International Journal of Industrial Engineering Computations 11 (2020) 429–442

Contents lists available at GrowingScience

International Journal of Industrial Engineering Computations

homepage: www.GrowingScience.com/ijiec

The use of labour flexibility for output control in workload controlled flow shops: A simulation analysis

Alberto Portioli-Staudacher[a], Federica Costa[a*] and Matthias Thürer[b]

[a]*Department of Management, Economics and Industrial Engineering, Via Lambruschini 4/b, 20156, Milano, Italy*
[b]*Institute of Physical Internet, School of Electrical and Information Engineering, Jinan University (Zhuhai Campus), 519070, Zhuhai, PR China*

C H R O N I C L E

Article history:
Received August 8 2019
Received in Revised Format November 28 2019
Accepted November 28 2019
Available online November 28 2019

Keywords:
Labour Flexibility
Workload Control
Output Control
Simulation
Flow Shop

A B S T R A C T

Workload control theory seeks to align capacity and demand to improve delivery performance. However, workload control researchers mainly focused on input control, which regulates the input of work to the production system, thereby neglecting output control, which uses capacity adjustments to regulate the outflow of the work. Moreover, few existing studies on output control investigate a temporarily increase in capacity. This paper introduces a new search direction for output control which does not require an increase in capacity – labour flexibility. Idle operators can move from their workstation to another, thus temporarily increasing the output of that workstation without extra capacity. Using simulation of a five workstations flow shop line, we highlight the positive performance effect of labour flexibility. However, this comes at the cost of high labour movement. Introducing a load-based constraint on when workers are allowed to move significantly reduces labour movement, while realizing most of the performance improvement observed for unconstrained labour movement. This has important implications for future research and practice.

1. Introduction

Due to changes in customer needs and increase of competition, customization has become the focus for more and more companies (Portioli-Staudacher & Tantardini, 2012). However, high degrees of customization can only be realized producing to-order (Linda & Kingsman, 1999; Stevenson et al., 2005; Manzini & Urgo, 2015), since highly customized orders do typically not repeat and cannot be kept in stock. A production control concept specifically developed for this kind of high variety make-to-order context is Workload Control (WLC). The concept has been shown to significantly improve production systems performances both through simulation (Kundu et al., 2018; Portioli-Staudacher & Tantardini, 2012; Thürer et al., 2012) and, on occasions, in practice. While there are several different approaches to WLC (Bergamaschi et al., 1997; Thürer et al., 2011), a major unifying principle driving WLC is input/output control (I/OC), i.e. that the input rate to a shop should be equal to the output rate (e.g. Wight, 1970; Plossl & Wight, 1971). Consequently, there are two control mechanisms within the WLC concept

* Corresponding author Tel :+57 3174420959
E-mail: federica.costa@polimi.it (F. Costa)

2020 Growing Science Ltd.
doi: 10.5267/j.ijiec.2019.11.004

(Land & Gaalman, 1996; Kingsman, 2000): i) input control (I/C), which regulates the work that can enter the shop and/or shop floor; and ii) output control (O/C), which uses capacity adjustments to regulate the outflow of work.

While I/C has received much attention in the WLC literature (Fredendall et al., 2010; Land et al., 2015; Melnyk & Ragatz, 1989; Philipoom et al., 1993; Sabuncuoglu & Karapinar, 1999; Kundu et al., 2018; Thürer et al., 2015), how O/C can be effectively realized has been largely neglected (Thürer et al., 2016). Only recently, more research has emerged that uses WLC theory to guide O/C decisions – in particular, when to *increase* capacity (Land et al., 2015; Thürer et al., 2014; Thürer et al., 2015). In this paper we present a research, that extends this recent stream of literature by investigating the impact of labour flexibility in WLC controlled shops. This means we assume that overall capacity of the system remains unchanged, but operators can temporarily be moved from one workstation to another. All variability in a production system is buffered by inventory, capacity or time (Hopp & Spearman, 2004). In a to-order context the use of an inventory buffer is limited while the lead time allowance should be as short as possible and is often determined by the customer. So, the main buffering mechanism is capacity. The capacity buffer itself consists of two components: its size and its flexibility (Hopp & Spearman 2001). Capacity flexibility has been a solution to align capacity and demand, because it provides managers the possibility of (re)allocating capacity, such as labour, according to the need (Iravani et al., 2005). i.e. from underload workstations to overload workstations. However, the impact of capacity flexibility is still widely neglected in the WLC literature. The limited existing literature is in the context of Dual Resource Constraint (DRC) shops in which capacity is constraint by both machine availability and labour availability. This literature highlights the positive performance impact of labour flexibility in context where there is less labour than machines; a relationship that is intuitively appealing (Stewart & Barrick, 2000). Meanwhile, also some research in assembly line balancing problem (ALBP) considered walking workers, who travel along the line. Theoretically, the ability of workers to walk to other workstations can bring some advantages in terms of productivity (Sikora et al., 2017). However, there is no clear explanation how labour flexibility improves performance in a shop where there is no resource constraint, labour is the main resource, and allocation of operators to workstations is dynamically managed according to the real-time workload at the workstations. In response, this research uses simulation to investigate the impact of different degrees of labour flexibility on a WLC controlled pure flow shop. Combining the recent stream of WLC literature on O/C with the stream of DRC and ALBP literature that considers labour flexibility, this research seeks to: (i) explore the impact of labour flexibility in this context; (ii) explain how labour flexibility realizes any performance improvements; and, (iii) identify solutions to take advantage of labour flexibility in practice.

The remainder of this paper is structured as follows. In Section 2, the relevant literature is reviewed and our research question introduced. The specific approach to WLC considered in this study, how labour flexibility is modelled, and the simulation model used to evaluate performance are then described in Section 3. Results are presented, discussed, and analysed in Section 4, before conclusions are summarized in Section 5.

2. Literature Review

In order to accept all customer orders arriving in a time period and still deliver them on time, companies often have to use O/C strategies, increasing their capacity for a period of time (Kingsman, 2000). The capacity is the maximum output rate and there are different ways to increase capacity; for example, overtime, worker allocation or subcontracting. The control of the output can be exerted in the short period (daily or temporarily), or over a medium-long term period (weekly or monthly). It is more likely that extra shifts and subcontracting have to be planned quite in advance, and so applied over the medium-long term period, while putting extra-workers on a workstation could be a feasible way to realize short term increases in capacity.

There are very few studies investigating the effect of O/C within the WLC theory (most of them used a job shop configuration), and, most of them do not focus on a specific O/C strategy, except for Thürer et al. (2014, 2015) who investigated how the subcontracting decision improves performance. Land et al. (2015) investigated short term increase of capacity in response to the violation of a given level of the workload. Using simulation, Land et al. (2015) showed that small but timely capacity adjustments targeted at high load periods significantly improve delivery performances. More recently, Thürer et al. (2016, 2018) combined this O/C mechanism with WLC order release and due date setting. However, a major assumption of all existing studies on O/C in the WLC literature is that capacity is increased (e.g. (Land et al., 2015), work subcontracted (e.g. Thürer et al., 2014), or work rejected (e.g. Kingsman & Hendry, 2002). This neglects an important factor often used in practice to deal with temporary overloads: labour flexibility. This will be discussed next in Section 2.1. with a specific focus on the DRC literature. Section 2.2 then shortly reviews another related research stream – ALBP. Finally, a discussion of the literature is presented in Section 2.3 where also our research question is outlined.

2.1 Labour Flexibility in DRC Shops

The term labour flexibility is used to indicate the relative ease in which workers can be relocated between organizational units (Frye, 1974). An important characteristic of the workforce is the development of multiple skills (Wallace et al., 2004; Costa et al., 2019); the larger a worker's range of skills, the more flexible is the worker, either in terms of the variety of goods and/or services he/she can produce, or in terms of the range of job assignments (Sawhney, 2013). This flexibility in turn can be used as a buffer protecting throughput from variability, specifically in production contexts with high variability (Treleven, 1989; Kher & Fredendall, 2004). Labour flexibility is typically modelled through cross-training or flexibility matrices. For example, Park and Bobrowski (1989) considered four labour flexibility matrices with an increasing level of labour flexibility in a job shop configuration (specifically, labour flexibility 1,2,3 and 5 to be intended as the number of different workstations a worker can work on). They found that performance improvement is not linearly associated with the number of different workstations at which a worker can operate. Meanwhile, Park (1991) tested an additional flexibility matrix, labour flexibility 4, and found that performances sharply increased as labour flexibility matrix changed from 1 to 2, but the increase of performance by moving from 2 to 5 was not significant. So, the minimum introduction of worker cross-training showed the most significant improvement. Brusco and Johns (1998) later showed that chaining structures obtained better performances.

2.2 Assembly Line Balancing Problem (ALBP)

This paper does not aim to review the literature on ALBP, for this the reader is referred to Gagnon and Ghosh (1991), Rekiek et al. (2002), Becker and Scholl (2006), Boysen et al. (2007) and Battaïa and Dolgui (2013). Rather we focus on multi-manned assembly lines. Amongst the different assembly line forms, multi-manned assembly line (MMAL) allows simultaneous operation of more than one worker at the same workstation (Kellegöz & Toklu, 2012; Chen, 2017; Chen et al., 2018). MMAL are commonly used in large-sized product manufacturing, and they satisfy several advantages over simple assembly lines, such as increased team-working, reduced line length, and reduced work-in-process (Dimitriadis, 2006). The balancing problem in MMAL line is more difficult since it must be determined which worker performs which tasks, besides the workstation assignment (Roshani & Giglio, 2017; Roshani et al., 2013). Some studies in the ALBP field of research addressed walking workers or travelling workers option, considering, thus, the possibility that workers travel along the line (Sikora et al., 2017; Nakade & Nishiwaki, 2008; Cevikcan, 2016; Al-Zuheri et al., 2013; Al-Zuheri et al., 2016). But, the ALBP literature mainly considers optimization algorithms used in the design phase of an assembly line with a limited number of different products, aimed at finding the allocation of tasks to workstations, and which worker has to perform them (in the case of walking workers), maximizing an objective function i.e. cycle time, number of workers etc. In contrast, we focus on dynamic worker allocation in a stochastic make-to-order environment, considering a multi-product flow shop, where allocation of worker to workstations is determined dynamically on the base of the workload present at each workstation.

2.3 Discussion of the Literature

The WLC literature on O/C is limited. Moreover, the existing literature is not concerned with the type of adjustment but simply assumes an increase in capacity. This completely neglects the possible impact of labour flexibility. For example, a temporary increase of capacity could be achieved not only by hiring another worker, but also by moving a worker from an under loaded workstation to an overloaded workstation. The former implies an increase in the overall capacity of the shop that receives the extra worker and it is subjected to the availability of extra workers. On the other hand, the latter exploits the unbalances generated in the shop and it transfers workers from under loaded workstations to overloaded workstations. In general, if a workstation is 90% utilized, the worker will be idle 10% of the time and could be temporarily used at another workstation. That there is a positive performance impact is suggested by the DRC literature, that considers labour flexibility, and by ALBP literature, that investigates how to design an assembly line under the consideration of walking workers. But this literature typically considers some resource constraint. It remains overlooked whether the impact also persists if there is enough labour. Moreover, the existing literature does not consider more advanced I/C mechanisms as presented in the WLC literature. In response, we start with the following research question:

Can labour flexibility improve performance in a make-to-order flow shop with WLC?

This study consequently differs from ALBP since it is placed in managing phase of a flow shop and it deals with the decision of dynamically allocating workers depending on their real-time workload to process at each workstation. It also differs from the DRC literature because the flow shop under analysis is not constrained by worker availability, being the number of workers equal to the number of workstations. Controlled simulation experiments will be used next to answer our research question.

3. Research Methodology

Due to a stronger spread of Lean Management and the consequent linearization of production flows, we will test our research question in a pure flow shop configuration (Kundu et al., 2018, 2019). This means, we consider companies whose production follows a dominant flow sequence. Such companies can be found, for example, in the ceramics industry and in furniture manufacturing (Portioli-Staudacher & Tantardini, 2012). We consider pure flow lines with multi manned workstations that produce large-sized products such as in the automotive or CNC machining industry (Dimitriadis, 2006). The product size is sufficient to allow two workers performing together on the same order avoiding any blocking situation. A stylized simulation model is used to prevent interactions that could interfere with the understanding of the main experimental factors. While every single shop in practice could differ from the stylized model, it captures the job and shop characteristics of MTO companies, i.e. high processing time variability and high arrival rate variability (Rossini et al., 2019). The shop and job characteristics modelled in the simulations are first summarized in Section 3.1. How I/C is modelled is then outlined in section 3.2 before labour flexibility is discussed in section 3.3. Finally, section 3.4 presents the experimental design and the performance measures.

3.1 Overview of modelled Shop and Job Characteristics

Discrete event simulation has been used as methodology since it is one of the most used techniques for analysing and understanding manufacturing systems (Negahban & Smith, 2014; Thomas et al., 2018). A simulation model of a pure flow shop has been implemented in Python using the SimPy module. We have kept our flow shop relatively small since this allows causal factors to be identified more easily. The shop is a U-shaped line composed of 5 workstations. There is one worker per workstation, so the shop is fully stuffed and no dual resource constraint exists. Considering a real case of an Italian manufacturing company leader worldwide in producing CNC machines, the five workstations of the line are the following ones: Mechanic, Hydraulic, Electric, Assembly and Testing.

Each workstation is a single resource with capacity for two workers. Each workstation can allow simultaneous operation of more than one worker at the same workstation. Maximum 2 workers are allowed to work simultaneously on the same workstation. The arrival of orders follows a Poisson distribution with an average of 1,875 orders per time unit. Operation processing times are stochastic and follow a lognormal distribution with an average processing time at each workstation of 0.5 time unit. As the context of MTO is characterized by a high variability, a coefficient of variation of 0.8 is assumed for processing times. Set-up times are considered sequence independent and part of the operation processing time. In order to keep the due date assignment method simple, a constant delivery time allowance of 56 time units is added to the order arrival date to calculate the due date. This allowance results in an average percentage of tardy order of 20%, computed across the different workload norms for the scenarios with no worker flexibility. In Table 1, shop and job characteristics are summarized.

Table 1
Shop and Job Characteristics of the simulation model

Shop Characteristics	Routing variability	Fixed
	Number of work stations	5
Job Characteristics	Arrivals	Poisson (λ =1,875 orders per time unit)
	Processing times	Lognormal; mean=0,5 time units, CV=0,8
	Due date	Due date= entry time + 56 time units

3.2 Order Review and Release

As in previous simulation studies on WLC, it is assumed that all jobs are accepted, materials are available, and all necessary information, e.g. regarding shop floor routings and processing times, are known. If order release is applied, jobs are not immediately released to the shop floor, but retained in a so-called pre-shop pool (PSP) from where they are released to meet certain performance targets. There are many order release methods in the WLC literature; for examples, see the reviews by Wisner (1995), Land and Gaalman (1996) and Bergamaschi et al. (1997). In this paper, a simple method is used that keeps the workload released but not yet completed at each workstation within limits or norms (see e.g. Oosterman et al. (2000), Thürer et al. (2011), Yan et al. (2016)). At periodic time intervals (every 8 time units) the following release procedure is executed.

1. *Job Sequencing*: All jobs in the set of jobs *J* in the PSP are sorted according to the pool sequencing rule, in this paper a First Come First Served (FCFS) dispatching rule is used (Portioli-Staudacher & Tantardini 2012). The job $j \in J$ with the highest priority is considered for release first.
2. *Job Selection*: If job *j*'s processing time p_{ij} at the ith operation at workstation *s*, together with the workload W_s released to workstation *s* and yet to be completed fits within the workload norm Ns, for each workstation, then the job *j* is selected for release. That means it is removed from *J*, and its workload contribution is included in the Ws of each workstation, otherwise the job remains in the PSP and its processing time does not contribute to Ws. The next job in the PSP is considered for release in the same way, until all jobs in the PSP are evaluated for release.

Nine different workload norms - Ns – have been used with the highest one, that corresponds to infinite release. The norm is multiplied by the workstation number to account for direct and indirect load (Oosterman et al., 2000). Once released to the shop floor, shop progress is controlled by the FCFS dispatching rule.

3.3 Labour Flexibility

We assume that the five workers are fully interchangeable. A worker is eligible for transfer when the queue β of the workstation at which she/he is working is empty (β=0). We consider that each worker has his/her "home" workstation and 4 more potential destination workstations where he/she can work when the queue of his/her "home" workstation is equal to zero time units. The worker returns to this "home" workstation when its queue increases above zero. We do not consider pre-emption, so the worker only returns after finishing the current job. This also means that it can happen that orders entering the queue of a workstation wait to be processed because the worker is away helping another workstation. The transfer of the idle worker is modelled as a temporary 50% decrease of the processing time at the workstation since the two workers are working together during the time period the additional worker is at the workstation. This decrease occurs immediately since we consider transfer time to be neglectable. Processing times are generally very long in multi-manned workstations since they are often used in the manufacturing of large-sized products (Dimitriadis, 2006). In these lines, it is generally allowed workers to pass from one workstation to another, and walking distance between workstations can be very small compared to the duration of the processing times at workstations (Şahin & Kellegöz, 2019). Note that the above decrease in processing times does not increase system capacity, since during its absence the worker's home workstation cannot process any job. Meanwhile, a workstation can receive "help" only from one additional worker, meaning that the maximum number of workers that process together a job is equal to 2. Before a job starts to be processed at a workstation, say workstation 3, the other workstations are checked and all the workers evaluated if they are idle or not, all idle workers are eligible for the transfer. The level of labour flexibility selects amongst the eligible workers and allows for the transfer to the workstation under consideration (workstation 3 in our example). In the case that there is more than one worker that could be transferred, the closest one is chosen for transfer. Four different levels of labour flexibility have been tested: Flex1, Flex2, Flex3 and Flex5. Flex1 is used as a baseline and means each worker can work on just one workstation (i.e. his/her home workstation). Flex2 allows each worker to be transferred only to the workstation immediately upstream. Flex3 allows each worker to be transferred to the workstations immediately upstream and the one immediately downstream. Finally, Flex5 allows each worker to be transferred to any workstation. In Table 2 we present four different levels tested, going from Flex1 to Flex5.

Table 2
Worker's transfers configurations according to Flex1, Flex2, Flex3 and Flex5

		Workstation1	Workstation2	Workstation3	Workstation4	Workstation5
Worker1	Flex1	√				
	Flex2	√				√
	Flex3	√	√			√
	Flex5	√	√	√	√	√
Worker2	Flex1		√			
	Flex2	√	√			
	Flex3	√	√	√		
	Flex5	√	√	√	√	√
Worker3	Flex1			√		
	Flex2		√	√		
	Flex3		√	√	√	
	Flex5	√	√	√	√	√
Worker4	Flex1				√	
	Flex2			√	√	
	Flex3			√	√	√
	Flex5	√	√	√	√	√
Worker5	Flex1					√
	Flex2				√	√
	Flex3	√			√	√
	Flex5	√	√	√	√	√

Finally, scheduling issues within each workstation, in the case of multi-manned workstation, are solved considering precedence restrictions.

3.4 Experimental Design and Performance Measures

The experimental factors are: (i) nine different workload norms and (ii) four levels of labour flexibility (from Flex1 to Flex5). A full factorial design with (9×4) 36 scenarios has been used. To reduce the variance between experiments and to focus only on the variations due to the parameters of the simulation, the common random technique was used. Results were collected over 5000 time units following a warm-up period determined through the Welch method, presented in Mahajan and Ingalls (2004), performed for all performance measures. To determine the number of runs, the Mean Square Pure Error (MSPE) has been implemented for the performance measures used. The number of runs chosen – 100 – is equal to the maximum that allows the convergence of the MSPE for all performance measures. The four main system performance measures considered in this study are the following: *Gross Throughput Time* (GTT), the time between order entry and completion*;* *Shop Floor Throughput Time* (SFTT), the time between order release from the pool and completion; *Percentage of Tardy Orders*, the percentage of orders with a positive lateness (given by the completion date minus the due date); and, *Mean Tardiness*, given by max(0;lateness).

4. Results

To aid the interpretation, the results of this study are presented in the form of performance curves. The left-hand starting point of the curves represents the lowest workload norm. The workload norm increases by moving from left to right in each graph, with each data point representing one norm level. Increasing the norm increases the level of work-in-process and, as a result, lengthens the SFTT. In Fig. 1 the GTT, the percentage of tardy jobs and the mean tardiness results are shown against the SFTT results.

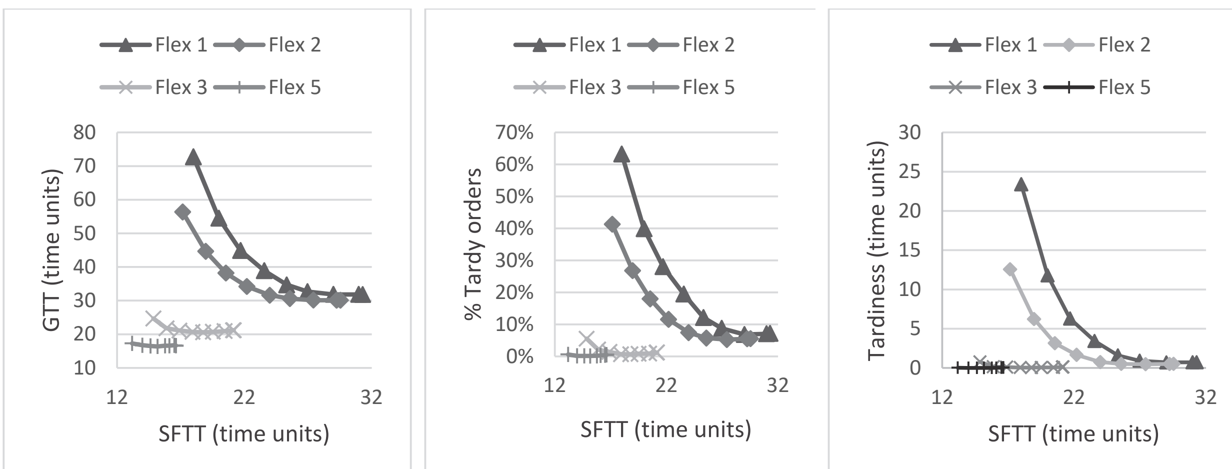

Fig. 1. Impact of Flex1, Flex2, Flex3 and Flex5 on GTT in time units, %Tardy Orders and Tardiness in time units, set against SFTT in time units on X-axis

The following can be observed from the results:

- *Impact of labour flexibility*: The impact of labour flexibility in isolation can be observed from the right starting point of the curves, which represent unrestricted release. Significant improvements across all performance measures considered in this study can be observed. Moving from Flex1 to Flex5, SFTT, and consequently GTT, are almost cut in half. This effect is obtained by using a worker at idle workstations to speed up jobs at other workstations where there is work. This in turns leads to much improved tardiness performance.
- *Impact of workload control*: This can be observed from the remaining data points on each curve. When norms are tightened, i.e. by moving from right to left on each curve, SFTT is reduced. But this may be at the expense of GTT and tardiness performance specifically when norms are tight. As recently observed in Thürer et al. (2018), the impact of O/C appears to be dominant over I/C.

Above results highlight the potential of worker flexibility. All four performance measures considered – GTT, SFTT, the percentage of tardy jobs and the mean tardiness – can be improved by postponing the idleness of workers transferring them from idle workstations to workstations with work. Note that we postpone idleness and not eliminate idleness, since helping at a workstation causes the worker at this workstation to become idle earlier.

4.1 Performance Analysis – Average Helping Time

To better understand the impact of labour flexibility, we first consider the average "helping time", i.e. the idleness time shifted. This average is computed across the WLC norms and workstations for the different levels of labour flexibility - Flex1, Flex2, Flex3 and Flex5. Results are presented in Fig. 2.

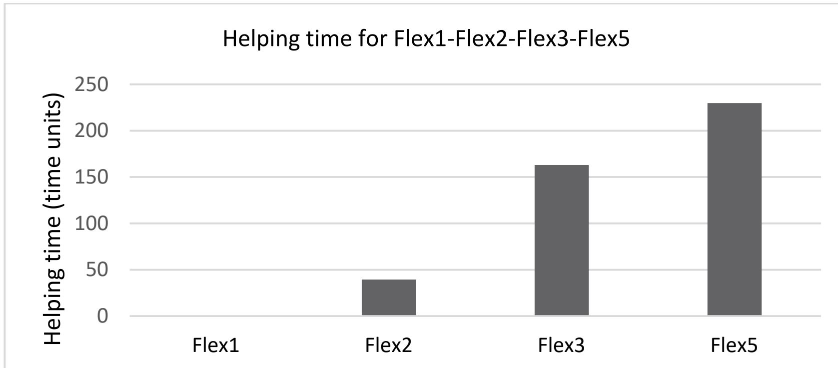

Fig. 2. Helping Time (average across all workload norms and the five workers) expressed in time units for Flex1, Flex2, Flex3 and Flex5

The increase in helping time from Flex2 to Flex3 appears to be more than double compared with the shift from Flex1 to Flex2 and from Flex3 to Flex5. This suggests that the average helping time is not linear with the labour flexibility, i.e. the number of different workstation a worker is able to work on. This finding is in line with Park and Bobrowski (1989). However, in our flow shop controlled by WLC the largest improvement is observed when there is the shift from Flex2 to Flex3, instead of shifting from Flex1 to Flex2, as suggested by Park and Bobrowski (1989).

4.2 Performance Analysis – Distribution of Helping Time over Time

The above analysis focused on the average helping time. This section focuses on the distribution of the helping time over time and how this effects performance. For this we measured the impact of Flex2, 3 and 5 on the queue length (measured in time units) of workstation 3 for unrestricted release over an arbitrary selected time period of 2400 time units, (4000 observations recorded in total). Flex1 is given as a baseline in each graph in Figure 3. Time is placed on the horizontal axis while the primary vertical axis depicts the queue length in time units. The secondary vertical axis depicts the number of workers at workstation 3, that can be equal to 0 worker, 1 worker or 2 workers. There are two important observations:

- Labour flexibility allows to reduce the duration of overload periods (duration of the queue) which explains the positive performance impact (Land et al., 2015); and,
- As Labour flexibility increases from Flex2 to Flex5 the number of peaks in queue length may increase and so does their height. Higher peaks emerge if we shift from Flex2 to Flex3 and even a higher number of peaks appears if we shift from Flex3 to Flex5. Round shapes are placed in the second graph of Fig. 3 to circle some peaks that come out passing from Flex2 to Flex3, as well for the round shapes on the third graph of Fig. 3 that circle some peaks that come out passing from Flex3 to Flex5.

The first effect is dominant over the second, particularly moving from Flex1 to Flex2, so the first effect masks the effect of the second, which is more evident moving from Flex2 to Flex3 to Flex5.

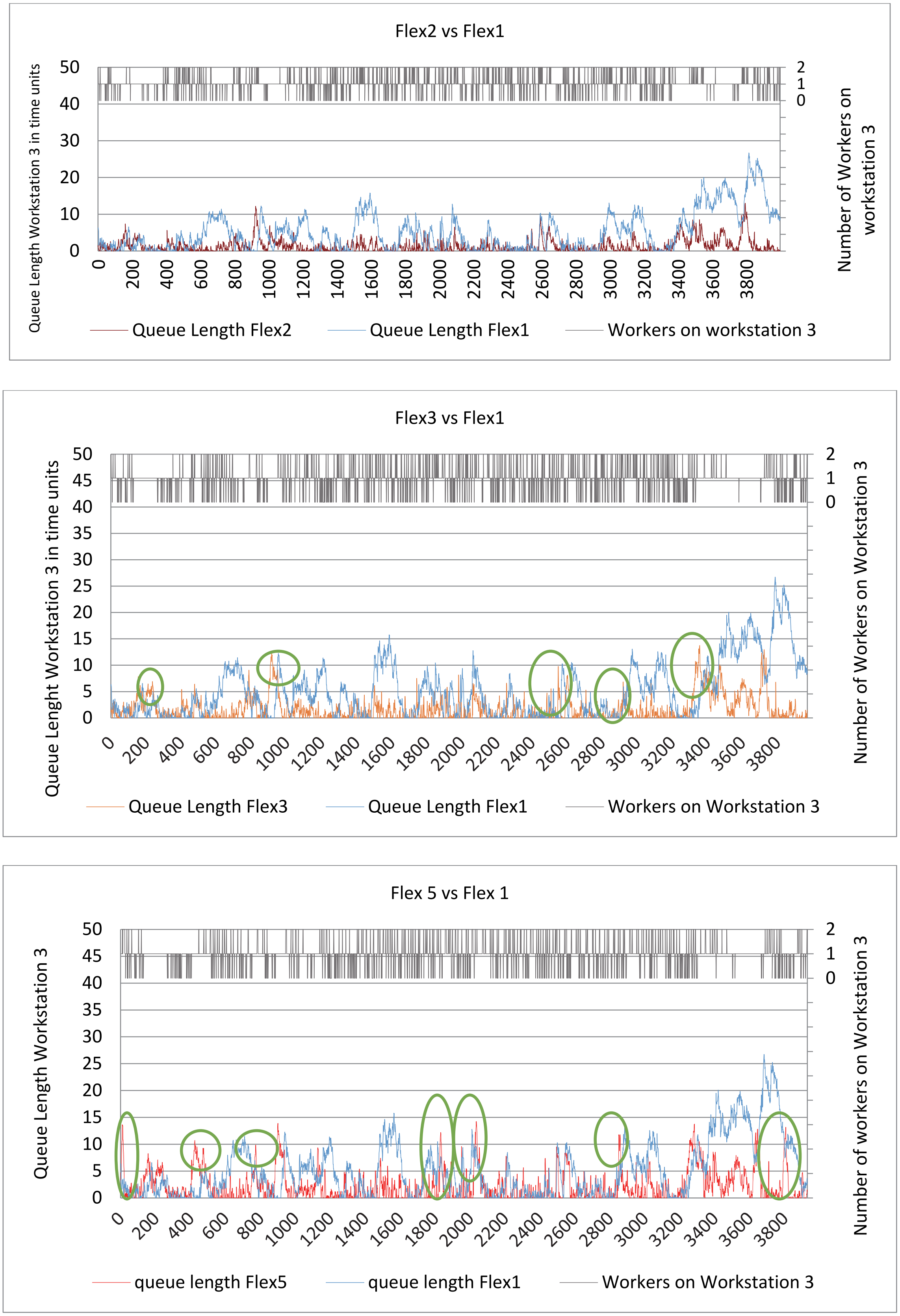

Fig. 3. For Flex2, Flex3 and Flex5, the queue length in time units of workstation 3 is presented on left Y-axis. The queue length is recorded every 0,6 time units, for a total of 4000 observations, reported on X-axis. In each graph on the right Y-axis the number of worker - that can be equal to 0 or 1 or 2 - on workstation 3 is presented. Flex1 is used as baseline in each graph

While the first effect was somewhat expected from the previous literature, the latter requires some explanations. One possible explanation would be the fact that we do not allow for preemption, which may result in a build up of the queue when the worker is away and still finsihing its job at another workstation. However, this build up cannot be more than the maximum processing time, so it does not represent an explanation. A more likely explanation is the increase of capacity at workstations upstream of an overloaded workstation, for example, if more than one workstation has an overload situation. In this case, speeding up work at the upstream workstation only aggravates the overload situation at the downstream workstation. This effect is aggravated by the pure flow shop environment considered in our study. Finally, from Fig. 3 we also notice that going from Flex2 to Flex3 to Flex5 the line of workers on workstation 3 is more nervous. In other words, the increase in average helping time observed in Section 4.1 is at the cost of a high number of worker transfers. This arguably questions the use of labour flexibility in practice.

4.3 Additional Experiments – Reducing the Number of Transfers

To make labour flexibility more practical, we introduce a refined transfer rule. This section describes this rule and presents additional experiments testing its performance impact. The transfer rule described in section 3.3 does not consider the queue length of the destination workstation. The refinement of the transfer rule, here proposed, consists in the insertion of the following condition: the queue length of the destination workstation should be above a given level. As explained in section 3.3, the worker considered eligible for the transfer, is moved to the destination workstation; with the refined transfer rule, the worker, considered eligible for the transfer, before being moved, checks the queue length of the destination workstation. If the queue length of the destination workstation is larger than a certain threshold in terms of time units, the worker is moved, otherwise not. 13 different Queue Length Thresholds (QLT), expressed in time units, have been tested. As starting value, we tested a threshold equal to zero - meaning that there is no constraint to worker's transfer in terms of queue length of the destination workstation – and a maximum threshold value that corresponds to the GTT value of Flex1– meaning that the QLT is so high that no worker's transfer is allowed. The additional experiments have been carried out considering Flex5 scenario and infinite release. Results are presented in Fig. 4, where the X-axis gives the average number of workstation's abandonments per worker in 8 time units and the Y-axis the GTT % improvement expressed as percentage improvement of Flex5 – corresponding to QLT=0 - with respect to Flex1's GTT. Worker's abandonment is intended as the number of time workers leave their home workstation.

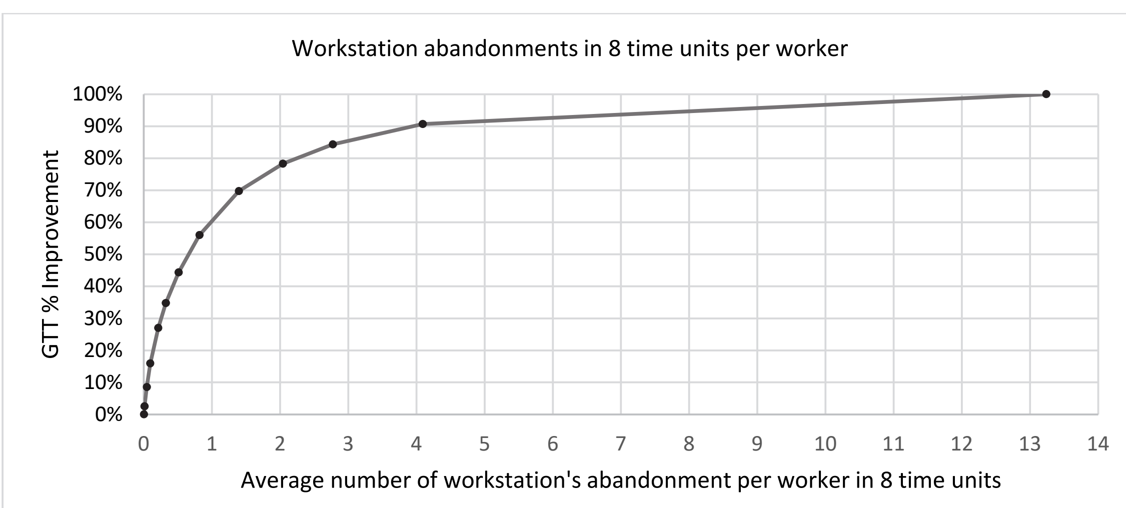

Fig. 4. GTT percentage improvement, on Y-axis, with respect to Flex1 scenario, set against the average number of workstation abandonments. The number of workstation abandonments is computed per worker in 8 time units

The right-hand starting point of the curve represents the lowest value of QLT, thus QLT=0. The QLT increases by moving from right to left on the curve in Figure 4, with each data point representing one QLT. From Fig. 4 it can be observed that with QLT=0, meaning no constraint in terms of worker's transfer, the average number of workstation's abandonments per worker in 8 time units is equal to 13 and the GTT improvement with respect to Flex1 is maximum. However, increasing the QLT, with the refined transfer rule, it is possibile to have a more reasonable number of workstation's abandonments per worker, retaining anyway a GTT percentage improvement. In particular, if we decrease the number of workstation's abandonments per worker to 2 in 8 time units, we observe that we still obtain 80% of the improvement in GTT compared to Flex1. With an average of 4 abandonments per worker, we retain 90% of the improvement. With the inclusion of the refined transfer rule that takes into account the queue length of the destination's workstation, we observe that we obtain viable and more applicable results, since we limit the number of worker's transfer. The number of workstation's abandonments are drastically reduced with the refined transfer rule applied, however with a GTT % improvement retained with respect to Flex1. Indeed we retain 90% GTT improvement, lowering the average workstation's abandonments from 13 to 4 per worker in 8 time units. These are relevant results for practical application.

5. Conclusions, limitations and future research

The main objective of WLC was to align the demand to the capacity. This was realized by two different mechanisms: I/C that regulates the input/flow to a production system and O/C that controls the output, acting on the capacity of a production system. While I/C has been largely investigated in the WLC literature, few studies has assessed the impact of O/C within WLC. Moreover, all of the existing literature address O/C by considering an increase in the overall capacity of the production system under analysis. An alternative solution was suggested in the literature on DRC and ALBP: labour flexibility. However, this research assumes less labour than machines and it widely neglects Workload Control. In response, we asked: *Can labour flexibility improve performance in a make-to-order flow shop with WLC?* Using simulation, we have shown that significant improvements in performances could be achieved without increasing the overall capacity of the system, rather creating and exploiting labour flexibility, and temporarily moving workers that are idle to help workers in overloaded workstation. Higher flexibility levels lead to an increase in helping time which in turn leads to a stronger reduction in the duration of overload periods. But this positive effect comes with two negative side effects. First, increasing capacity at workstations upstream of an overloaded station may lead to an aggravation of the overload situation. Second, helping time is not continuous, but rather a high amount of labour transfers occurs. Specifically, the latter questions the use of labour flexibility in practice. In response, a simple load threshold was introduced, which allowed to significantly reduce the amount of labour abandonments from a workstation while retaining most of the performance benefits. A first important limitation of our study is our focus on a pure flow shop. While this is justified by the practical relevance of this shop type in the context of multi-manned work stations, future research is needed to extend our findings to more complex shops. Another important limitation of our study is that we consider workers to be fully interchangeable thereby neglecting potential differences in worker efficiency. We are currently working on the modelling of more complex transfers rule that could take into consideration different factors such as worker's different efficiency, heterogeneous labour, learning and forgetting (Renna, 2019), and different degrees of labour flexibility. However, this research paper has been intended to open a new field of research aiming at investigating labour flexibility as a new output control mechanism that improves performances without increasing the overall capacity of the system.

Acknowledgements

This work was supported by Lean Excellence Centre, observatory of Politecnico di Milano, www.lean.polimi.it.

References

Al-Zuheri, A., Luong, L., & Xing, K. (2013). Impact of walking worker assembly line configuration on ergonomics performance. *Australian Journal of Multi-Disciplinary Engineering*, *9*(2), 165-175.

Al-Zuheri, A., Luong, L., & Xing, K. (2016). Developing a multi-objective genetic optimisation approach for an operational design of a manual mixed-model assembly line with walking workers. *Journal of intelligent manufacturing*, *27*(5), 1049-1065.

Battaïa, O., & Dolgui, A. (2013). A taxonomy of line balancing problems and their solution approaches. *International Journal of Production Economics, 142*(2), 259–277.

Becker, C., & Scholl, A. (2006). A survey on problems and methods in generalized assembly line balancing. *European Journal of Operational Research*, *168*(3), 694-715.

Bergamaschi, D., Cigolini, R., Perona, M., & Portioli, A. (1997). Order review and release strategies ina job shop environment: A review and a classification. *International Journal of Production Research*, *35*(2), 399-420.

Boysen, N., Fliedner, M., & Scholl, A. (2007). A classification of assembly line balancing problems. *European Journal of Operational Research*, *183*(2), 674-693.

Brusco, M.J., & Johns, T.J. (1998). Staffing a multiskilled workforce with varying levels of productivity: An analysis of cross-training policies. *Decision Sciences,* 29(2), 499–515.

Cevikcan, E. (2016). An optimization methodology for multi model walking-worker assembly systems: an application from busbar energy distribution systems. *Assembly Automation*, *36*(4), 439-459.

Chen, Y. Y. (2017). A hybrid algorithm for allocating tasks, operators, and workstations in multi-manned assembly lines. *Journal of Manufacturing Systems*, *42*, 196-209.

Chen, Y. Y., Cheng, C. Y., & Li, J. Y. (2018). Resource-constrained assembly line balancing problems with multi-manned workstations. *Journal of Manufacturing Systems*, *48*, 107-119.

Costa, F., Lispi, L., Staudacher, A. P., Rossini, M., Kundu, K., & Cifone, F. D. (2019). How to foster Sustainable Continuous Improvement: A cause-effect relations map of lean soft practices. *Operations Research Perspectives*, *6*, 100091.

Dimitriadis, S. G. (2006). Assembly line balancing and group working: A heuristic procedure for workers' groups operating on the same product and workstation. *Computers & Operations Research*, *33*(9), 2757-2774.

Fredendall, L.D., Ojha, D., & Patterson, J. W. (2010). Concerning the theory of workload control. *European Journal of Operational Research* 201 (1). 99–111.

Frye, J. S. (1974). Labor flexibility in multiechelon dual-constraint job shops. *Management Science* 20 (7), 1073–1081.

Gagnon, R. J., & Ghosh, S. (1991). Assembly line research: Historical roots, research life cycles and future directions. *Omega,* 19(5), 381–399.

Hopp, W.J., & Spearman, M.L. (2001). *Factory Physics: Foundations of Manufacturing Management*. 2nd ed. Boston, MA: Irwin/McGraw-Hill.

Hopp, W. J., & Spearman, M.L. (2004). To pull or not to pull: What is the question?. *Manufacturing & Service Operations Management,* 6(2), 133–148.

Hopp, W. J., Tekin, E., & Van Oyen, M. P. (2004). Benefits of skill chaining in serial production lines with cross-trained workers. *Management Science,* 50(1), 83–98.

Iravani, S.M., Van Oyen, M.P., & Sims, K.T. (2005). Structural flexibility: A new perspective on the design of manufacturing and service operations. *Management Science, 51*(2), 151–166.

Kellegöz, T., & Toklu, B. (2012). An efficient branch and bound algorithm for assembly line balancing problems with parallel multi-manned workstations. *Computers and Operations Research, 39*(12), 3344–3360.

Kher, H.V., & Fredendall, L.D. (2004). Comparing variance reduction to managing system variance in a job shop. *Computers and Industrial Engineering, 46*(1), 101–120.

Kingsman, B.G. (2000). Modelling input-output workload control for dynamic capacity planning in production planning systems. *International Journal of Production Economics,* 68(1), 73–93.

Kingsman, B.G., & Hendry, L.C. (2002). The relative contributions of input and output controls on the performance of a workload control system in make-to-order companies. *Production Planning and Control, 13*(7), 579–590.

Kundu, K., Rossini, M., & Portioli-Staudacher, A. (2018). Analysing the impact of uncertainty reduction on WLC methods in MTO flow shops. *Production & Manufacturing Research, 6*(1), 328–344.
Kundu, K., Rossini, M & Portioli-Staudacher, A. (2019). A study of a kanban based assembly line feeding system through integration of simulation and particle swarm optimization. *International Journal of Industrial Engineering Computations*, *10*(3), 421-442.
Land, M.J., & Gaalman, G.J.C. (1996). Workload control concepts in job shops a critical assessment. *International Journal of Production Economics, 46–47*, 535–548.
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.
Linda, G.A., & Kingsman, H.B. (1999). Competitive advantage, customisation and a new taxonomy for non make-to-stock companies. *International Journal of Operations & Production Management, 19* (4), 349–371.
Mahajan, S., & Ingalls, R.G. (2004). Evaluation of methods used to detect warm-up period in steady state simulation. In *Proceedings of the 2004 Winter Simulation Conference*.
Manzini, M., & Urgo, M. (2015). Makespan estimation of a production process affected by uncertainty: Application on MTO production of NC machine tools. *Journal of Manufacturing Systems* 37. 1–16.
Melnyk, S.A. & Ragatz, G.L. (1989). Order Review_release, Research Issues and Perspectives. *International Journal of Production Research, 27* (7), 1081–1096.
Nakade, K., & Nishiwaki, R. (2008). Optimal allocation of heterogeneous workers in a U-shaped production line. *Computers & Industrial Engineering*, *54*(3), 432-440.
Negahban, A., & Smith, J.S. (2014). Simulation for manufacturing system design and operation: Literature review and analysis. *Journal of Manufacturing Systems, 33*(2), 241–261.
Oosterman, B., Land, M.J., & Gaalman, G.J.C. (2000). Influence of shop characteristics on workload control. *International Journal of Production Economics, 68*(1), 107–119.
Park, P.S., & Bobrowski, P.M. (1989). Job release and labor flexibility in a dual resource constrained job shop. *Journal of Operations Management, 8*(3), 230–249.
Park, P.S. (1991). The examination of worker cross-training in a dual resource constrained job shop. *European Journal of Operational Research, 52*(3), 291–299.
Philipoom, P.R., Malhotra, M.K., & Jensen, J.B. (1993). An evaluation of capacity sensitive order review and release procedures in job shops. *Decision Sciences, 24*(6), 1109–1134.
Plossl, G., & Wight, O. (1971). Materials requirements planning by computer. (American Production and Inventory Control Society, Washington, DC).
Portioli-Staudacher, A., & Tantardini, M. (2012). A lean-based ORR system for non-repetitive manufacturing. *International Journal of Production Research,* 50(12), 3257–3273.
Rekiek, B., Delchambre, A., Dolgui, A. & Bratcu, A. (2002). Assembly line design : a survey. IFAC.
Renna, P. (2019). Flexible job-shop scheduling with learning and forgetting effect by Multi-Agent System. *International Journal of Industrial Engineering Computations*, *10*(4), 521-534.
Roshani, A., & Giglio, D. (2017). Simulated annealing algorithms for the multi-manned assembly line balancing problem: minimising cycle time. *International Journal of Production Research*, *55*(10), 2731-2751.
Roshani, A., Roshani, A., Roshani, A., Salehi, M., & Esfandyari, A. (2013). A simulated annealing algorithm for multi-manned assembly line balancing problem. *Journal of Manufacturing Systems, 32*(1), 238.
Rossini, M., Audino, F., Costa, F., Cifone, F.D., Kundu, K., & Portioli-Staudacher, A. (2019). Extending lean frontiers : A kaizen case study in an Italian MTO manufacturing company. The International Journal of Advanced Manufacturing Technology, 1-20.
Sabuncuoglu, I., & Karapınar, H. Y. (1999). Analysis of order review/release problems in production systems. *International Journal of Production Economics*, *62*(3), 259-279.
Şahin, M., & Kellegöz, T. (2019). Balancing multi-manned assembly lines with walking workers: Problem definition, mathematical formulation, and an electromagnetic field optimisation algorithm. *International Journal of Production Research*. doi:10.1080/00207543.2019.1566672.
Sawhney, R. (2013). Implementing labor flexibility: A missing link between acquired labor flexibility

and plant performance. *Journal of Operations Management, 31*(1–2), 98–108.
Sikora, C.G.S., Lopes, T.C. & Magatão, L. (2017). Traveling worker assembly line (re)balancing problem: Model, reduction techniques, and real case studies. *European Journal of Operational Research,* 259(3), 949–971.
Stevenson, M., Hendry, L.C., & Kingsman, B.J. (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.
Stewart, G.L., & Barrick, M.R. (2000). Team structure and performance: Assessing the mediating role of intrateam process and the moderating role of task type. *Academy of Management Journal, 43*(2), 135–148.
Thomas, T., Sherman, S.R., & Sawhney, R.S. (2018). Application of lean manufacturing principles to improve a conceptual 238Pu supply process. *Journal of Manufacturing Systems, 46,* 1–12.
Thürer, M., Land, M.J., Stevenson, M, Fredendall, L.D., & Godinho Filho, M. (2015). Concerning workload control and order release: The pre-shop pool sequencing decision. *Production and Operations Management, 24*(7), 1179–1192.
Thürer, M., Silva, C., & Stevenson, M. (2011). Optimising workload norms: the influence of shop floor characteristics on setting workload norms for the workload control concept. *International Journal of Production Research*, *49*(4), 1151-1171.
Thürer, M., Stevenson, M., & Land, M.J. (2016). On the integration of input and output control: Workload control order release. *International Journal of Production Economics, 174*, 43–53.
Thürer, M., Stevenson, M., Land, M. J., & Fredendall, L. D. (2019). On the combined effect of due date setting, order release, and output control: an assessment by simulation. *International Journal of Production Research*, *57*(6), 1741-1755.
Thürer, M., Stevenson, M., & Qu, T. (2015). Simple subcontracting rules for make-to-order shops with limited subcontractor capacity: An assessment by simulation."*Production Planning and Control, 26* (13), 1145–1161.
Thürer, M., Stevenson, M., Qu, T., & Godinho Filho, M. (2014). The design of simple subcontracting rules for make-to-order shops: An assessment by simulation. *European Journal of Operational Research, 239*, 854–864.
Thürer, M., Stevenson, M., Silva, C., Land, M.J. & Fredendall, L.D. (2012). Workload control and order release: A lean solution for make-to-order companies. *Production and Operations Management, 21* (5), 939–953.
Treleven, M. (1989). A review of the dual resource constrained system research. In *IIE Transactions*, *21*(3), 279-287.
Wight, O. (1970). Input/output control a real handle on lead time. *Production Inventory Management Journal, 11*(3), 9–31.
Wisner, J. D. (1995). A review of the order release policy research. *International Journal of Operations & Production Management, 15*(6), 25–40.
Yan, H., Stevenson, M., Hendry, L.C., & Land, M.J. (2016). Load-oriented order release (LOOR) revisited: Bringing it back to the state of the art. *Production Planning and Control, 27*(13), 1078–1091.