that penalties may be explicitly agreed with the customer at the contractual level, but even if this were not the case, an implicit cost should be considered, at least to cover for backlog issues such as: customer compensation for delay, loss of goodwill, expedited shipping, potential loss of future orders, as well as for other indirect costs. Also note that storage and penalty costs can be considered as mutually exclusive, as either one or the other must be null for every order. Indeed, it can be assumed that late orders are never held in stock (or are held for a very short time) but are shipped immediately.

Operatively, akin the holding cost for WIP, also the holding cost of the finished goods are obtained by multiplying the time spent by a job in the finished goods inventory for its monetary value. Since jobs have been completed and are ready to be delivered, their value coincides with the total transformation costs. So, we have that:

$$ST_j = \text{Inventory Time} \cdot \text{Unit Holding Cost per unit of time} = (IT_j \cdot \hat{h}_j) = IT_j h (k_h \cdot TR_j) \quad (10)$$

where IT_j is the total inventory time and the unit holding cost \hat{h}_j is again obtained as a fraction of the total transformation cost TR_j .

Penalties costs are harder to be estimated, and this topic has been intensively studied in the Make-To-Stock sector, whose literature on Out-Of-Stock could provide some interesting suggestions to the reader (see for example^{66,67} for quantitative studies on Out-Of-Stock in the fast-moving consumer goods sector). Yet, neglecting costs of penalties would lead to an inaccurate cost structure. So, our suggestion is to calculate PN by multiplying a unitary penalty cost (or fee) per unit of time, say f , for the total delay D , expressed as the positive difference between the shipping time ST and the due date DD agreed upon with the customer. So, we have:

$$PN_j = \text{Total Delay} \cdot \text{Unit Penalty Cost per unit of time} = (D_j \cdot f_j) = k_p \cdot P_j \cdot \max\{0, (ST_j - DD_j)\} \quad (11)$$

where in analogy with the holding cost, the coefficient k_p is used to express the penalty f as a fraction of the sale price P , per unit of delay time. We note that it could be wise to limit the penalty cost to a maximum value, such as a given percentage of the sales price, to rule out the risk of unrealistically higher penalty costs due to very long delays.

Market related items. For a fair comparison, the generated turnover should be included in the analysis, too. Different PPC systems, in fact, may lead to a different throughput rate and, consequently, to non-negligible changes in the production volumes and turnover. Note that this issue is even more stringent when WLC is considered, as different 'job acceptance' rules may modify not only the productive volumes, but also the productive mix, with strong impact on the overall profitability.

In case of MTO companies, production is activated upon customer's demand and each accepted job is sold at the selling price contractually defined with the customer. So, turnover (TV) can be easily obtained summing up the selling prices of all the jobs delivered in the considered time interval. Instead, in case of simulation, we suggest compute the turnover of job as in Eq. (12):

$$TV_j = TR_j \cdot (1 + \delta) \quad (12)$$

where $\delta > 0$ is a mark-up percentage applied to the total transformation cost.

The mark-up should cover the indirect industrial costs, a quota of the overhead costs and, eventually, it should generate a profit. This percentage is generally available in the cost accounting system of the firm.

Performance assessment of a WLC system

To demonstrate the need to introduce the economic model to assess the quality of a WLC system, we reproduced the High-Variability Low Volumes (HVLV) six-machines job-shop, which is typically used as a benchmark in technical literature⁶⁸. To this aim we used Simpy®, an open-source discrete event simulation package, developed in Python 3.7©.

Job-Shop configuration. Concerning the simulated job-shops:

- All the six machines are equally loaded, with an average utilization rate or 90%;
- They all have constant capacity and full availability;
- Materials and resources are always available; machines can be idle, but starvation never occurs;
- The job shop is ‘pure and randomly routed’. In this regard:
 - Each job is processed by N machines, with N uniformly distributed in the range $[1, 6]$;
 - The order in which machines are visited is random, but half of the generated jobs has a ‘directional routing’, while the other half has a ‘purely random routing’. Specifically, machines are labelled with a progressive number (from 1 to 6) and, in case of directional routing, for each couple of successive machines (i, k) in the routing of job j , k must be greater than i .
 - Rerouting is not allowed, i.e., a machine cannot appear on the routing of job j more than once.
- Processing time $t_{j,m}$ (of each job j on each machine m) follows a Gamma Distribution with shape parameter $k = 2$ and scale parameter $\theta = 0.52$. Hence the average processing time on each machine is $\bar{t} = 1.05$ time units and the coefficient of variation is $CV = 0.707$. This choice is typical to reproduce the processing time variability of HVLV job-shops.
- Set-up times must be performed any time a machine processes a new job, but they are not sequence dependent. Hence, set-up times are not explicitly defined, but simply included in the processing times $t_{j,m}$ generated from the Gamma distribution described above;
- Jobs are generated according to a Poisson process, with exponential distributed inter-arrival times. An arrival rate $\lambda = 1.45$ units of time was chosen to obtain, in a purely push-operating system (i.e., jobs never wait in the PSP), the desired 90% utilization rate, for each machine.

WLC settings. Concerning the WLC used to regulate the job-shop, aiming to investigate how performance and costs change as the operating setting changes, a rather ample set of alternative were tested, as detailed next:

- *Job entry* – All generated jobs are always accepted (i.e., total acceptance rule) and an exogenous due date, imposed by the customer, is assigned to each job upon arrival. The value of the due date is randomly generated following the same logic proposed by ample WLC research. Specifically, as shown in Eq. (13) the due date of job j is obtained adding to the arrival time AT_j the expected total processing time $(\bar{t} \cdot N_j)$ and a uniformly distributed random allowance U .

$$DD_j = AT_j + U + (N_j + n) \cdot \bar{t} \quad (13)$$

where n is a natural number used to include the extra allowance factor $n\bar{t}$ in the expected processing time $N_j\bar{t}$.

- *Job Release*, the following sub setting were used:
 - Jobs in the PSP are considered for possible release every 24 units of time;
 - Norms are defined as upper bound of the workload (expressed in units of expected working time) either using the Total Shop (TSL) load or the load at each machine approach (LEM). So, the decision of releasing a job depends on the value of W_t or of $W_{m,t}$ in the first and second case, respectively. Also, in case of load at each machine the upper bound set on $W_{m,t}$ is the same for all machine m ; a natural choice for a perfectly symmetric job-shop with equivalent and equally loaded machines.
 - The workload contribution of each entering job is quantified either using the aggregated (AGG) or the corrected aggregated (CAGG) approach, computed as in Eqs. (3) and (5), respectively.
- *Job Dispatching*—Two alternative dispatching rules, namely First In First Out (FIFO) and Planned operation Start Time (PST) are used. The first one is extremely easy (jobs simply queue up) and is the one that minimizes work in process; the second one is very effective in minimizing tardiness or other time related performances⁶⁹. Specifically, when the PST rule is in use, jobs are sorted in ascending order of their planned operation start time. The later one is computed as in Eq. (14):

$$PST_j = DD_j - (\tilde{N}_j + n) \cdot \bar{t} \quad (14)$$

Where n is a natural number and $\tilde{N}_j \leq N_j$ is the number of machines that are yet to be visited by job j , so that $\tilde{N}_j\bar{t}$ is the remaining expected processing time. In brief, therefore, PST_j is an estimate of the net time that is still available to complete the job in time.

Concerning the allowance factors of Eqd. (13) and (14), we consider U as uniformly distributed on the interval $[0, 30]$ and we set $n = 2$. Both values were empirically defined via simulation, to maximize the percentage

Parameter	Variable	Description
System's parameters		
Number of Machines	No	6 machines
Routing	No	Random length (1 to 6) with random sequence, no re-entrant loops
Processing time	No	Gamma distributed on all machines, with mean 1.05 and variance 0.54
Setup time	No	Included in the operation time
Machines' utilization	No	90% for all machines
Machines' availability	No	100% for all machines
Job's arrival rate	No	Exponentially distributed, $\lambda = 1.45$ job every time units
Exogenous Due Dates	No	Computed as in Eq. (13), average value of 21, net of the entry time
WLC parameters		
Jobs Acceptance	No	All jobs are assumed to be feasible, the 'total acceptance' criterion is used
Jobs Consideration	No	Discrete timing convention, jobs in the PSP evaluated every 24 units of time
Aggregation of workload measure	Yes	1. Total Shop Load (baseline)
		2. Load at each machine
Workload accounting over time	Yes	1. Aggregated load (baseline)
		2. Corrected aggregated load
Workload control	No	Norms defined as upper bound values
Dispatching rule	Yes	1. FIFO (baseline)
		2. PST

Table 2. Configuration of the simulated Job-shop and WLC settings.

of on time deliveries when the job-shop operates in purely push way and the PST rule is used, provided that the percentage of on time deliveries does not drop below 75% (considered as a realistic value) when the FIFO rule is used instead. For the sake of clarity, the settings used for the Job-Shop and for the WLC system are summarized in Table 2.

In brief, a total of eight alternatives configurations were obtained, by combining 2 aggregations of workload measure, 2 workload accounting systems and 2 dispatching rules, that is 2 different options for each variable parameter that we considered in our simulation (namely Aggregation of workload measure, Workload accounting over time and Dispatching rule).

Hence, we have 3 variables, each of which can assume 2 alternative levels. Also, if we consider the baseline WLC policies (i.e., total shop load, aggregated approach, FIFO rule) as 'low factor levels', and the alternative policies as 'high factor levels' (i.e., load at each machine, corrected aggregated approach, PST rule), the design scheme that we tested corresponds to a 2^3 full factorial design with repetitions.

Performance analysis. For each one of the eight investigated configurations, we recorded the performance observed for different levels of the norms. More precisely, starting from an almost infinite level, corresponding to a purely push operating system, we progressively constrained the system by reducing the norms' level proceeding with step of 1%. This procedure was interrupted as soon as the observed throughput dropped below 1.44 [jobs per time unit], that is the lower end of the 95% confidence interval of the throughput of the unconstrained system. In this way the full domain of the solutions ensuring the maximum throughput is obtained, with the last generating one corresponding to the condition that minimizes work in process and queues. For each investigated point we performed a simulation trial of 10 runs, each of 3650 units of time. A warm-up period of 1200 units of time, enough to reach the steady state of the system, was also included in each simulation run. These simulation parameters are consistent with those generally used in case of job-shop simulations and provide stable results in an adequate amount of time.

An example is given in Fig. 1, where the percentage of jobs delivered in time (performance measure) is plotted on the y axis, versus the norm level plotted on the x axis. Note that the latter one is expressed as a percentage of the unconstrained case, that is a value equal to 100% correspond to an unconstrained system (i.e., norms are never violated, and jobs are always released upon arrival, without pending in the PSP).

A first consideration concerns norm levels, marking a significant difference between FIFO rule, performing much better in terms of on time jobs for low values of the norms, that is a relatively starved system, and PST rule, which overperforms FIFO at higher norm levels (i.e., roughly above 55%).

Figure 1 is also in line with WLC literature to confirm that better on time performances are obtained with the load at each machine approach and with PST dispatching rule to sort jobs both in the pre-shop pool and in the shop floor. However, in this case, a clear maximum does not exist, as the red curves flatten out for high values of the norms, when the WLC system approximates a purely push-operating job shop. In other words, in terms of the percentage of on time orders, the performance of a purely push system based on the PST rule does not deviates much from that of a WLC system. The system, in fact, is undersaturated (90% utilization level) and so, even if jobs are pushed in production, queues do not diverge, and the use of an appropriate rule (such as the PST) is enough to assure a good performance. Also note that the red curves fall very quickly when work in process is reduced moving from right to left, that is when moving away from the purely push condition. Hence,

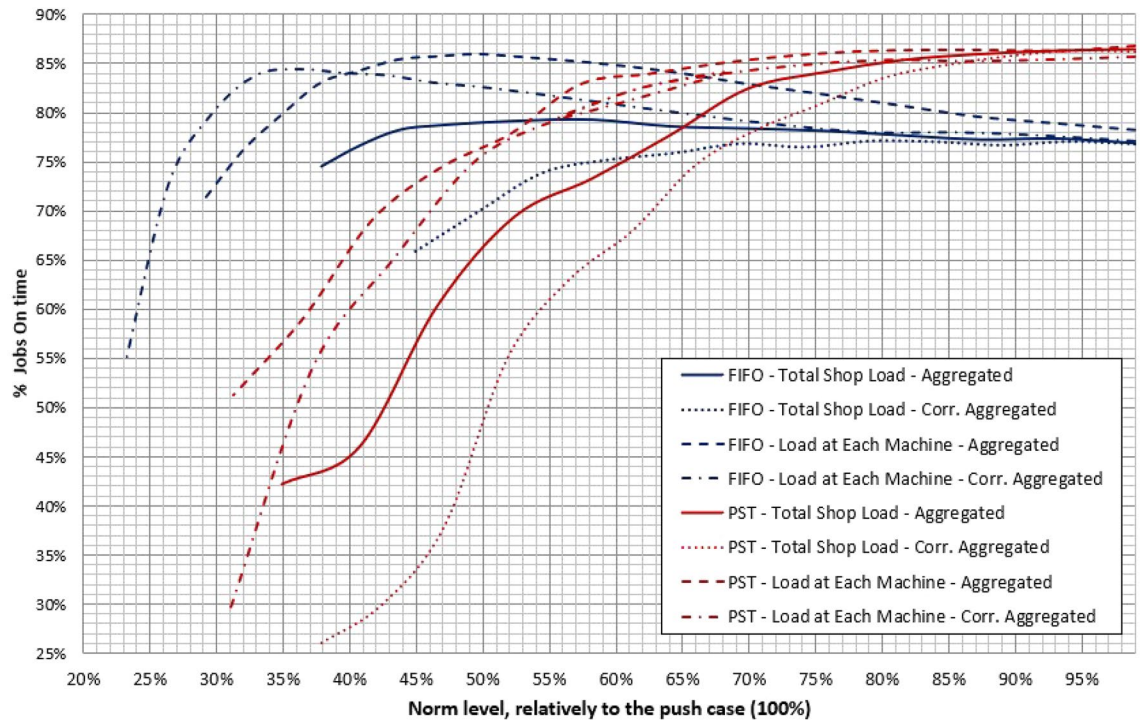


Figure 1. Trend of the percentage of tardy job of eight WLC configurations.

improvements achievable in terms of WIP reduction are limited. Conversely, as the blue curves demonstrate, the FIFO rule coupled with the load at each machine approach ensures a high reduction in WIP at the expense of a modest deterioration of the percentage of on time deliveries. In this case, in fact, there is a clear maximum point, which is located in the left-hand side of the graph.

So, as anticipated in “Performance assessment of a WLC system” section, there is a clear trade off among different solutions, and choosing one of them based on a single criterion could be somehow misleading. This concept is expressed in an even clearer way by Table 3 where, for each investigated WLC configuration, the average value of the following performance measures is given:

- *Gross Throughput Time—GTT*, the overall time spent by a job in the system both as an order pending in the PSP, or as an actual job being processed on the shop floor.
- *Shop-Floor Throughput Time—SFTT*, the overall time spent by a job in the shop floor (also known as production Lead Time).
- *Work In Process—WIP*, the amount of jobs processed in the shop floor, measured as units of remaining working time.
- *Percentage of jobs in time—% IN TIME*, the number of jobs completed and delivered before their due date over the total number of accepted jobs.
- *Tardiness—Tar*, the difference between the completion (or delivery) time C_j and the due date DD_j , if the difference is positive, zero otherwise., i.e., $T_j = \max\{0, (C_j - DD_j)\}$.
- *Earliness—Ear*, the difference between the due date and the completion time, if the difference is positive, zero otherwise., i.e., $E_j = \max\{0, (DD_j - C_j)\}$.

As anticipated, all these metrics are evaluated in average terms, with the exception of the tardiness, for which we have also provided its maximum value and its conditional expectation $\mathbb{E}[T|T>0]$, that is the mean taken on jobs with non-zero delay only. We also note that, since both the earliness and the tardiness were provided, we decided not to include the earliness $E_j = (C_j - DD_j)$ in Table 3. For the sake of completeness and of clarity the overall best is displayed immediately underneath the corresponding metric, and it is also highlighted in light blue color.

Lastly, we note that, for each considered configuration, norms have been fine tuned relatively to three alternative objective functions that, as indicated in the fifth column of the table are: (i) the minimum level of WIP, (ii) the maximum percentage of on time delivery, and (iii) the minimum tardiness. For this reason, the table is organized in blocks of three rows, one for each configuration and objective function. Note that sometimes, the values of two adjacent rows may be identical (as for rows 13 and 14); in these cases, although fine tuning was made with respect to different performance parameters, the same configuration (in terms of norm levels) was found.

As a benchmark, the same parameters have also been computed for the purely push-operating system equipped with FIFO and PST rules, respectively.

#	PPC	Contr.	Norm.	Disp.	Opt. Cr.	E[GTT]		E[SFTT]		E[WIP]		E[% IN TIME]		E[Tar]		E[Earl.]		E[T>0]		Max[Tar]		Min[Tar]	
						Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
1	PUSH	-	-	FIFO	-	24.97	24.97	24.97	24.97	107.38	107.38	76.63%	76.63%	3.99	3.99	22.01	22.01	16.64	16.64	64.00	64.00	64.00	64.00
2	PUSH	-	-	PST	-	24.23	24.23	24.23	24.23	101.24	101.24	86.57%	86.57%	1.33	1.33	17.47	17.47	9.23	9.23	26.00	26.00	26.00	26.00
3	WLC	TSL	AGG	FIFO	Min E[WIP]	43.39	15.13	15.13	61.49	61.49	74.51%	74.51%	23.16	23.16	23.64	23.64	90.36	90.36	722.91	722.91	722.91	722.91	
4	WLC	TSL	AGG	FIFO	Max E[% In Time]	27.09	20.95	20.95	88.15	88.15	79.31%	79.31%	6.45	6.45	21.81	21.81	30.07	30.07	300.65	300.65	300.65	300.65	
5	WLC	TSL	AGG	FIFO	Min E[Tar]	24.66	24.11	24.11	103.08	103.08	77.37%	77.37%	3.78	3.78	21.96	21.96	16.34	16.34	201.00	201.00	201.00	201.00	
6	WLC	TSL	AGG	PST	Min E[WIP]	75.00	16.61	16.61	56.50	56.50	42.16%	42.16%	46.58	46.58	20.54	20.54	80.17	80.17	641.35	641.35	641.35	641.35	
7	WLC	TSL	AGG	PST	Max E[% In Time]	24.38	23.93	23.93	99.00	99.00	86.48%	86.48%	1.40	1.40	17.43	17.43	9.44	9.44	73.00	73.00	73.00	73.00	
8	WLC	TSL	AGG	PST	Min E[Tar]	24.38	23.93	23.93	99.00	99.00	86.48%	86.48%	1.40	1.40	17.43	17.43	9.44	9.44	73.00	73.00	73.00	73.00	
9	WLC	TSL	CAGG	FIFO	Min E[WIP]	48.39	16.81	16.81	69.71	69.71	65.60%	65.60%	24.80	24.80	21.34	21.34	70.38	70.38	563.03	563.03	563.03	563.03	
10	WLC	TSL	CAGG	FIFO	Max E[% In Time]	25.20	23.27	23.27	99.09	99.09	77.14%	77.14%	4.08	4.08	21.72	21.72	17.35	17.35	327.00	327.00	327.00	327.00	
11	WLC	TSL	CAGG	FIFO	Min E[Tar]	24.83	24.11	24.11	103.07	103.07	77.07%	77.07%	3.81	3.81	21.86	21.86	16.22	16.22	171.50	171.50	171.50	171.50	
12	WLC	TSL	CAGG	PST	Min E[WIP]	96.87	16.94	16.94	60.79	60.79	25.93%	25.93%	65.60	65.60	20.92	20.92	88.15	88.15	511.50	511.50	511.50	511.50	
13	WLC	TSL	CAGG	PST	Max E[% In Time]	24.46	23.99	23.99	99.82	99.82	86.27%	86.27%	1.44	1.44	17.41	17.41	9.44	9.44	36.00	36.00	36.00	36.00	
14	WLC	TSL	CAGG	PST	Min E[Tar]	24.46	23.99	23.99	99.82	99.82	86.27%	86.27%	1.44	1.44	17.41	17.41	9.44	9.44	36.00	36.00	36.00	36.00	
15	WLC	LEM	AGG	FIFO	Min E[WIP]	49.12	10.27	10.27	43.95	43.95	71.22%	71.22%	29.24	29.24	25.10	25.10	100.93	100.93	857.87	857.87	857.87	857.87	
16	WLC	LEM	AGG	FIFO	Max E[% In Time]	23.02	16.44	16.44	71.67	71.67	85.95%	85.95%	4.44	4.44	22.72	22.72	30.47	30.47	676.50	676.50	676.50	676.50	
17	WLC	LEM	AGG	FIFO	Min E[Tar]	22.81	20.42	20.42	88.24	88.24	82.90%	82.90%	2.93	2.93	21.88	21.88	16.77	16.77	302.50	302.50	302.50	302.50	
18	WLC	LEM	AGG	PST	Min E[WIP]	66.27	13.65	13.65	49.94	49.94	51.22%	51.22%	40.27	40.27	21.98	21.98	82.03	82.03	239.00	239.00	239.00	239.00	
19	WLC	LEM	AGG	PST	Max E[% In Time]	24.21	23.76	23.76	96.66	96.66	86.81%	86.81%	1.30	1.30	17.47	17.47	9.09	9.09	60.50	60.50	60.50	60.50	
20	WLC	LEM	AGG	PST	Min E[Tar]	24.21	23.76	23.76	98.66	98.66	86.81%	86.81%	1.30	1.30	17.47	17.47	9.09	9.09	60.50	60.50	60.50	60.50	
21	WLC	LEM	CAGG	FIFO	Min E[WIP]	59.42	11.04	11.04	46.94	46.94	54.18%	54.18%	33.49	33.49	20.93	20.93	70.36	70.36	576.98	576.98	576.98	576.98	
22	WLC	LEM	CAGG	FIFO	Max E[% In Time]	23.09	16.16	16.16	69.58	69.58	84.42%	84.42%	3.51	3.51	21.89	21.89	22.06	22.06	396.00	396.00	396.00	396.00	
23	WLC	LEM	CAGG	FIFO	Min E[Tar]	23.32	20.32	20.32	87.17	87.17	82.13%	82.13%	3.00	3.00	21.54	21.54	16.52	16.52	322.50	322.50	322.50	322.50	
24	WLC	LEM	CAGG	PST	Min E[WIP]	70.69	15.90	15.90	63.10	63.10	29.29%	29.29%	39.27	39.27	18.16	18.16	54.39	54.39	155.50	155.50	155.50	155.50	
25	WLC	LEM	CAGG	PST	Max E[% In Time]	24.52	24.10	24.10	100.73	100.73	85.76%	85.76%	1.43	1.43	17.43	17.43	9.21	9.21	77.00	77.00	77.00	77.00	
26	WLC	LEM	CAGG	PST	Min E[Tar]	24.52	24.10	24.10	100.73	100.73	85.76%	85.76%	1.43	1.43	17.43	17.43	9.21	9.21	77.00	77.00	77.00	77.00	

- Optimized metrics are reported in the light blue colored row, overall bests are in highlighted in bold.
 - Configurations displayed in the light gray colored rows are the most performing ones. They are also displayed in the radar plot of figure 3.

Table 3. Performance measures of the investigate WLC configurations. Optimized metrics are reported in the light blue colored row, overall bests are in highlighted in bold, and local best (i.e., best results for each specific WLC configuration). Configurations displayed in the light gray colored rows are the most performing ones. They are also displayed in the radar plot of Fig. 3.

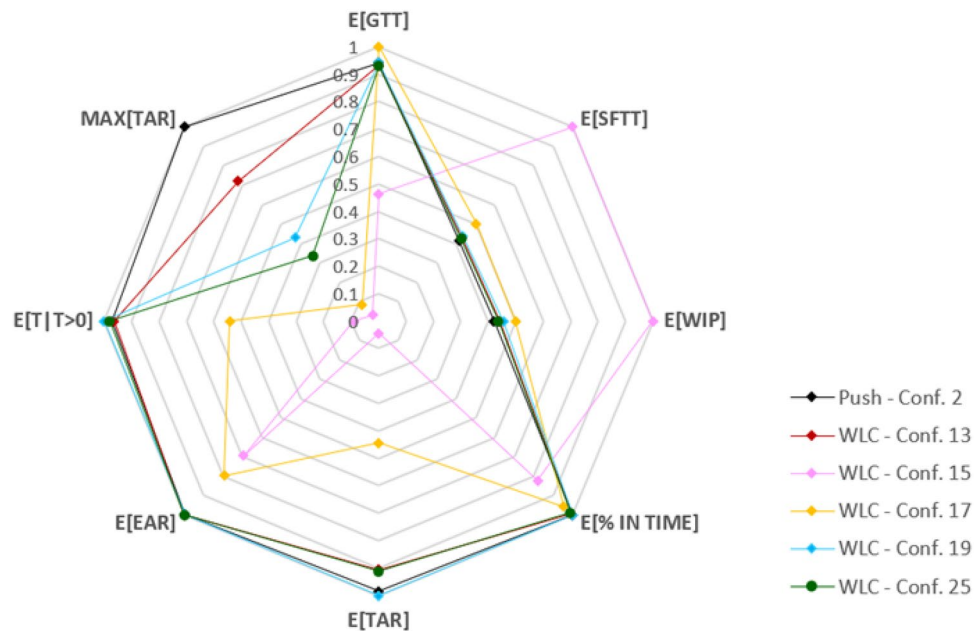


Figure 2. Radar plot of the observed performance indicators.

The WLC configuration based on the load at each machine approach, with corrected aggregated workload and PST dispatching rule, is the one that seems to perform best, as it optimizes three out of eight metrics. Yet, a clear dominant solution does not exist and, even more surprisingly, the minimum maximum tardiness is provided by the push operating system with PST dispatching rule.

An overall view of this fact is finally given by the radar chart of Fig. 2, where obtained performances are shown in relative terms. Specifically, we included in the chart all the configurations that achieved the absolute maximum for at least one performance. These configurations correspond to the rows 2, 13, 15, 17 and 19 that are highlighted in grey in Table 3. Note that configuration 2 is the push case, while configurations 13, 19 and 25 are the ones that optimize time related performances. As already noted, they all have a trend very similar to that of the push case, except for the maximum tardiness, that is reduced. We note that Fig. 2 also includes configuration 25, a WLC configuration that achieved good performances, on average, on all metrics.

Even this plot clearly shows that, if WLC is fine tuned to minimize lateness related objectives the difference between WLC and Push are marginal, and perhaps not enough to justify WLC implementation in the industry. Besides, relevant improvements can be obtained focusing on WIP reduction, especially when dispatching rules less prone to the prioritization of urgent orders are used. In this case, however, while it is still possible to keep the percentage of jobs in time at a fairly good level, the number of outliers with a very high tardiness increases significantly (see for example row 15 and 17 of Table 3). These are the jobs with high workload that wait a long time in the PSP before being released and that, once in the shop floor, although being late, are not prioritized and go through all the queues.

It is thus clear that selecting and fine tuning a WLC configuration based on a single metric is a myopic approach, that may lead to a sub-optimal solution that does not exploits the full potentiality of WLC. Hence, in the following section we will discuss how the use of the economic approach herein proposed can be effective to solve this problem.

Performance assessment in monetary terms

Parameters' setting. To introduce adequate costs in the simulated scenario, the following parameters were used:

- Cost of direct material \mathcal{M} uniformly distributed on the interval $[10, 30]$.
- Processing cost per unit of time $p_m = 17.75$, for each machine m .
- Stock holding coefficient $k_h = 0.12\%$.
- Penalties coefficient $k_p = 3k_h = 0.36\%$.
- Mark Up coefficient $\delta = 17.5\%$.

We also note that:

- All direct materials are supposed to be added to a job at its first operation, so Eq. (9) can be used.
- Since setup time were included in the processing time of each job, we also assumed that the quota of the setup cost per job is included in the processing cost per unit of time p_m .

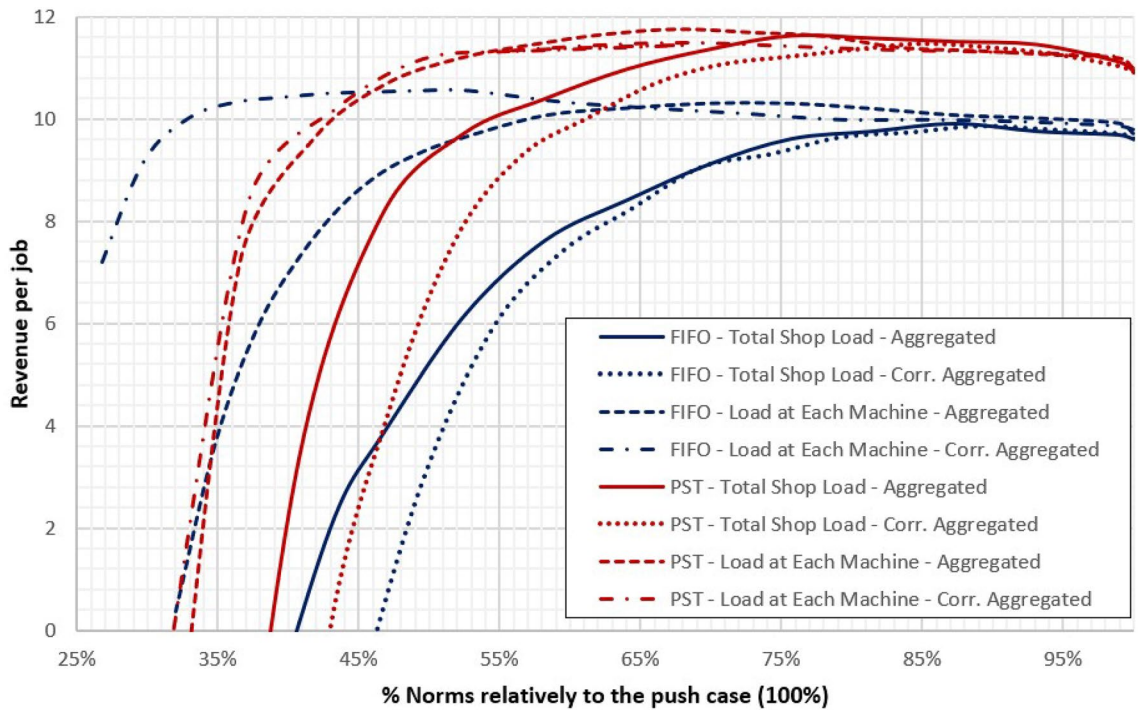


Figure 3. Trend of revenues per job of eight WLC configurations.

Groups	Count	Sum	Mean	Variance
PUSH—FIFO	10	96.505	9.650	0.5517
WLC—FIFO—TSL	10	96.875	9.687	0.5400
WLC—FIFO—LEM	10	103.333	10.333	0.1716
PUSH—PST	10	109.552	10.955	0.0711
WLC—PST—TSL	10	117.388	11.738	0.0186
WLC—PST—LEM	10	113.806	11.380	0.0742

Table 4. One-ways ANOVA summary results. *TSL* Total Shop Load, *LEM* Load at Each Machine.

Source of variance	Sum of squares	Deg. of freedom	Mean of squares	F value	p-value	F crit
Between groups	38.340	5	7.6681	32.228	4.608E-15	2.386
In groups	12.848	54	0.2379			
Total	51.188	59				

Table 5. One-ways ANOVA results.

- We limited the maximum penalty to the contractual selling price P_j , that is $PN_j = \min\{P_j, (k_p P_j T_j)\}$, where T_j is the tardiness. Hence, if the tardiness is so high that the penalty equals the contractual price the income is null (as for a lost sale) and there is a loss equal to all costs incurred.

Using these settings, in the baseline case (i.e., the push operating system with FIFO dispatching rule) the average selling price is 100, the average total cost is 90, for an expected operating revenue, which we will simply refer to as revenue per job, of 10 units for each job sold. We note that the system we proposed could be easily adapted to different cost and income settings.

Obtained results. The trend of the average revenue per job (on the y axis) versus the norm level (on the x axis) is shown in Fig. 3, for all eight investigated configurations (see also Fig. 1 and Table 2). We note at first that the PST norm significantly outperform FIFO in terms of Revenue per Job. Also, WLC configurations with Aggregated load perform slightly better than those with Corrected aggregated load, and the Aggregated load

Config. #1	Config. #2	Mean diff	Critical value	Significant
WLC—PST—TSL	PUSH—FIFO	2.088	0.751	Yes
WLC—PST—TSL	WLC—FIFO—TSL	2.051	0.751	Yes
WLC—PST—TSL	WLC—FIFO—LEM	1.405	0.751	Yes
WLC—PST—TSL	PUSH—PST	0.783	0.751	Yes
WLC—PST—TSL	WLC—PST—LEM	0.358	0.751	No

Table 6. Tukey–Kramer test.

PPC	Optimiz. criterium	NORM	%NORM (%)	E[GTT]	E[SFTT]	E[WIP]	E[% IN TIME] (%)	E[Tar]	E[Ear]	E[T>0]	Max[TAR]	E[Rev]
PUSH	[-]	40	100	24.23	24.23	101.24	86.57	1.33	17.47	9.23	26.00	10.97
WLC (LEM-AGG-PST)	Min E[% In Time]	38	95	24.21	23.76	96.66	86.81	1.30	17.47	9.09	60.50	11.12
WLC (LEM-AGG-PST)	Max E[Revenue]	26	65	24.89	21.11	79.20	85.12	1.59	17.29	9.83	80.40	11.74

Table 7. Summary results.

at Each machine achieves the highest value of Revenue per job. However, we believe important to note that, although the curves, and specifically PST curves (i.e. the red-brown ones), are rather flat for high values of the norms, a global maximum always exists as all curves are convex. This fact means that all WLC configuration we tested achieved a maximum in the Revenue per job function for some value of the norm below 100%, i.e. in a non-push condition.

To assess the impact that our the operational parameters we selected have on costs and incomes, we performed a preliminary 4-ways ANOVA. Specifically, we used the revenue per job as response variable and (i) Norm level, (ii) Aggregation of workload measure, (iii) Workload accounting over time and (iv) Dispatching rule as factors. The analysis confirmed all the hypotheses and indeed all factors and interactions were significant with a p -value lower than 0.005, with the only exception of the Workload accounting over time criterium, i.e. aggregated load vs. corrected aggregated load. For this reason, we excluded it from the analysis from now on. For a more detailed explanation on this point, please refer to Online Appendix A.

This fact is confirmed by the one-way ANOVA of Tables 4 and 5, that reveals that the points of maximum revenue per job of the alternative configurations differ significantly, in statistical terms. Note that, since the effect of the Workload accounting over time resulted non-significant in the preliminary 4-ways ANOVA, the analysis was limited to four WLC alternatives and the push system was taken as benchmark.

As Table 4 reports, revenues per job clearly differ when the PST rule is used instead of the FIFO one. This fact was also confirmed by a Tukey–Kramer post-hoc test (at level $\alpha = 0.1$); a synthesis of the pairwise comparison is provided in Table 6, which also highlights a significant difference between the optimal WLC configuration and the corresponding push condition (reported with bold character in Table 6).

It is interesting to note that the WLC configuration that maximizes the expected Revenue per job is the same one that maximized the percentage of on time deliveries (as we found in “Performance assessment of a WLC system” section), namely Aggregation of workload method: Load at each machine; Workload accounting over time: Aggregated; and Dispatching rule: PST. Yet, as it is shown by Table 7, the operating point that maximises the value of Revenue per job clearly differs from the one that maximizes the percentage of jobs delivered in time. Specifically, as it is shown in Table 7, which displays the norm level at each machine, either in units of time, or as a percentage relative to the norm of the push case, to maximise revenues in our specific scenario it is advisable to reduce the level of the norms to reduce WIP and stockholding costs. This improvement is large enough to positively balances the extra costs related to a small deterioration of the percentage of on time deliveries, that are necessarily incurred when the norms’ level is reduced. It is also worth noting that the difference in expected revenue, obtained at these two alternative operating points, is statistically significant with a p -value of a one-tailed T-test $\ll 0.01$. On the contrary, the expected revenue of the WLC system maximizing the percentage of jobs delivered in time does not significantly differ from that of the benchmark push-system, with a p -value > 0.1 . This is another clear indication of the opportunity to fine-tune a WLC system considering the expected Revenue per job, rather than a single performance measure.

Sensitivity analysis. Clearly, the results we discussed above depend on the cost structure under analysis and, most of all, on the values of the coefficients k_h and k_p used to estimate the stockholding and the penalties cost, respectively. As above discussed, in fact, these costs directly reflect antithetical operating performances (mainly WIP level and tardiness) for which a tradeoff must be achieved. Hence, for a further insight, we repeated the optimization analysis for different values of the cost ratio $r = \left(\frac{k_p}{k_h}\right)$, by fixing k_h at its original value of 0.12% and by progressively reducing k_h from $8k_h$ to $0.25k_h$. Obtained results are graphically summarized in Fig. 4,

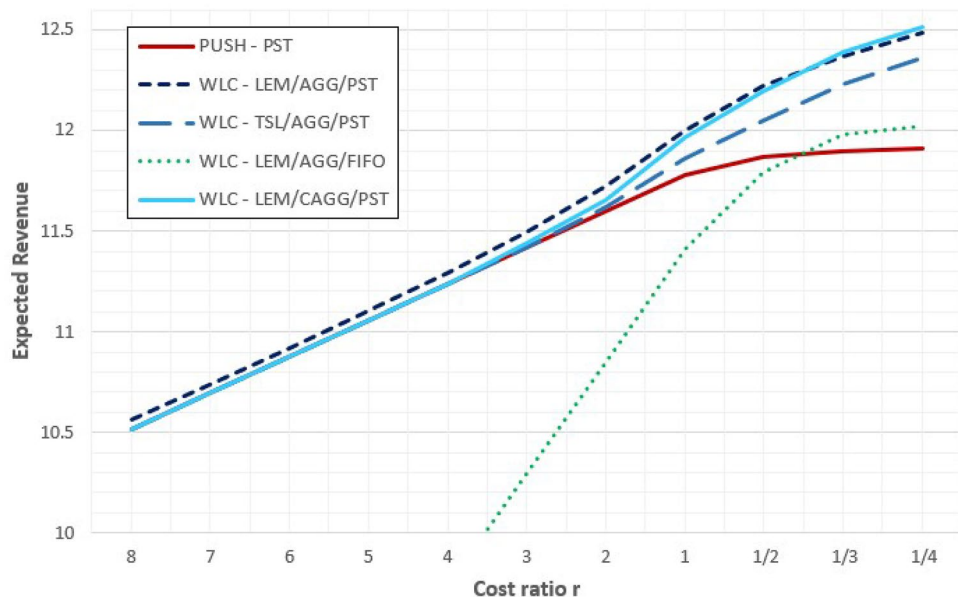


Figure 4. Sensitivity analysis.

which shows the maximum expected revenue that can be achieved using the most effective WLC configurations, each one fine-tuned for each level of the investigated cost ratio r .

As it can be seen, the WLC configurations based on the aggregate approach, load for every machine and the PST rules is almost ever the dominant solution and the positive gap, increases as the ratio r decreases. As already noted, indeed, this configuration is the one that tends to optimize the trade-off between low WIP and short and stable gross throughput time, and this fact becomes particularly relevant as the stockholding costs increase relative to the penalty ones. The only exception is the WLC configuration based on corrected aggregate approach, load for every machine and the PST rules, that becomes the best option for very high levels of the stockholding cost (the indifference point is found around $r = 2/5$). The use of the corrected aggregate approach, in fact, is more precise, but also more selective and binding; hence, relatively to the aggregate approach, it is more performant in terms of WIP, but less in terms of percentage of tardy job. In terms of cost, this condition becomes valuable only when r is small. For the same reason for low levels r , even the use of the FIFO policies becomes quite performing, as it outperforms the benchmark push approach, yet it remains lower than the other two configurations. Conversely, the economical gap among the benchmark push approach and the WLC policies based on the PST rule becomes almost inappreciable when r is big.

Conclusions, managerial implications, and future works

This paper deals with WLC, a hybrid push–pull PPC system that has seen a great deal of attention in the last decades. Focusing on performance assessment of rule-based WLC systems, the paper shows that setting the norms and other operating parameters by considering the percentage of tardy job as the only optimization criterion is generally sub-optimal. To get the most out of WLC, in fact, and to achieve benefits that could justify the required initial investment, practitioners should look for a trade-off between WIP reduction and due date compliance. To simplify this multi-criteria optimization problem, we propose to homogenize alternative performance criteria, even conflicting ones, by translating them in monetary terms. To this aim, we developed a general economic model to properly consider the economic valorization of the operating performances regarding a WLC system, namely the Revenue per job. It is the authors' opinion, in fact, that such a trade-off might only be identified and analyzed by means of an economic model, which enables the comparison of different system performances. Also, without an economic model that justifies the benefits of WLC, companies, and especially SMEs, will not be supported to implement 'off-the-shelves' WLC systems. We do believe, instead, that an adequate and realistic economic calculation of WLC benefits might help SMEs to decide for its implementation. Industrial sectors that could benefit from the approach we propose in this study are sub-contract machining shops, producers of mechanical components (e.g. for the automotive industry) and producers of electronic components (e.g. wafer fabrication).

The model has been finally tested via simulation and validated through a comprehensive statistical analysis. What emerged from the analysis is the fact that the economic optimum point (i.e. the maximum expected revenue per job) hardly ever coincides with the point that optimizes a single performance measure. The difference is even more prominent in case of high stock holding costs, comparable to those related to penalties. We note that the existence of these economic optimum points is calculated in a simulated system taken from the mainstream of WLC research, whose parameters and values have been used with profit by previous studies aiming at minimizing shop floor throughput times and percentages of tardy jobs in WLC systems. Thus, we believe that the existence of said optimum values in the Revenue per job function of such a simulated system justifies the need to look for

it from a broader, and more general, perspective. Also, we reported how the specific performances of different WLC configurations might impact on the economic method we propose, and we reported a sensitivity analysis that, starting from the preliminary observations of^{20,22}, shows the results of a continuous variation of the cost ratio r , and it allows the extension of our WLC evaluation approach to different sectors and industries. As an example, if we recall the above-mentioned sectors, a low r value is expected in companies producing electronic components, where penalty costs are higher than holding costs. Conversely, a high r value is typical of subcontract machining shops and of producers of mechanical components, where WIP value and holding costs are higher. Thus, these sectors could use the model we propose to better understand cost savings or other benefits expected from WLC adoption.

In general, with respect to the WLC configurations we investigated, the best economic performance can be achieved with a WLC system based on aggregate load, load at each machine, and PST dispatching rule. Yet, a WLC system based on the corrected aggregate approach becomes preferable for lower levels of the cost ratio $r = \left(\frac{k_h}{k_p}\right)$. Indeed, in this condition the additional WIP reduction that can be achieved is enough to offset the unavoidable worsening of delivery times. In this regard, it has to be said that our model underestimates the benefits related to WIP reduction. Less WIP, indeed, reduces errors (as visual management becomes easier), and assures other savings related to improved tidiness and reduced occupied space that could be dedicated to other activities. Since these savings have not yet been captured by our model, it is likely that the corrected aggregated approach becomes advantageous also at higher value of the ratio r . For the same reason, advantages of WLC over the benchmark push-oriented job shop might become evident also for values of r lower than 3, where, at the moment, the economic difference between WLC and push is very small. In this respect, we finally note that we limited the analysis to a rather unsaturated system with an average utilization rate of 90%. This decision, which is in line with the mainstream of WLC literature, suggests that the economic advantages achievable by applying WLC to almost saturated systems (i.e., utilization rate of 95% or higher, at least for certain amounts of time), with respect to push-operated systems, could be bigger than what we report. In such conditions, in fact, push systems will greatly diverge from hybrid or pull ones, in terms of queues and shopfloor throughput time. All these issues could be further investigated in subsequent works. Also, aligned with research mainstream on WLC, our model considers only machines as constrained resources. In some situations, this might not be the case, as human resources could be equally critical bottlenecks (e.g. manual assembly station or testing benches). We note that, in these cases, the general approach we followed still holds true, but some improvements are needed, and scientific research on this topic is still scarce. Developing a broader WLC approach capable of operating with a combination of human and machine resources, which takes into consideration the peculiarities of each one of them, could definitely be a topic of future research. Another important aspect that is currently undervalued in most WLC studies is the preliminary identification and mapping, and the subsequent management of critical constrained resources. As we mentioned above, the focus is (almost) only on machines, which are often considered identical in terms of utilization and availability rates. Thus, much WLC research does not consider stable or shifting bottlenecks. Therefore, considering an early identification and management of critical constrained resources, be them human or machines, could bridge the gap among WLC and other PPC systems (e.g. theory of constraints and drum-buffer-rope), and extend the impact of WLC on flexibility, by covering other aspects of flexibility, such as labor and material handling flexibility.

To conclude, we provide a last remark concerning the FIFO policy. As shown in the present paper, although this rule assures significant advantages in terms of WIP, it is always penalizing in economic terms. As discussed in “Performance analysis” section, the problem is mainly related to a minority of jobs requiring very high processing times. The fact is that these outliers wait for very long time in the PSP, before being admitted in the shop floor; also, once in the system, although already late, they are forced to queuing up because of the FIFO policy. Hence, incurred penalties are very high and detrimental in terms of revenue per job. To limit this issue, it could be wise to devise an acceptance policy that rejects new orders with a heavy workload, anytime the workload in the system approaches norms’ level, or to incorporate some other expediting corrections to the order release policy that could improve the system. As an example, only regular job should be sorted using the FIFO rule, whereas jobs that have been waiting for long in the PSP should be prioritized. The economic evaluation of this approach could be the topic of a future work.

Data availability

The datasets generated and/or analysed during the current study are available in the ‘wlc-economic-valorization’ repository, available at the current web link: <https://github.com/dmezzogori/wlc-economic-valorization>. Please, note that the repository might be amended and improved at a later stage, upon paper’s acceptance or by answering to reviewers comments.

Received: 27 May 2022; Accepted: 3 August 2022

Published online: 26 August 2022

References

1. Thürer, M., Stevenson, M. & Silva, C. Three decades of workload control research: A systematic review of the literature. *Int. J. Prod. Res.* **49**(23), 6905–6935 (2011).
2. Thürer, M., Stevenson, M. & Land, M. J. On the integration of input and output control: Workload Control order release. *Int. J. Prod. Econ.* **174**, 43–53 (2016).
3. M. Bertolini, G. Romagnoli, F. Zammori, and D. Uckelmann, “An industrial perspective on Workload Control: Economic evaluation and suitable technologies,” in *Proceedings of 2017 International Conference on Industrial Engineering and Systems Management, IEEE IESM 2017*, 2018, p. Paper in press.

4. Hendry, L., Huang, Y. & Stevenson, M. Workload control: Successful implementation taking a contingency-based view of production planning & control. *Int. J. Oper. Prod. Manag.* **33**(1), 69–103 (2013).
5. Raaymakers, W. H. M., Bertrand, J. W. M. & Fransoo, J. C. Performance of workload rules for order acceptance in batch chemical manufacturing. *J. Intell. Manuf.* **11**(2), 217–228 (2000).
6. M. Bertolini, G. Romagnoli, and F. Zammori, “Simulation of two hybrid production planning and control systems: A comparative analysis,” in *Proceedings of 2015 International Conference on Industrial Engineering and Systems Management, IEEE IESM 2015*, 2016, pp. 388–397.
7. M. Bertolini, G. Romagnoli, and F. Zammori, “Assessing performance of Workload Control in High Variety Low Volumes MTO job shops: A simulative analysis,” in *Proceedings of 2015 International Conference on Industrial Engineering and Systems Management, IEEE IESM 2015*, 2016, pp. 362–370.
8. Zammori, F., Ferretti, C., Ganugi, P. & Mezzogori, D. Workload control with shifting bottlenecks: Norms optimisation through design of experiments. *Int. J. Math. Model. Numer. Optim.* **11**(4), 355 (2021).
9. Bertolini, M., Romagnoli, G. & Zammori, F. 2MTO, a new mapping tool to achieve lean benefits in high-variety low-volume job shops. *Prod. Plan. Control* **28**(5), 444–458 (2017).
10. Stevenson, M., Huang, Y., Hendry, L. C. & Soepen, E. The theory and practice of workload control: A research agenda and implementation strategy. *Int. J. Prod. Econ.* **131**(2), 689–700 (2011).
11. Fernandes, N. O., Land, M. J. & Carmo-Silva, S. Workload control in unbalanced job shops. *Int. J. Prod. Res.* **52**(3), 679–690 (2014).
12. Thürer, M., Fernandes, N. O., Stevenson, M. & Qu, T. On the backlog-sequencing decision for extending the applicability of ConWIP to high-variety contexts: An assessment by simulation. *Int. J. Prod. Res.* **55**(16), 4695–4711 (2017).
13. Petroni, A., Zammori, F. & Marolla, G. World class manufacturing in make-to-order batch-production SMEs: An exploratory analysis in northern Italy. *Int. J. Bus. Excell.* **11**(2), 241 (2017).
14. Gómez Paredes, F. J., Godinho Filho, M., Thürer, M., Fernandes, N. O. & Jabbour, C. J. C. Factors for choosing production control systems in make-to-order shops: A systematic literature review. *J. Intell. Manuf.* **33**(3), 639–674 (2022).
15. Mezzogori, D., Romagnoli, G. & Zammori, F. Deep learning and WLC: How to set realistic delivery dates in high variety manufacturing systems. *IFAC-PapersOnLine* **52**(13), 2092–2097 (2019).
16. Melnyk, S. & Ragatz, G. L. Order review/release: Research issues and perspectives. *Int. J. Prod. Res.* **27**(7), 1081–1096 (1989).
17. Singgih, I. K., Yu, O., Kim, B. I., Koo, J. & Lee, S. Production scheduling problem in a factory of automobile component primer painting. *J. Intell. Manuf.* **31**(6), 1483–1496 (2020).
18. Moon, I., Lee, S., Shin, M. & Ryu, K. Evolutionary resource assignment for workload-based production scheduling. *J. Intell. Manuf.* **27**(2), 375–388 (2016).
19. Hopp, W. W. J. & Spearman, M. L. M. To pull or not to pull: What is the question?. *Manuf. Serv. Oper. Manag.* **6**(2), 133–148 (2004).
20. Neuner, P. & Haeussler, S. Rule based workload control in semiconductor manufacturing revisited. *Int. J. Prod. Res.* **59**(19), 5972–5991 (2021).
21. E. Albey and R. Uzsoy, Lead time modeling in production planning, in *2015 Winter Simulation Conference (WSC)*, 2015, pp. 1996–2007.
22. Schneckeneither, M., Haeussler, S. & Gerhold, C. Order release planning with predictive lead times: A machine learning approach. *Int. J. Prod. Res.* **59**(11), 3285–3303 (2021).
23. Kutanoglu, E. An analysis of heuristics in a dynamic job shop with weighted tardiness objectives. *Int. J. Prod. Res.* **37**(1), 165–187 (1999).
24. M. Schneckeneither and S. Haeussler, *Reinforcement learning methods for operations research applications: The order release problem*, vol. 11331 LNCS, no. 2016. Springer International Publishing, 2019.
25. Schneckeneither, M., Windmueller, S. & Haeussler, S. Smart Short Term Capacity Planning: A Reinforcement Learning Approach. In *Advances in Production Management Systems. Artificial Intelligence for Sustainable and Resilient Production Systems: IFIP WG 5.7 International Conference, APMS 2021, Nantes, France, September 5–9, 2021, Proceedings, Part I* (eds Dolgui, A. et al.) 258–266 (Springer, 2021).
26. T. Ziarnetzky, N. B. Kacar, L. Monch, and R. Uzsoy, Simulation-based performance assessment of production planning formulations for semiconductor wafer fabrication. in *2015 Winter Simulation Conference (WSC)*, 2015, pp. 2884–2895.
27. Kacar, N. B., Monch, L. & Uzsoy, R. Planning wafer starts using nonlinear clearing functions: A large-scale experiment. *IEEE Trans. Semicond. Manuf.* **26**(4), 602–612 (2013).
28. Kacar, N. B., Irdem, D. F. & Uzsoy, R. An experimental comparison of production planning using clearing functions and iterative linear programming-simulation algorithms. *IEEE Trans. Semicond. Manuf.* **25**(1), 104–117 (2012).
29. Pürgstaller, P. & Missbauer, H. Rule-based vs. optimisation-based order release in workload control: A simulation study of a MTO manufacturer. *Int. J. Prod. Econ.* **140**(2), 670–680 (2012).
30. Haeussler, S. & Netzer, P. Comparison between rule- and optimization-based workload control concepts: A simulation optimization approach. *Int. J. Prod. Res.* **58**(12), 3724–3743 (2020).
31. Wight, O. Input/Output control a real handle on lead time. *Prod. Invent. Manag. J.* **11**(3), 9–31 (1970).
32. Thürer, M., Stevenson, M. & Qu, T. Job sequencing and selection within workload control order release: An assessment by simulation. *Int. J. Prod. Res.* **54**(4), 1061–1075 (2016).
33. Land, M. J. & Gaalman, G. J. C. Workload control concepts in job shops. A critical assessment. *Int. J. Prod. Econ.* **46–47**, 535–548 (1996).
34. Day, G. S. The capabilities of market-driven Organizations. *J. Mark.* **58**(4), 37 (1994).
35. Dwivedi, A. et al. Addressing the challenges to sustainable initiatives in value chain flexibility: Implications for sustainable development goals. *Glob. J. Flex. Syst. Manag.* **22**(S2), 179–197 (2021).
36. Ming, L. T. Value chain flexibility with RFID: A case study of the octopus card. *Int. J. Eng. Bus. Manag.* **3**, 3 (2011).
37. Zhang, Q., Vonderembse, M. A. & Lim, J. S. Value chain flexibility: A dichotomy of competence and capability. *Int. J. Prod. Res.* **40**(3), 561–583 (2002).
38. Moreira, M. R. A. & Alves, R. A. F. S. A methodology for planning and controlling workload in a job-shop: A four-way decision-making problem. *Int. J. Prod. Res.* **47**(10), 2805–2821 (2009).
39. Bergamaschi, D., Cigolini, R., Perona, M. & Portioli, A. Order review and release strategies in a job shop environment: A review and a classification. *Int. J. Prod. Res.* **35**(2), 399–420 (1997).
40. G. Marangoni, G. Romagnoli, and F. Zammori, “Multiple Value Stream Mapping : How to implement Work Load Control in complex systems,” in *IESM’2013, International Conference on Industrial Engineering and Systems Management*, 2013.
41. Mezzogori, D., Romagnoli, G. & Zammori, F. Defining accurate delivery dates in make to order job-shops managed by workload control. *Flex. Serv. Manuf. J.* **33**(4), 956–991 (2021).
42. Wiendahl, H.-P. *Load-Oriented Manufacturing Control* (Springer, 1995).
43. Bechte, W. Theory and practice of load-oriented manufacturing control. *Int. J. Prod. Res.* **26**(3), 375–395 (1988).
44. Land, M. J. & Gaalman, G. J. C. The performance of workload control concepts in job shops: Improving the release method. *Int. J. Prod. Econ.* **56–7**, 347–364 (1998).
45. Cigolini, R. & Portioli-Staudacher, A. An experimental investigation on workload limiting methods within ORR policies in a job shop environment. *Prod. Plan. Control* **13**(7), 602–613 (2002).

46. Yan, H., Stevenson, M., Hendry, L. C. & Land, M. J. Load-Oriented Order Release (LOOR) revisited: Bringing it back to the state of the art. *Prod. Plan. Control* **27**(13), 1078–1091 (2016).
47. Stevenson, M. Refining a workload control (WLC) concept: A case study. *Int. J. Prod. Res.* **44**(4), 767–790 (2006).
48. Thüner, M. & Stevenson, M. Workload control in job shops with re-entrant flows: An assessment by simulation. *Int. J. Prod. Res.* **54**(17), 5136–5150 (2016).
49. van Ooijen, H. P. G. & Bertrand, J. W. M. Economic due-date setting in job-shops based on routing and workload dependent flow time distribution functions. *Int. J. Prod. Econ.* **74**(1–3), 261–268 (2001).
50. Chiang, T. C. & Fu, L. C. Using dispatching rules for job shop scheduling with due date-based objectives. *Int. J. Prod. Res.* **45**(14), 3245–3262 (2007).
51. Hutter, T., Haeussler, S. & Missbauer, H. Successful implementation of an order release mechanism based on workload control: A case study of a make-to-stock manufacturer. *Int. J. Prod. Res.* **56**(4), 1565–1580 (2018).
52. Land, M. J., Stevenson, M. & Thüner, M. Integrating load-based order release and priority dispatching. *Int. J. Prod. Res.* **52**(4), 1059–1073 (2014).
53. Kundu, K. & Portioli-Staudacher, A. Investigation of effectiveness of order review and release models in make to order supply chain. *MATEC Web Conf.* **81**(06003), 06003 (2016).
54. Kundu, K., Land, M. J., Portioli-Staudacher, A. & Bokhorst, J. A. C. Order review and release in make-to-order flow shops: Analysis and design of new methods. *Flex. Serv. Manuf. J.* **33**(3), 750–782 (2021).
55. Renna, P. Allocation improvement policies to reduce process time based on workload evaluation in job shop manufacturing systems. *Int. J. Ind. Eng. Comput.* **8**, 373–384 (2017).
56. Thüner, M. *et al.* Lean control for make-to-order companies: Integrating customer enquiry management and order release. *Prod. Oper. Manag.* **23**(3), 463–476 (2014).
57. Chen, J. C., Chen, T. L., Pratama, B. R. & Tu, Q. F. Capacity planning with ant colony optimization for TFT-LCD array manufacturing. *J. Intell. Manuf.* **29**(8), 1695–1713 (2018).
58. Huang, Y. Information architecture for effective workload control: An insight from a successful implementation. *Prod. Plan. Control* **28**(5), 351–366 (2017).
59. Dumetz, L. *et al.* A simulation framework for the evaluation of production planning and order management strategies in the sawmilling industry. *IFAC-PapersOnLine* **28**(3), 622–627 (2015).
60. Fu, M. C., Herrmann, J. W. & Narayanaswamy, M. Setting thresholds for periodic order release. *J. Intell. Manuf.* **8**(5), 369–383 (1997).
61. Asmundsson, J., Rardin, R. L., Turkseven, C. H. & Uzsoy, R. Production planning with resources subject to congestion. *Nav. Res. Logist.* **56**(2), 142–157 (2009).
62. J. Liu, C. Li, F. Yang, H. Wan, and R. Uzsoy, Production planning for semiconductor manufacturing via simulation optimization, *Proc. 2011 Winter Simul. Conf.*, pp. 3617–3627, 2011.
63. M. Li, A Metamodeling-Based Approach for Production Planning, *Proc. 2014 Winter Simul. Conf.*, pp. 2204–2215, 2014.
64. W. J. Stevenson, *Operations management*, 10th Editi. Milan, Italy: McGraw-Hill/Irwin, 2009.
65. Zammori, F., Braglia, M. & Castellano, D. Just-in-time parts feeding policies for paced assembly lines: Possible solutions for highly constrained layouts. *Int. Trans. Oper. Res.* **23**(4), 691–724 (2016).
66. Bottani, E., Bertolini, M., Rizzi, A. & Romagnoli, G. Monitoring on-shelf availability, out-of-stock and product freshness through RFID in the fresh food supply chain. *Int. J. RF Technol.* **8**(1–2), 33–55 (2017).
67. T. Gruen and D. Corsten, “A Comprehensive Guide to Retail Out-of-Stock Reduction in the Fast-Moving Consumer Goods Industry,” 2007.
68. Thüner, M., Stevenson, M., Land, M. J. & Fredendall, L. D. On the combined effect of due date setting, order release, and output control: An assessment by simulation. *Int. J. Prod. Res.* **57**(6), 1741–1755 (2019).
69. Thüner, M., Stevenson, M., Silva, C., Land, M. J. & Fredendall, L. D. Workload control and order release: A lean solution for make-to-order companies. *Prod. Oper. Manag.* **21**(5), 939–953 (2012).

Author contributions

G.R. mostly wrote “Introduction” to “Performance assessment of a WLC system” sections and prepared Table 1, D.M. mostly wrote “A proposal for an economic evaluation of WLC systems” and “Performance assessment of a WLC system” sections and prepared Table 2 and Fig. 1, F.Z. prepared and run the simulations, prepared all other figures and tables, and mostly wrote “Performance assessment in monetary terms” section. All authors wrote the “conclusions” section and reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-17968-5>.

Correspondence and requests for materials should be addressed to G.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022