

Workload Control Order Release in General and Pure Flow Shops with Limited Buffer Size Induced Blocking: An Assessment by Simulation

Matthias Thürer (corresponding author), Lin Ma, and Mark Stevenson

Name: Prof. Matthias Thürer
Institution: Jinan University
Address: School of Intelligent Systems Science and Engineering
Jinan University (Zhuhai Campus)
519070, Zhuhai, PR China
E-mail: matthiasthurer@workloadcontrol.com

Name: Lin Ma
Institution: Jinan University
Address: School of Management
Jinan University
510632, Guangzhou, PR China
E-mail: malin15102939217@163.com

Name: Prof. Mark Stevenson
Institution: Lancaster University
Address: Department of Management Science
Lancaster University Management School
Lancaster University
LA1 4YX - U.K.
E-mail: m.stevenson@lancaster.ac.uk

Keywords: *Workload Control; Order Release; Job Shop; Flow Shop; Blocking.*

Workload Control Order Release in General and Pure Flow Shops with Limited Buffer Size Induced Blocking: An Assessment by Simulation

Abstract

Most manufacturing shops in practice have limited physical space in front of each workstation, due, for example, to physical, economical or operational constraints. As a result, a job may cause blocking because it has to remain at a given station after an operation has been completed until space in front of the next station in its routing becomes available. Despite this practical reality, the Workload Control literature typically assumes infinite buffer limits and therefore neglects the impact of blocking. Using simulation, we highlight the direct, detrimental impact of blocking in both the pure and general flow shop. Workload Control order release dampens the effect of blocking and improves overall performance. This makes Workload Control order release even more important in the context of shops with blocking or physical space constraints. Further analysis reveals that the impact of blocking is less pronounced in the pure flow shop given its directed routing. Finally, most of the blocking that occurs is because jobs cannot enter the shop, i.e. there is no space in front of the gateway station. This re-emphasizes the close relationship between blocking and release methods that limit the workload, and it highlights the importance of workload balancing.

Keywords: *Workload Control; Order Release; Job Shop; Flow Shop; Blocking.*

1. Introduction

This study explores the impact of Workload Control order release in high-variety make-to-order shops with limited queue space (or buffer size) and, as a result, the impact of blocking. The motivation for this study arose from a project that sought to improve process flow within one plant of a leading paint manufacturer in the Pearl River Delta of the People's Republic of China. The manufacturer produces a high variety of different products on a to-order basis. It has seven production bases and more than 2,500 employees, realizing a net annual sales income of approximately 2.3 billion RMB. While the present study was originally motivated by this single company, the problem and subsequent findings are argued to be of much broader relevance. For example, many production lines observe blocking due to a limited buffer size (e.g. Roser *et al.*, 2014) while buffer space allocation is a core task in line design (Gershwin & Schor, 2000; Shi & Gershwin, 2016; Xi, *et al.*, 2019).

Workload Control is a production planning and control concept developed for high-variety make-to-order contexts (Zäpfel & Missbauer, 1993; Stevenson *et al.*, 2005). A key production control function of Workload Control is order release control. When order release control is applied, jobs do not enter the shop floor directly. Instead, they are retained in a pre-shop pool and released using criteria that allow the shop to achieve certain performance targets, e.g. to restrict the level of work-in-process inventory and/or to maximize due date adherence. Given its importance, a broad literature has emerged to assess the performance of Workload Control order release methods both through simulation (e.g. Land & Gaalman, 1998; Perona & Portioli, 1998; Sabuncuoglu & Karapinar, 2000; Cigolini, & Portioli-Staudacher, 2002; Portioli-Staudacher & Tantardini, 2011; Thürer *et al.*, 2012, 2014; Fernandes *et al.*, 2016, 2020; Gonzalez-R *et al.*, 2018; Haeussler & Netzer, 2019) and in practice (e.g. Wiendahl, 1992; Bechte 1994; Hendry *et al.*, 2013; Silva *et al.*, 2015; Perona *et al.*, 2016; Huang, 2017; Hutter *et al.*, 2018).

A common assumption in the literature on Workload Control order release is that queue space is infinite, i.e. there is no physical limitation on the amount of work that can queue in front of a station, irrespective of the limits applied by order release control. In practice however, queue space is often limited (Leisten, 1990; Sabuncuoglu & Karapinar, 2000; Haskose, *et al.*, 2004; Liu *et al.*, 2018); for example, due to physical, economical, or operational constraints. This in turn may result in blocking since there is no space or outlet for one station to discharge its jobs to another (Buzacott, 1976). Blocking is hereby defined as the situation where a job, having completed all of

its processing requirements at a station, remains at the station (and thus blocks station capacity) until space in the queue at the next station in its routing becomes available (Roser *et al.*, 2014). The job simply has nowhere to go if it is to follow its planned route. While these situations may commonly arise in practice, constraints of queue space and thus the occurrence of blocking has not been considered to be a factor that affects the applicability of Workload Control (see, e.g. Henrich *et al.* 2003; Cransberg *et al.*, 2016).

In general, shop control problems under finite buffer constraints have received little attention (Mascis & Pacciarelli, 2002). In the scheduling literature, for example, it is typically considered too complex to analyze (Hall & Sriskandarajah, 1996; Haskose *et al.*, 2002 and 2004). Meanwhile, the existing scheduling literature focuses on deterministic contexts and neglects how blocking may affect the performance of order release control. To the best of our knowledge the only two studies considering the impact of order release in shops with finite buffer space are those by Buzacott (1976) and Thüerer *et al.* (2013). Buzacott (1976) however only focused on a job shop comprised of two machines with equal processing rates to analytically evaluate the maximum output rate. Meanwhile, although Thüerer *et al.* (2013) focused on the impact of order release on the required storage space in high-variety make-to-order job shops, the authors did not consider the occurrence of blocking. Thüerer *et al.* (2013) showed that the average of the storage space requirements can be reduced by controlled order release, but peaks that will lead to blocking still remain.

In general, both of the aforementioned studies can be criticized for the use of a job shop environment, since undirected routings may lead to the mutual blocking of stations. This is similar to the blocking behavior of the Paired Cell Overlapping Loops of Cards with Authorization (POLCA) approach observed, for example, in Lödding *et al.* (2003) and Harrod & Kanet (2013). This type of blocking did not occur in Buzacott (1976) since an analytical study was used which does not capture dynamic behavior; the same criticism holds for the existing scheduling literature (see, e.g. Mascis & Pacciarelli, 2002). Meanwhile, it did not occur in Thüerer *et al.* (2013) either since the authors did not consider blocking. Moreover, the best-performing release rule in Buzacott (1976), which releases jobs whenever a station runs idle, becomes dysfunctional in shops with more directed routings and a gateway station (Thüerer *et al.*, 2015a). Thus, there is a need to further extend existing literature on the impact of blocking by considering dynamic behavior and the impact of blocking on order release. In response, this study uses simulation to assess, for the first

time, the performance of Workload Control order release in high-variety make-to-order (general and pure) flow shops with blocking.

The simulation model used to assess performance will be described next in Section 2. This includes a description of the release methods being considered. The results are then presented in Section 3 before final conclusions are provided in Section 4 together with the managerial implications, limitations, and future research directions.

2. Methodology

This study started by asking:

What is the impact of Workload Control order release in general and pure flow shops with blocking?

To answer this question, we explore the performance of Workload Control order release using a simulation model of a general flow shop and a pure flow shop (Oosterman *et al.*, 2000). Note that the pure job shop is not considered since undirected routings may lead to the mutual, or two-way, blocking of stations. We first describe how both shop types were modeled in Section 2.1 before the Workload Control order release mechanism applied is outlined in Section 2.2. Section 2.3 then outlines the dispatching rule used to prioritize jobs on the shop floor. Finally, Section 2.4 summarizes our experimental set-up and the main performance measures considered.

2.1 Shop and Job Characteristics

Our two simulation models have been implemented in the Python[®] programming language using the SimPy[®] simulation module. Both shops contain six stations, where each station is a single, constant capacity resource. For both shops, the buffer size in front of each station is limited. Five different limits are applied: 15 jobs, 20 jobs, 25 jobs, 30 jobs, and no limit (i.e. infinite jobs). To improve the generalizability of the findings, and to avoid interactions that might inhibit a full understanding of the effects of the experimental factors, we use stylized models. The five buffer limits considered in this study are set in the context of these stylized models, i.e. the tightest level is set such that the simulations remain stable and the percentage tardy within reasonable limits. As is typical for make-to-order shops, there is no finished goods inventory and jobs are delivered to the customer as soon as they have been completed.

The main difference between the general flow shop and pure flow shop is the routing of jobs. The routing length of jobs varies uniformly from one to six operations for the *general flow shop*. The routing length is first determined before the routing sequence is generated randomly without replacement; this means re-entrant flows are prohibited. The resulting routing vector (i.e. the sequence in which stations are visited) is then sorted such that the routing becomes directed and there are typical upstream and downstream stations. In contrast, for the *pure flow shop* all jobs visit all six stations in the same sequence. Operation processing times follow a truncated 2-Erlang distribution with a mean of 1 time unit after truncation and a maximum of 4 time units. The inter-arrival time of jobs to the shop follows an exponential distribution with a mean of 0.648 time units for the general flow shop and 1.111 time units for the pure flow shop. Both values deliberately result in a utilization level of 90%. Due dates are set exogenously by adding a uniformly distributed random allowance factor to the job entry time. This factor was set arbitrarily between 30 and 45 time units for the general flow shop and between 40 and 60 time units for the pure flow shop.

2.2 Workload Control Order Release

There are many order release methods in the Workload Control literature; for examples, see the reviews by Wisner (1995), Land & Gaalman (1996), Bergamaschi *et al.* (1997), and Fredendall *et al.* (2010). In this paper, the workload W_s released to a station s is kept within a pre-established workload norm via the use of a continuous release procedure that is executed whenever the system state changes (Fernandes & Carmo-Silva, 2011). In other words, all jobs are considered for release whenever an operation is complete, or whenever a new job arrives at the shop as follows:

- (1) All jobs in the set of jobs J in the pre-shop pool are sorted according to planned release dates.
- (2) The job $j \in J$ with the highest priority is considered for release first.
- (3) Take R_j to be the ordered set of operations in the routing of job j . If job j 's corrected processing time $\frac{p_{ij}}{i}$ at the i^{th} operation in its routing together with the workload W_s released to station s (corresponding to operation i) and yet to be completed fits within the workload norm N_s at this station, that is $\frac{p_{ij}}{i} + W_s \leq N_s \quad \forall i \in R_j$, then the job is selected for release. That means it is removed from J and its load contribution is included, i.e. $W_s := W_s + \frac{p_{ij}}{i} \quad \forall i \in R_j$. Otherwise, the job remains in the pool and its processing time does not contribute to the station load. A released job contributes to W_s until its operation at this station is completed. The load

contribution to a station is therefore calculated by dividing the processing time of the operation at a station by the station's position in a job's routing (Oosterman *et al.*, 2000).

- (4) If the set of jobs J in the pool contains any jobs that have not yet been considered for release, then return to Step 2 and consider the job with the next highest priority. Otherwise, the release procedure is complete and the selected jobs are released to the shop floor.

As in previous simulation studies on Workload Control (e.g. Land & Gaalman, 1998; Fredendall *et al.*, 2010; Thüerer *et al.*, 2012), it is assumed that all jobs are accepted, materials are available, and all necessary information regarding shop floor routings, processing times, etc. is known. Jobs flow into a pre-shop pool to await release according to the continuous release procedure described above. Eleven workload norms – from 5 to 15 time units – are considered. As a baseline measure, experiments without controlled order release have also been executed, i.e. where workload norms are infinite, and jobs are released onto the shop floor immediately upon arrival. Finally, the planned release date of a job is given by its due date minus an allowance for the operation throughput time for each operation in its routing. The allowance for the operation throughput time at each station is given by the cumulative moving average, i.e. the average of all operation throughput times realized until the current simulation time.

2.3 Shop Floor Dispatching Rule

Jobs in the queue or waiting for space so that they can enter the queue are prioritized according to operation due dates. The operation due date for the last operation in the routing of a job is equal to the due date, while the operation due date of each preceding operation is determined by successively subtracting an allowance for the operation throughput time from the operation due date of the next operation. In this study, the allowance for the operation throughput time at each station is given by the cumulative moving average, i.e. the average of all operation throughput times realized until the current simulation time.

2.4 Experimental Design and Performance Measures

The experimental factors are: (i) the two different shop types (the general flow shop and pure flow shop); (ii) the twelve levels of the workload norm (from 5 to 15 time units and infinite); and, (iii) the five levels of the limit on the queue (the buffer size). A full factorial design was used with 120 (2x12x5) scenarios, where each scenario was replicated 100 times. Results were collected over 10,000 time units following a warm-up period of 3,000 time units. Since we focus on a make-to-

order shop, our main performance indicator will be delivery performance. Delivery performance will be measured by: the *percentage tardy* – i.e. the percentage of jobs completed after the due date; and, the *mean tardiness*, that is $T_j = \max(0, L_j)$, with L_j being the lateness of job j (i.e. the actual delivery date minus the due date of job j). We also measure the mean of the *total throughput time* – i.e. the mean of the completion date minus the pool entry date across jobs – and, as an instrumental performance variable, the mean of the *shop floor throughput time*. While the total throughput time includes the time that an order waits before being released, the shop floor throughput time only measures the time after an order has been released to the shop floor.

3. Results

To obtain a first indication of the relative impact of the experimental factors, statistical analysis has been conducted by applying ANOVA. The results for the general flow shop and the pure flow shop are summarized in Table 1 and Table 2, respectively. All main effects and two-way interactions were shown to be statistically significant in the general flow shop (i.e. Table 1). Meanwhile, the significance of the performance effects is less pronounced in the pure flow shop (i.e. Table 2); the main effect of the buffer size is found to be not significant at $\alpha = 0.05$ for the percentage tardy and there are no significant two-way interactions. Detailed performance results will be presented next in Section 3.1 and Section 3.2 for the general flow shop and the pure flow shop, respectively. Section 3.3 then presents a discussion of the results.

[Take in Table 1 & Table 2]

3.1 Performance Assessment: General Flow Shop

The results are presented in the form of performance curves. The left-hand starting point of the curves represents the tightest workload limit (5 time units). The limit increases stepwise by moving from left to right in each graph, with each data point representing one limit (from 5 to 15 time units). Loosening the limits increases the workload on the shop floor and, as a result, increases the shop floor throughput times. The results obtained when orders are released immediately are given by single points. They are located to the right of the curves as they lead to the longest shop floor throughput times. Figure 1 shows the total throughput time, percentage tardy, and mean tardiness results over the shop floor throughput time results. The following can be observed from the results:

- *The Impact of Blocking*: The occurrence of blocking has a direct detrimental impact on shop performance. This can be observed from the results for immediate release (i.e. the single points to the right). A limit of 15 jobs at each queue increases the percentage tardy compared to the scenario without blocking (no limit) by 60% and results in a five-fold increase in mean tardiness. In practice, space constraints will often be much tighter.
- *The Impact of Workload Control Order Release*: Workload Control attenuates the impact of blocking while making significant improvements across all performance measures considered in this study. Tightening the workload norms, e.g. moving from right to left, reduces performance differences across the different levels of the buffer limit considered. In our experimental setting, performance differences across limit levels become negligible at the best performing norm level.

[Take in Figure 1]

Workload Control order release reduces the number of jobs on the shop floor, which leads to two effects. First, the risk of blocking reduces since the queue lengths at stations reduces. Second, and even if blocking occurs, the negative impact of blocking is less as blocking is more detrimental at stations with a high workload (i.e. large queue length), since these stations determine tardiness performance (Land *et al.*, 2015). Both effects can be observed from Table 3, which gives the average blocking duration, the number of occurrences per 10.000 time units, and the number of jobs in the queue at the station that is blocked. Note that we did not observe blocking at Station 5 or Station 6.

The fact that there is no blocking at Station 6 can be explained by our assumption that there is no finished goods inventory, with jobs being delivered to the customer as soon as they have been completed. Meanwhile, that there are no occurrences at Station 5 can be explained by the general decrease in the occurrence of blocking when moving downstream for loose workload norms. For tight workload norms, this is no longer true; but here the number of occurrences of blocking is, in general, reduced. Meanwhile, most of the blocking occurred when a job is to enter the shop, i.e. when there is no space at the gateway station. This is called ‘release blocking’ in this study. In fact, Workload Control order release changes the type of blocking that occurs, since an order release approach that establishes a work-in-process limit can in itself be considered a form of blocking. This will be discussed further in Section 3.3.

[Take in Table 3]

3.2 Performance Assessment: Pure Flow Shop

Figure 2 shows the total throughput time, percentage tardy, and mean tardiness results over the shop floor throughput time results for the pure flow shop. The following can be observed from our results:

- *The Impact of Blocking:* Compared to the general flow shop, the direct detrimental effect of blocking is much less. In general, there is a decrease in the frequency and duration of blocking, as can be observed from Table 4, which gives the same results as Table 3 but for the pure flow shop. One explanation is the impact of starvation in pure flow shops where the input of work to a station depends on processing times at upstream stations (Thürer *et al.*, 2015a). This will be discussed further in Section 3.3.
- *The Impact of Workload Control Order Release:* Workload Control order release again attenuates the impact of blocking while leading to significant performance improvements across all performance measures considered in this study when workload norms are set appropriately.

[Take in Figure 2 and Table 4]

3.3 Discussion of Results

Two issues emerged during the presentation of our results: the limited impact of blocking in pure flow shops and the importance of release blocking. The finding that blocking has a limited impact in pure flow shops (as modelled in our study) comes somewhat as a surprise given that most of the work on blocking in the scheduling literature focusses on flow shop environments (Hall & Sriskandarajah, 1996). To better understand the limited impact of blocking in pure flow shops compared to general flow shops, we recorded the operation throughput times for each station. Results are given in Table 5 for a queue limit of 15 jobs. In addition, Table 5 also includes the results for immediate release and no limit as a reference. Note that summing the operation throughput times for the pure flow shop in Table 5 results in the shop floor throughput times given in Figure 2. But this does not hold for the general flow shop since here routing lengths and routings vary across jobs. The results given in Figure 1 provide the mean across jobs (which have different routings and different routing lengths) meaning they cannot be directly derived from the operation throughput times at each station.

If we compare the results for immediate release with a queue limit and with no limit in the general flow shop, then we observe a significant increase in operation throughput times for upstream stations if a queue limit is applied. The increase in operation throughput times at Station 1 caused by the occurrence of blocking appears to explain most of the negative performance effect in the general flow shop. If we take a look at the results for immediate release with no limit in the pure flow shop, then we observe that the structure of the shop itself already creates longer operation throughput times at upstream stations since the input of work to a station depends on a single upstream station. Station 1 is the single gateway station and acts as a release function for downstream stations. Thus, the effect of blocking is almost negligible in our pure flow shop.

[Take in Table 5]

Meanwhile, when norms are tightened, for both shops the pattern of operation throughput times across stations in Table 5 turns around: at loose norms and for immediate release, upstream stations have larger operation throughput times than downstream stations, but for tighter norms upstream stations have shorter operation throughput times than downstream stations. This is an effect of the workload balancing mechanism (Germs & Riezebos, 2010) that is incorporated in Workload Control order release.

Our results highlighted that most of the blocking that occurred in a shop without Workload Control order release (i.e. under immediate release) is ‘release blocking’. In other words, a job cannot enter the system since there is no space in the queue at the first station in its routing. As a result, the job has to wait in a backlog or pool. This is equivalent to controlled order release where jobs also have to wait in a backlog or pool until the workload allows for their release. Workload Control uses an explicit work-in-process cap to limit the workload in the system. This means that Workload Control also ‘blocks’ the release of jobs to the system. In fact, the common element that distinguishes pull systems, such as Kanban, is a limit on in-process inventory buffers and therefore station blocking (Berkley, 1992). Thus, Workload Control order release appears to substitute one form of blocking caused by limited buffer size for another deliberate type of blocking that seeks to improve the workload balance together with other performance measures.

However, Workload Control is different from blocking in two ways: it provides a centralized release mechanism, which provides a globalized view; and, it limits the actual workload of jobs. This realizes workload balancing across stations. In contrast, blocking is similar to Kanban and

POLCA, which are decentralized and typically do not consider job workloads resulting in limited workload balancing capabilities (Germes & Riezebos, 2010). Note that another means of improving workload balancing in the context of order release is the sequence in which jobs are considered for release from the pool (Thürer *et al.*, 2015b, 2017a and 2017b). This factor does however not affect our conclusions given the relationship between the operation due date at the first station (used to prioritize jobs subject to release blocking) and planned release dates (used to prioritize jobs in the order pool).

4. Conclusions

Most shops in practice have some physical limit on the queue space (or buffer size), i.e. the number of jobs that can queue in front of a station. The Workload Control literature, however, typically assumes the infinite availability of space. This neglects the impact of blocking, which occurs in a situation where a job, having completed processing at a station, has to remain at the station until space at the next station in its routing becomes available. In response, this study started by asking: *What is the impact of Workload Control order release in general and pure flow shops with blocking?* Using simulation, we have highlighted the direct, detrimental performance effect of blocking, specifically in general flow shops. In general, we have observed an increase in the detrimental impact when routings become undirected, with the worst case being for pure job shops in which actual ‘deadlocks’ may occur. Workload Control order release enables this effect to be dampened whilst simultaneously improving shop performance. In other words, Workload Control order release is even more important in shops with a limited buffer size than it is in shops with an unlimited buffer size, which is the scenario assumed in the majority of prior studies. For example, in our study blocking increased the percentage tardy in the general flow shop by up to 60% (from 15% to 24%) compared to the scenario with an infinite buffer size (i.e. no blocking). The use of Workload Control order release reduced the percentage tardy to 5% when workload norm levels were set appropriately regardless of whether the buffer size was limited or not.

4.1 Managerial Implications

Our study has highlighted the importance of deliberate management decisions. Workload Control order release, just like any other pull system, is essentially a blocking system. The main difference between blocking caused by limits on the queue length and blocking caused by controlled order release is the load balancing capability incorporated in the latter. In other words, significant

performance improvements are obtained by substituting one form of blocking for another. A main message from our study is therefore that if there is blocking (since queue space or buffer size is limited) then a deliberate decision should be taken on how blocking shall occur and which jobs should be blocked. This decision itself may be more important than the buffer allocation decision.

In general, our study suggests that Workload Control even allows for an overall reduction in buffer size required to meet a certain performance level, which may result in significant additional savings given the cost of buffer space. However, another important managerial implication of our study is the relationship between the buffer size and the workload norm. It is apparent that a smaller buffer size requires a tighter workload norm to overcome the detrimental performance effect of blocking due to a limited buffer size. But if the workload norm becomes too tight then the workload norm itself may introduce a negative performance effect since the workload norm at one station may hinder the release of work to another station that is starving (so-called premature station idleness). At this point, it may be more advisable to actually increase buffer space to provide Workload Control with the flexibility it needs to realize workload balancing thereby reaping the full performance benefits of the Workload Control concept.

4.2 Limitations and Future Research

A main limitation of our study is its sole focus on Workload Control order release. While this is justified by our focus on high-variety make-to-order shops, future research could consider other pull systems, such as Kanban or ConWIP (Constant Work-In-Process). This however is likely to require more repetitive contexts to be modeled. Another limitation is our focus on manufacturing whereas blocking is a common phenomenon also relevant to many non-manufacturing contexts. For example, in hospitals it is often observed that although patients are ready to be discharged they cannot be released because no aftercare is available. Finally, our study highlights that blocking is a dynamic phenomenon that takes time to occur. Thus, our study calls for more research using dynamic modelling techniques to better understand the occurrence of blocking and how it can be avoided.

References

Bechte, W., 1994, Load-oriented manufacturing control just-in-time production for job shops, *Production Planning & Control*, 5, 3, 292 – 307.

- Bergamaschi, D., Cigolini, R., Perona, M., and Portioli, A., 1997, Order review and release strategies in a job shop environment: A review and a classification, *International Journal of Production Research*, 35, 2, 399-420.
- Berkley, B.J., 1992; A review of the kanban production control research literature, *Production & Operations Management*, 1, 4, 393-411.
- Buzacott, J.A., 1976, The production capacity of job shops with limited storage space, *International Journal of Production Research*, 14, 5, 597 – 605.
- Cigolini, R., and Portioli-Staudacher, A., 2002, An experimental investigation on workload limiting methods with ORR policies in a job shop environment, *Production Planning & Control*, 13, 7, 602– 613.
- Cransberg, V., Land, M.J., Hicks, C., and Stevenson, M., 2016, Handling the complexities of real-life job shops when implementing workload control: a decision framework and case study, *International Journal of Production Research*, 54, 4, 1094-1109.
- Fernandes, N.O., Thürer, M., Pinho, T.M., Torres, P., and Carmo-Silva, S., 2020, Workload Control and Optimised Order Release: An Assessment by Simulation, *International Journal of Production Research*, (in print)
- Fernandes, N.O., Land, M.J., and Carmo-Silva, S., 2016; Aligning workload control theory and practice: lot splitting and operation overlapping issues, *International Journal of Production Research*, 54, 10, 2965-2975.
- Fernandes, N.O., and Carmo-Silva, S., 2011, Workload Control under continuous order release, *International Journal of Production Economics*, 131, 257 – 262.
- Fredendall, L.D., Ojha, D., and Patterson, J.W., 2010, Concerning the theory of workload control, *European Journal of Operational Research*, 201, 1, 99-111.
- Germes, R., and Riezebos, J., 2010, Workload balancing capability of pull systems in MTO production, *International Journal of Production Research*, 48, 8, 2345-2360.
- Gershwin S.B., and Schor, J.E., 2000, Efficient algorithms for buffer space allocation, *Annals of Operations Research*, 93, 117-144.
- Gonzalez-R, P.L., Calle, M., and Andrade-Pineda, J.L., 2018, Job shop management of products under internal lifespan and external due date, *International Journal of Production Research*, 56, 16, 5457-5474.

- Haeussler, S., and Netzer, P., 2019, Comparison between rule- and optimization-based workload control concepts: a simulation optimization approach, *International Journal of Production Research*, in press.
- Hall, N.G., and Sriskandarajah, C., 1996, A survey of machine scheduling problems with blocking and no-wait in process, *Operations Research*, 44, 3, 510–525.
- Harrod, S., and Kanet, J.J., 2013, Applying work flow control in make-to-order shops, *International Journal of Production Economics*, 143, 620-626.
- Haskose, A., Kingsman, B.G., and Worthington, D., 2002, Modelling flow and jobbing shops as a queuing network for workload control, *International Journal of Production Economics*, 78, 271–285.
- Haskose, A., Kingsman, B.G., and Worthington, D., 2004, Performance analysis of make-to-order manufacturing systems under different workload control regimes, *International Journal of Production Economics*, 90, 169 – 186.
- Hendry, L.C., Huang, Y., and Stevenson, M., 2013, Workload control: Successful implementation taking a contingency-based view of production planning & control, *International Journal of Operations & Production Management*, 33, 1, 69-103.
- Henrich, P., Land, M., and Gaalman, G., 2003, Exploring applicability of the workload control concept, *International Journal of Production Economics*, 90, 187-198.
- Huang, Y., 2017, Information architecture for effective Workload Control: an insight from a successful implementation, *Production Planning & Control*, 28, 5, 351-366
- Hutter, T., Haeussler, S., & Missbauer, H., 2018, Successful implementation of an order release mechanism based on workload control: a case study of a make-to-stock manufacturer, *International Journal of Production Research*, 56, 4, 1565-1580.
- Land, M.J., Stevenson, M., Thürer, M., & Gaalman, G.J.C., 2015; Job Shop Control: In Search of the Key to Delivery Improvements, *International Journal of Production Economics*, 168, 257-266.
- Land, M.J., and Gaalman, G.J.C., 1998, The performance of workload control concepts in job shops: Improving the release method, *International Journal of Production Economics*, 56-57, 347-364.
- Land, M.J., and Gaalman, G.J.C, 1996, Workload control concepts in job shops: A critical assessment, *International Journal of Production Economics*, 46–47, 535–538.

- Leisten, R., 1990, Flowshop sequencing problems with limited buffer storage, *International Journal of Production Research*, 28, 11, 2085–2100.
- Liu, S.Q., Kozan, E., Masoud, M., Zhang, Y., and Chan, F.T.S., 2018, Job shop scheduling with a combination of four buffering constraints, *International Journal of Production Research*, 56, 9, 3274-329.
- Lödging, H., Yu, K.-W., and Wiendahl, H.-P., 2003, Decentralized WIP-oriented manufacturing control (DEWIP), *Production Planning & Control*, 14, 1, 42-54.
- Mascis, A., and Pacciarelli, D., 2002, Job-shop scheduling with blocking and no-wait constraints, *European Journal of Operational Research*, 143, 498–517.
- Oosterman, B., Land, M.J., and Gaalman, G., 2000, The influence of shop characteristics on workload control, *International Journal of Production Economics*, 68, 1, 107-119.
- Perona, M., Saccani, N., Bonetti, S., and Bacchetti, A., 2016, Manufacturing lead time shortening and stabilisation by means of workload control: an action research and a new method, *Production Planning & Control*, 27, 7-8, 660-670
- Perona, M., and Portioli, A., 1998, The impact of parameter setting in load oriented manufacturing control, *International Journal of Production Economics*, 55, 133 – 142.
- Portioli-Staudacher, A., and Tantardini, M., 2011, A lean-based ORR system for non-repetitive manufacturing, *International Journal of Production Research*, 50, 12, 3257-3273.
- Roser, C., Lorentzen, K., and Deuse, J., 2014, Reliable Shop Floor Bottleneck Detection for Flow Lines through Process and Inventory Observations, *Procedia CIRP*, 19, 63-68.
- Sabuncuoglu, I., and Karapinar, H.Y., 2000, A load-based and due-date-oriented approach to order review/release in job shops, *Decision Sciences*, 31, 2, 413–447.
- Shi, C., and Gershwin, S.B., 2016, A segmentation approach for solving buffer allocation problems in large production systems, *International Journal of Production Research*, 54, 20, 6121-6141.
- Silva, C., Stevenson, M., and Thürer, M., 2015, A Case Study of the Successful Implementation of Workload Control: A Practitioner-led Approach, *Journal of Manufacturing Technology Management*, 26, 2, 280-296.
- Stevenson, M., Hendry, L.C., and Kingsman, B.G., 2005, A review of production planning and control: The applicability of key concepts to the make to order industry, *International Journal of Production Research*, 43, 5, 869-898.

- Thürer, M., Fernandes, N.O., Stevenson, M., and Qu, T., 2017, On the Backlog-sequencing Decision for Extending the Applicability of ConWIP to High-Variety Contexts: An Assessment by Simulation, *International Journal of Production Research*, 55, 16, 4695-4711.
- Thürer, M., Fernandes, N.O., Carmo-Silva, S., and Stevenson, M., 2017b, Improving Performance in POLCA Controlled High Variety Shops: An Assessment by Simulation, *Journal of Manufacturing Systems*, 44, 143-153.
- Thürer, M., Stevenson, M., and Protzman, C.W., 2015a, COBACABANA (Control of Balance by Card Based Navigation): An Alternative to *Kanban* in the Pure Flow Shop?, *International Journal of Production Economics*, 166, 143-151.
- Thürer, M., Land, M.J., Stevenson, M., Fredendall, L.D., and Godinho Filho, M., 2015b, Concerning Workload Control and Order Release: The Pre-Shop Pool Sequencing Decision, *Production & Operations Management*, 24, 7, 1179-1192.
- Thürer, M., Stevenson, M., Silva, C., Land, M.J., Fredendall, L.D., and Melnyk, S.A., 2014, Lean control for make-to-order companies: Integrating customer enquiry management and order release, *Production & Operations Management*, 23, 3, 463-476.
- Thürer, M., Godinho Filho, M., and Stevenson, M., 2013, Coping with Finite Storage Space in Job Shops through Order Release Control: An Assessment by Simulation, *International Journal of Computer Integrated Manufacturing*, 26, 9, 830-838.
- Thürer, M., Stevenson, M., Silva, C., Land, M.J., and Fredendall, L.D., 2012, Workload control (WLC) and order release: A lean solution for make-to-order companies, *Production & Operations Management*, 21, 5, 939-953.
- Wiendahl, H.P., Gläßner, J., and Petermann, D., 1992, Application of load-oriented manufacturing control in industry, *Production Planning & Control*, 3, 2, 118 – 129.
- Wisner, J. D., 1995, A review of the order release policy research, *International Journal of Operations & Production Management*, 15, 6, 25-40.
- Xi, S., Chen, Q., MacGregor Smith, J., Mao, N., Yu, A. and Zhang, H., 2019, A new method for solving buffer allocation problem in large unbalanced production lines, *International Journal of Production Research*, in press
- Zäpfel, G. and Missbauer, H., 1993, New concepts for production planning and control, *European Journal of Operational Research*, 67, 297-320.

Table 1: ANOVA Results for the General Flow Shop

	Source of Variance	Sum of Squares	Degrees of freedom	Mean Squares	F-Ratio	p-Value
Total Throughput Time	Norm (N)	14267.75	11	1297.07	245.82	0.00
	Buffer Size (B)	2393.36	4	598.34	113.40	0.00
	N x B	2956.22	44	67.19	12.73	0.00
	Error	31342.89	5940	5.28		
Percentage Tardy	Norm (N)	13.03	11	1.18	540.60	0.00
	Buffer Size (B)	0.92	4	0.23	105.28	0.00
	N x B	0.90	44	0.02	9.30	0.00
	Error	13.02	5940	0.00		
Mean Tardiness	Norm (N)	1100.09	11	100.01	85.71	0.00
	Buffer Size (B)	484.49	4	121.12	103.80	0.00
	N x B	1064.21	44	24.19	20.73	0.00
	Error	6931.30	5940	1.17		

Table 2: ANOVA Results for the Pure Flow Shop

	Source of Variance	Sum of Squares	Degrees of freedom	Mean Squares	F-Ratio	p-Value
Total Throughput Time	Norm (N)	8945.22	11	813.20	54.64	0.00
	Buffer Size (B)	161.68	4	40.42	2.72	0.03
	N x B	80.92	44	1.84	0.12	1.00
	Error	88400.54	5940	14.88		
Percentage Tardy	Norm (N)	6.17	11	0.56	92.72	0.00
	Buffer Size (B)	0.05	4	0.01	2.21	0.07
	N x B	0.03	44	0.00	0.09	1.00
	Error	35.94	5940	0.01		
Mean Tardiness	Norm (N)	1087.53	11	98.87	28.47	0.00
	Buffer Size (B)	53.26	4	13.32	3.83	0.00
	N x B	31.78	44	0.72	0.21	1.00
	Error	20626.06	5940	3.47		

Table 3: Blocking Analysis for Buffer Size of 15 – General Flow Shop

Norm	Release		Station 1			Station 2			Station 3			Station 4		
	Dur. ¹⁾	Count ²⁾	Dur.	Count	Load ³⁾	Dur.	Count	Load	Dur.	Count	Load	Dur.	Count	Load
IMM	1.1	4326.1	1.2	381.4	12.6	1.2	246.2	8.7	1.2	169.5	7.6	1.2	85.5	6.8
N15	1.0	1912.4	1.0	356.2	10.8	1.1	241.5	8.5	1.1	161.3	7.5	1.2	82.3	6.9
N14	1.1	1222.1	1.0	326.3	9.8	1.1	231.8	8.3	1.1	158.6	7.4	1.2	81.7	6.8
N13	1.3	691.6	1.0	289.0	8.6	1.1	218.9	8.2	1.1	155.7	7.3	1.2	81.3	6.8
N12	1.7	368.5	1.0	225.2	7.3	1.0	201.1	7.5	1.1	146.8	7.1	1.2	77.7	6.7
N11	2.1	215.3	1.0	158.3	6.0	1.0	171.5	6.8	1.1	135.3	6.8	1.1	74.9	6.4
N10	2.3	132.4	1.0	89.3	5.1	1.0	129.5	5.9	1.0	117.6	6.2	1.1	68.0	6.1
N9	2.6	81.1	1.0	44.6	4.3	0.9	82.2	5.0	1.0	91.8	5.5	1.0	59.5	5.6
N8	2.7	42.7	0.9	16.6	3.6	0.9	38.7	4.1	0.9	54.5	4.5	1.0	40.7	4.6
N7	2.5	18.0	0.9	4.9	3.1	0.9	12.6	3.2	0.9	22.3	3.5	0.9	21.9	3.9
N6	2.0	4.9	0.5	0.9	1.3	0.6	2.3	2.1	0.8	4.9	2.6	0.7	7.1	3.1
N5	0.7	0.5	0.1	0.1	0.2	0.1	0.1	0.3	0.2	0.4	0.8	0.2	0.7	0.6

Dur.¹⁾ - average blockage duration; Count²⁾ - average occurrences per 10.000 time units; Load³⁾ - average number of jobs queuing at blocked station when blocking occurred

Table 4: Blocking Analysis for Buffer Size of 15 – Pure Flow Shop

Norm	Release		Station 1			Station 2			Station 3			Station 4		
	Dur. ¹⁾	Count ²⁾	Dur.	Count	Load ³⁾	Dur.	Count	Load	Dur.	Count	Load	Dur.	Count	Load
IMM	1.0	1756.4	0.7	185.3	10.4	0.7	149.3	7.5	0.7	117.2	6.7	0.7	79.1	5.9
N15	0.8	736.1	0.7	177.9	9.2	0.7	150.6	7.5	0.7	116.1	6.8	0.7	80.9	6.0
N14	0.8	450.4	0.7	173.8	8.4	0.7	149.0	7.4	0.7	114.5	6.7	0.7	79.8	6.0
N13	0.7	217.5	0.7	167.0	7.3	0.7	148.9	7.4	0.7	114.4	6.8	0.7	79.5	5.9
N12	0.7	80.8	0.7	158.9	6.1	0.7	146.9	7.2	0.7	113.3	6.7	0.7	80.5	6.1
N11	0.5	23.4	0.7	141.3	4.9	0.7	140.8	6.7	0.7	109.8	6.6	0.7	78.9	6.0
N10	0.3	4.2	0.7	106.2	3.6	0.7	130.7	6.0	0.7	105.8	6.4	0.7	76.8	5.8
N9	0.0	0.5	0.6	66.2	2.5	0.7	112.6	4.9	0.7	95.7	5.8	0.7	73.9	5.7
N8	0.0	0.0	0.6	29.9	1.7	0.7	86.0	3.6	0.7	83.0	4.9	0.7	67.9	5.3
N7	0.0	0.0	0.5	6.4	0.8	0.6	51.2	2.5	0.7	64.0	3.9	0.7	59.3	4.5
N6	0.0	0.0	0.1	0.3	0.1	0.6	16.1	1.3	0.7	35.9	2.3	0.7	41.8	3.6
N5	0.0	0.0	0.0	0.0	0.0	0.2	1.4	0.3	0.5	8.8	1.1	0.6	18.6	2.2

Dur.¹⁾ - average blockage duration; Count²⁾ - average occurrences per 10.000 time units; Load³⁾ - average number of jobs queuing at blocked station when blocking occurred

Table 5: Operation Throughput Times for Buffer Size of 15

	Norm	Station 1	Station 2	Station 3	Station 4	Station 5	Station 6
General Flow Shop	IMM (no limit)	7.60	7.00	6.76	6.68	6.48	6.48
	IMM	10.56	7.60	6.79	6.31	5.80	6.23
	N15	9.22	7.58	6.80	6.31	5.80	6.23
	N14	8.64	7.47	6.76	6.31	5.80	6.23
	N13	7.95	7.35	6.73	6.31	5.80	6.21
	N12	7.22	7.09	6.65	6.28	5.79	6.19
	N11	6.51	6.75	6.51	6.22	5.78	6.14
	N10	5.91	6.28	6.29	6.13	5.76	6.06
	N9	5.43	5.76	5.99	5.97	5.69	5.95
	N8	4.97	5.28	5.56	5.69	5.56	5.79
	N7	4.50	4.80	5.10	5.33	5.34	5.53
	N6	3.98	4.31	4.60	4.86	4.99	5.17
	N5	3.38	3.73	4.04	4.31	4.48	4.68
Pure Flow Shop	IMM (no limit)	7.79	6.22	5.99	5.77	5.53	5.52
	IMM	7.91	6.18	5.82	5.52	5.09	5.41
	N15	7.57	6.20	5.82	5.53	5.10	5.41
	N14	7.35	6.18	5.81	5.52	5.10	5.41
	N13	7.08	6.17	5.82	5.52	5.09	5.41
	N12	6.71	6.13	5.81	5.53	5.09	5.41
	N11	6.29	6.06	5.79	5.51	5.09	5.40
	N10	5.81	5.92	5.73	5.48	5.08	5.38
	N9	5.32	5.68	5.64	5.44	5.08	5.37
	N8	4.83	5.36	5.47	5.36	5.03	5.30
	N7	4.33	4.93	5.20	5.20	4.98	5.19
	N6	3.78	4.48	4.83	4.96	4.84	5.02
	N5	3.17	3.95	4.35	4.54	4.57	4.72

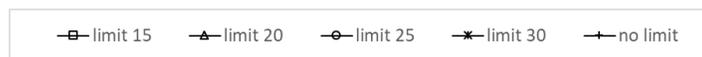
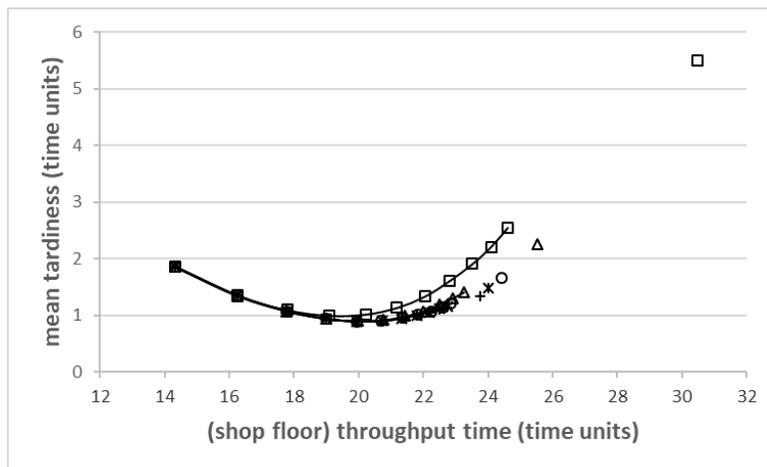
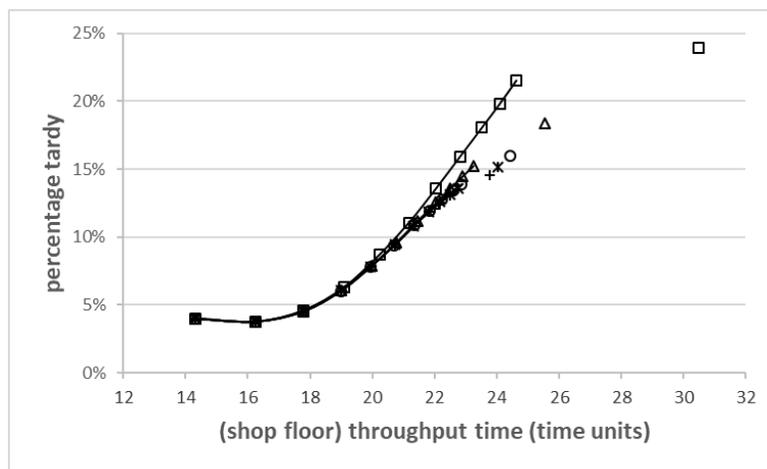
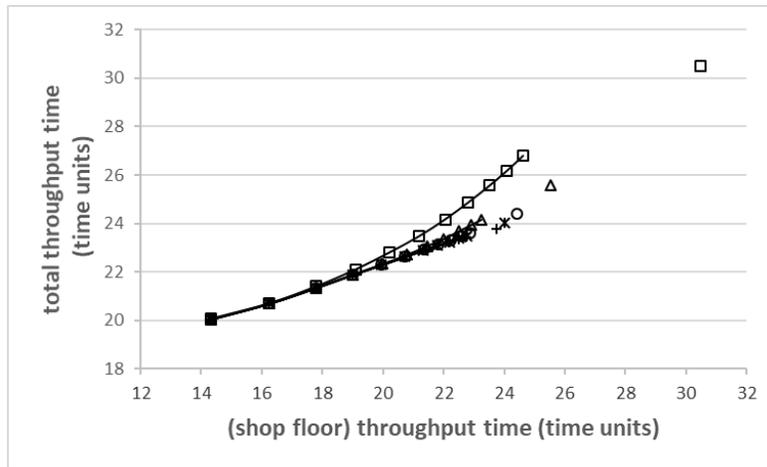


Figure 1: General Flow Shop Performance

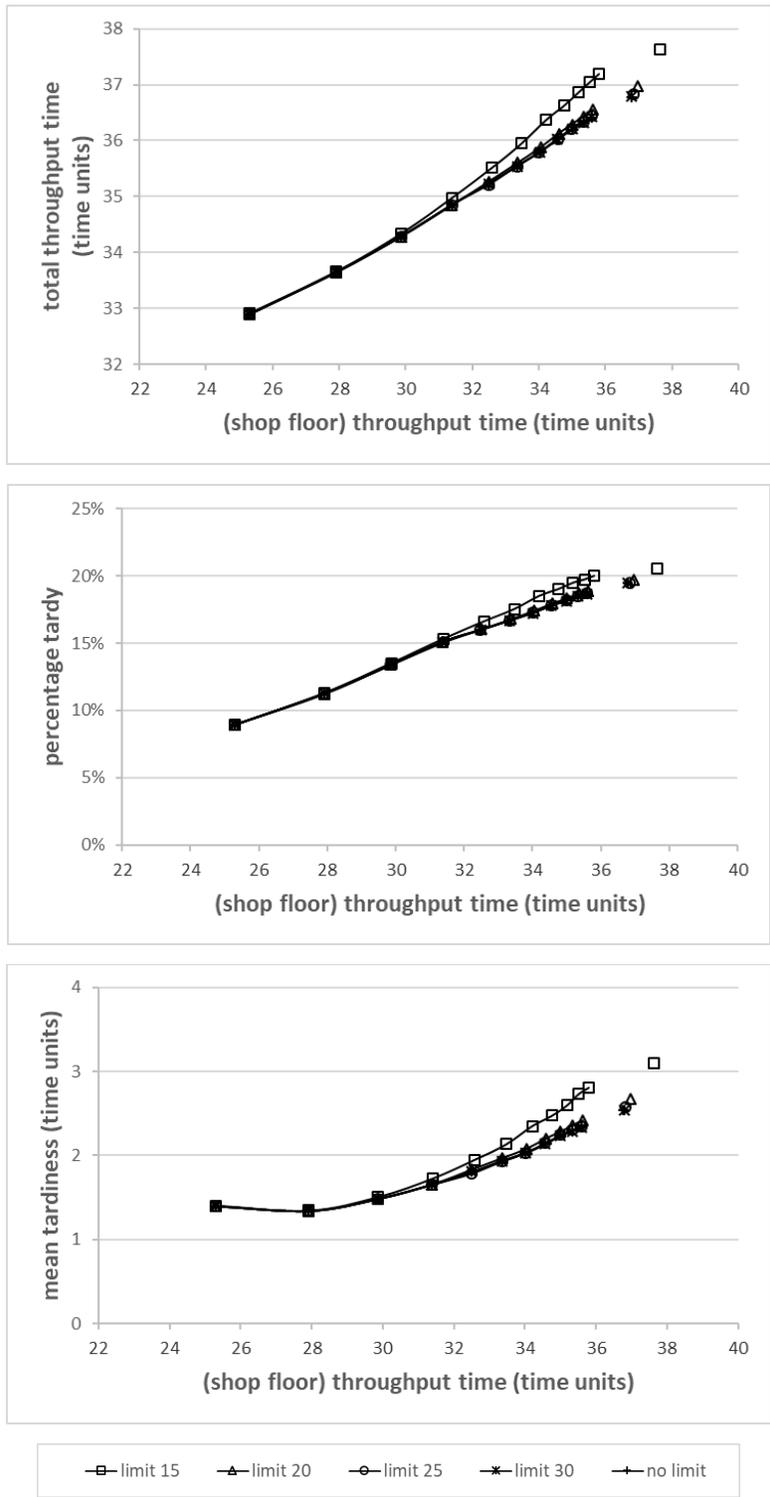


Figure 2: Pure Flow Shop Performance