On the combined effect of due date setting, order release, and output control: an assessment by simulation

Matthias Thürer, Mark Stevenson, Martin J. Land & Larry D. Fredendall

To cite this article: Matthias Thürer, Mark Stevenson, Martin J. Land & Larry D. Fredendall (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, DOI: 10.1080/00207543.2018.1504250

To link to this article: https://doi.org/10.1080/00207543.2018.1504250

Published online: 03 Aug 2018.

Submit your article to this journal

Article views: 125

View Crossmark data

Citing articles: 1 View citing articles
On the combined effect of due date setting, order release, and output control: an assessment by simulation

Matthias Thürera*, Mark Stevensonb, Martin J. Landc and Larry D. Fredendallld

aSchool of Electrical and Information Engineering, Jinan University, Institute of Physical Internet, Zhuhai, People's Republic of China; bDepartment of Management Science, Lancaster University, Lancaster University Management School, Lancashire, UK; cDepartment of Operations, Faculty of Economics and Business, University of Groningen, Groningen, Netherlands; dDepartment of Management, Clemson University, Clemson, SC, USA

(Received 30 March 2018; accepted 28 June 2018)

Workload Control is a production control concept for high-variety shops built on the principle of input/output control. The literature, however, has argued that input/output control overemphasises throughput improvements to the detriment of the timing of individual orders and, consequently, that it needs to be supplemented by a preceding customer enquiry stage where due dates are set. Yet, although there are broad separate literature streams on due date setting, order release, and output control, there is a lack of research on the three functions together. In response, this study uses simulation to assess the combined performance effect of all three functions. Results show that each control function can be related to a specific performance objective. The degree of emphasis that should be placed on each function, therefore, depends on a company’s specific performance needs. Due date setting and capacity adjustments (output control) are shown to support each other as they address different performance objectives. Meanwhile, order release (input control) is effective in reducing work-in-process and can play a role in making throughput improvements when capacity adjustments are not possible. Findings enhance existing literature on the diagnosis of delivery reliability performance in high-variety shops, with important implications for research and practice.

Keywords: Workload Control; due date assignment; order release; capacity planning; job shop

1. Introduction

Workload Control is a production planning and control concept developed for high-variety contexts (Zäpfel and Missbauer 1993; Stevenson, Hendry, and Kingsman 2005) that has been shown to significantly improve the performance of job shops both through simulation (e.g. Thürer et al. 2012, 2014) and in practice (e.g. Wiendahl, Gläßner, and Petermann 1992; Bechte 1994; Hendry, Huang, and Stevenson 2013; Silva, Stevenson, and Thürer 2015; Hutter, Haeussler, and Missbauer 2018). In order to realise the principle of input/output control, i.e. that the input rate to the shop should be equal to the output rate (e.g. Wight 1970; Plossl and Wight 1971), Workload Control integrates two control mechanisms (e.g. Land and Gaalman 1996; Kingsman 2000): (i) input control, which regulates the work that can enter the shop and/or shop floor, and (ii) output control, which uses capacity adjustments to regulate the outflow of work. Furthermore, there are two performance objectives that together determine delivery performance in make-to-order job shops (e.g. Land 2006): (i) throughput improvement, which aims to reduce the average lateness of orders, and (ii) timing, which aims to reduce the variance of lateness among orders. As noted by Kingsman, Tatsiopoulos, and Hendry (1989), input/output control mechanisms have a strong bias towards throughput time stabilisation and thus towards improving throughput performance. This overlooks the important role of the timing function and due date setting, the latter specifically if customer due dates can be determined internally by the company, i.e. are negotiable. In response, this study makes a contribution to the literature by using simulation to assess the combined effect of due date setting, order release, and output control.

According to Bertrand and Wortmann (1981), there are three main production control functions in job shops: due date setting, order release (input control), and output control. However, while there exists a broad literature on the different functions in isolation, literature that assesses the combined effect of the different control functions is scarce. Few studies, for example, have examined the interactions between due date setting and order release control. Bertrand (1983a) argued that controlled order release does not have a significant impact on performance if an effective due date setting procedure is applied. Ahmed and Fisher (1992) later found that the true impact of order release depends on the set of due date setting

*Corresponding author. Email: matthiasthurer@workloadcontrol.com

© 2018 Informa UK Limited, trading as Taylor & Francis Group
and order release rules applied, but the authors did not consider the procedure described in Bertrand (1983a). More recently, Thürer et al. (2014) showed that an effective due date setting rule and order release control can and should play complementary roles in an integrated Workload Control system, while Thürer, Stevenson, and Land (2016a) showed that output control can and should complement order release. Hence, there are a few papers that have examined pairs of control functions; but to the best of our knowledge, no study has examined the combined effect of all three control functions together. This is a major shortcoming since it remains unknown whether all three control functions should be implemented together or if a subset is sufficient (as argued in Bertrand 1983a), and, if so, which subset should be implemented. Addressing this research need is, therefore, the first objective of our study.

A second objective of our study is to link the control functions to throughput and timing performance objectives. Soepenberg, Land, and Gaalman (2012) recently argued that a diagnosis of delivery reliability performance is vital if production planning and control decisions are to be improved. The authors outlined a diagnosis framework that allows poor delivery reliability to be attributed to poor throughput or timing capabilities. The present study seeks to enhance this framework by providing an indication of which control function to use to improve delivery reliability. To summarise, our study contributes to the existing literature in two ways. First, by assessing the combined performance impact of all three control functions: due date setting, order release, and output control. Second, by relating the three control functions to the performance objectives thereby supporting managers in finding the right trade-off between the degree to which a control function is implemented in line with the costs, the additional complexity, and any resulting performance improvements (Cransberg et al. 2016).

The remainder of this paper is structured as follows. In Section 2, we first discuss the Workload Control literature, outlining Workload Control and its control functions as applied in our study. The simulation model used to evaluate performance is then described in Section 3. The results are presented, discussed, and analysed in Section 4 before conclusions are provided in Section 5, where managerial implications, limitations, and future research directions are also outlined.

2. Workload control

This study started by asking:

What is the combined performance effect of due date setting, order release control, and output control?

To answer this question, we explore the performance of a comprehensive Workload Control concept in a pure job shop using controlled simulation experiments. Workload Control was chosen since it integrates all three production control functions, as illustrated in Figure 1. The relevant literature on due date setting, order release control, and output control will be discussed in Sections 2.1–2.3, respectively. Here, we also outline how each control function is modelled in our study. The simulation model follows the flow of orders, as shown horizontally in Figure 1, while it contains modules to take the decisions at each stage of the order flow, as shown vertically in Figure 1.

2.1. Due date setting

The due date is the date when the order is placed plus a lead time allowance (i.e. the time that the customer is willing to wait). In terms of setting lead time allowances, two types of jobs can be identified: (i) jobs where the lead time allowance is proposed or quoted by the company and, therefore, negotiable, and (ii) jobs where the lead time allowance is specified by the customer and, therefore, relatively fixed (e.g. Ragatz and Mabert 1984; Cheng and Gupta 1989; Kingsman 2000). Five different scenarios are modelled to assess the effect of the due date setting rule: 100%, 75%, 50%, 25% of due dates set using the internal due date setting rule, and no due dates set internally (i.e. 0% of due dates set by the due date setting rule). The probability that a due date can be set internally for an enquiry is modelled as a Bernoulli trial.

2.1.1. Internally (or endogenously) set due dates

When a new job $j$ arrives, a feasible due date ($d_j$) is determined via forward scheduling whereby the following three elements – which constitute the lead time allowance – are added to the current time $t$ (see Equation (1)): an allowance $a_j$ for the time that a job has to wait in the pre-shop pool prior to release; an allowance $b_{ij}$ for the operation throughput time of each operation $i$ in the routing $R_j$ of a job to allow for the shop floor throughput time; and an external allowance $c_j$ that compensates for
variability between the estimated lead time and the delivery time that is ultimately realised.

\[ d_j = t + a_j + \sum_{i \in R_j} b_{ij} + c_j \]  

(1)

The Workload Control literature on due date setting typically assumes that jobs are released immediately, i.e. that the pool waiting time \( a_j \) is zero (e.g. Ahmed and Fisher 1992; Enns 1995a) or constant for all jobs (Hendry, Kingsman, and Cheung 1998; Thürer et al. 2013, 2014). An exception is Land (2009) who presented a method of estimating a dynamic allowance for the pool waiting time, which will be used in this study. Following Little’s Law (1961), Land (2009) estimated the pool waiting time based on the total processing time units waiting in the pool to be released to the station that is most likely to restrict the release of a job, i.e. the station that has the largest load waiting to be released across the stations in the routing of a job. The pool waiting time is given by the quotient of this maximum pool load and the maximum output of the station.

Most due date setting rules presented in the Workload Control literature differ in the way that allowances are determined for the shop floor throughput time. For example, forward infinite loading assumes operation throughput times are constant (e.g. Weeks 1979; Ragatz and Mabert 1984; Vig and Dooley 1993). Meanwhile, other studies link the processing time and shop load to the delivery time based on historical data via regression (e.g. Ragatz and Mabert 1984; Ahmed and Fisher 1992; Vig and Dooley 1993; Moses et al. 2004) or link the workload at a station to the allowance for the operation throughput time (e.g. Nyhuis and Wiendahl 2009).

Bertrand (1983a, 1983b) determined a dynamic allowance for operation throughput times by fitting a cumulative input curve to a cumulative output curve. Operation due dates \( d_{ij} \) for each operation \( i \) in the routing of a job \( j \), where \( d_{0j} \) is defined as the current date, are successively scheduled using the time-phased accepted workload \( W_{st}^j \) and time-phased capacity \( C_{st} \) of the corresponding station \( s \) — both measures calculated cumulatively up to time bucket \( t \) — as follows. Starting with the first station in the routing of a job:

- If the time bucket into which the operation due date would fall if capacity were infinite — that is \( d_{ij} = d_{i-1,j} + p_{ij} \) — has enough free capacity to include the workload \( p_{ij} \) of the \( i^{th} \) operation of job \( j \) at the relevant station \( s \) — that is
In addition to the above periodic release mechanism, LUMS COR incorporates a 

$$W_{st}^R + p_{ij} \leq C_{st} \cdot u_s$$ with \(u_s\) equal to the utilisation rate – then the operation is loaded into the time bucket and the operation due date is given by this time bucket.

- If no or insufficient capacity is available, the next time bucket \(t+1\) is considered until the workload has been successfully loaded.

This procedure is then repeated at the next station in a job’s routing until all operation due dates have been determined. An operation remains loaded into a time bucket – and thus contributes to the cumulative workload – until it has been completed. This forward finite loading procedure was recently identified as the best solution for the Workload Control concept (see, e.g. Thürer et al. 2013) and will thus be included in our study to set allowances for operation throughput times. The time buckets are arbitrarily set to one time unit.

Finally, the external allowance \(c_j\) is often included in the allowance for operation throughput times in the literature (Hopp and Sturgis 2000). Notable exceptions that have differentiated between an internal (or production) due date and an external (or customer) due date – which is the internal due date plus the external allowance – were presented by Bertrand (1983a), Enns (1995b), and Hopp and Sturgis (2000). The latter compared the use of a constant external allowance with the use of alternative, dynamic external allowances. Numerical results suggested that there are no significant performance differences. Therefore, a constant allowance will be used. This allowance is set through preliminary simulation experiments such that the average of the lead time allowance is 32 time units, i.e. the average of (exogenously set) lead time allowances, as will be described next.

### 2.1.2. Externally (or exogenously) set due dates

Due dates are set exogenously by adding a random lead time allowance, uniformly distributed between 28 and 36 time units, to the job entry time. The minimum value will be sufficient to cover a minimum shop floor throughput time corresponding to the maximum processing time (4 time units) for the maximum number of possible operations (6) plus an allowance for the waiting or queuing times of 4 time units. The maximum value has been set such that the percentage of tardy jobs is 20% if jobs are released immediately upon arrival and capacity adjustments are not applied. Like in all prior Workload Control simulation studies, backward scheduling is performed based on the assumption of infinite capacity. The allowance for operation throughput times is set to 3 time units as this value resulted in the best performance during preliminary simulation experiments.

### 2.2. Input Control – order release (and dispatching)

There are many order release methods in the Workload Control literature; for example, see the reviews by Wisner (1995), Land and Gaalman (1996), Bergamaschi et al. (1997), and Fredendall, Ojha, and Patterson (2010). In this paper, the LUMS COR (Lancaster University Management School Corrected Order Release) method is used because it was recently shown to be the best order release solution for Workload Control (Thürer et al. 2012). LUMS COR uses a periodic release procedure to keep the workload \(W_s^R\) released to a station \(s\) within a pre-established workload norm as follows:

1. All jobs in the set of jobs \(J\) in the pre-shop pool are sorted according to the highest priority determined by a pool sequencing rule (as will be described below).
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 processing time \(p_{ij}\) at the \(i^{th}\) operation in its routing – corrected for station position \(i\) – together with the workload \(W_s^R\) 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 \(\frac{p_{ij}}{t} + W_s^R \leq N_s \ \forall i \in R_j\), then the job is selected for release. That means it is removed from \(J\) and its load contribution is included, i.e. \(W_s^R := W_s^R + \frac{p_{ij}}{t} \ \forall i \in R_j\).
   - Otherwise, the job remains in the pool and its processing time does not contribute to the station load.
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.

Since a released job contributes to \(W_s^R\) until its operation at this station is completed, the load contribution to a station in LUMS COR is calculated by dividing the processing time of the operation at a station by the station’s position in a job’s routing (Oosterman, Land, and Gaalman 2000). In addition to the above periodic release mechanism, LUMS COR incorporates a continuous workload trigger. If the load of any station falls to zero, the next job in the pool sequence with that station as
the first in its routing is released irrespective of whether this would exceed the workload norms of any station to avoid premature station idleness (see, e.g. Kanet 1988; Land and Gaalman 1998).

As in previous simulation studies on Workload Control (e.g. Land and Gaalman 1998; Fredendall, Ojha, and Patterson 2010; Thürr et al. 2012), it is assumed that all jobs are accepted, materials are available, and all necessary information regarding shop floor routings, processing times, and so on is known. Jobs flow into a pre-shop pool to await release according to LUMS COR. The time interval between releases for the periodic element of LUMS COR is set to 4 time units and eleven workload norms – from 4 to 14 time units – are considered. As a baseline measure, experiments without controlled order release have also been executed, i.e. where jobs are released onto the shop floor immediately upon arrival.

2.2.1. Pool sequencing rule
Two pool sequencing rules are considered in this study: the Planned Release Date (PRD) and Modified Capacity Slack (MODCS) rules. PRD sequences jobs according to planned release dates. The planned release date of a job is given by its due date minus a constant allowance for the operation throughput time for each operation in its routing. The constant allowance of the operation throughput time has been set to 5 time units since this value resulted in the best overall performance in preliminary simulation experiments. MODCS, as introduced by Thürr et al. (2015), uses: (i) a load-oriented Capacity Slack CORrected (CSCOR – as described below) element to speed up production when multiple jobs become urgent, and (ii) a time-oriented PRD element to ensure non-urgent jobs are released so the mix of released jobs can be produced on time. MODCS can be summarised as follows:

(i) Jobs are divided into two classes: urgent jobs, i.e. jobs with a planned release date that falls within the next release period or has already passed; and non-urgent jobs. Urgent jobs will always receive priority over non-urgent jobs.
(ii) Within the class of urgent jobs, jobs are sequenced according to the CSCOR rule.
(iii) Then, within the class of non-urgent jobs, jobs are sequenced according to the PRD rule.

CSCOR is a load-oriented rule that sequences jobs according to a capacity slack ratio based on corrected aggregate load measures of the workload, as given by Equation (3) below. This rule integrates three elements into one priority measure: (i) the workload contribution of the job (i.e. the corrected processing time); (ii) the load gap at a station (i.e. the remaining capacity that is available for orders in the pool to fill); and (iii) the routing length \(n_j\), i.e. the number of operations in the routing of a job \(j\), which is used to average the ratio between the load contribution and load gap elements over all operations in the routing of the job. The lower the capacity slack ratio \(S_j\) of job \(j\), the higher the priority of the job. Note that the same rule – but based on an uncorrected measure for calculating the load contribution and load gap elements – was originally proposed by Philipoom, Malhotra, and Jensen (1993).

\[
S_j = \frac{\sum_{i \in k} \left( \frac{p_{ij}}{N_s - W^R_s} \right)}{n_j}
\]

Finally, the capacity slack ratio could become negative due to the continuous starvation trigger incorporated in LUMS COR. This could result in the sequencing rule prioritising a job that contributes to the workload of an already overloaded station. Therefore, if the workload of a station is equal to or exceeds the workload norm, that is \(N_s - W^R_s \leq 0\), then the job is positioned at the back of the queue by replacing the component \(\frac{p_{ij}}{N_s - W^R_s}\) related to this station in the priority value \(S_j\) by \(\left(\frac{p_{ij}}{N_s - W^R_s}\right) \cdot M\), where \(M\) is a sufficiently large number.

2.2.2. Priority dispatching
To ensure the alignment of due date setting and dispatching, the dispatching rule applied on the shop floor is based on operation due dates, i.e. the job with the earliest operation due date is chosen from the queue in front of a station.

2.3. Output Control – capacity adjustments
Output Control evaluates the availability and need for capacity resources in the near future, e.g. through overtime or extra shifts (Lödding 2012). The main interest of this study in terms of output control is in the operational impact of capacity
adjustments. We are consequently not interested in the specific adjustment mechanisms used but in the timing of the capacity adjustment, i.e. when and where to adjust capacity. To model a capacity adjustment, we simply decrease the operation processing time by a predetermined percentage.

Several capacity planning techniques exist; for a review, the reader is referred to Wortman et al. (1996). In this study, we use the procedure outlined by Land et al. (2015) where capacity adjustments are guided by three parameters:

1. The size of the processing time reduction ($\alpha$);
2. The load threshold that triggers the commencement of a capacity adjustment ($\beta$); and
3. The load threshold signalling that the load has reduced sufficiently to cease the adjustment ($\gamma$).

Five different scenarios for the adjustment size $\alpha$ are considered to assess the impact of output control: 0% (i.e. no capacity adjustment), 10%, 20%, 30%, and a 40% adjustment. We only use one level of $\beta$ (90th percentile) and one level of $\gamma$ (85th percentile), which is justified by the performance frontier observed in Land et al. (2015).

The two load thresholds specify which periods will be distinguished as high-load periods and thus are in need of capacity adjustments. These thresholds have been determined numerically based on preliminary simulation experiments where we recorded the cumulative frequency distribution of the planned workload (measured in terms of the corrected load) that emerges without capacity adjustment for each workload norm being considered and for the scenario where all due dates are set externally. The planned workload is the workload in the pool plus the workload released to a station that is not yet completed. The load threshold for each workload norm is derived using a percentile of this distribution. The different levels for $\beta$ and $\gamma$ are summarised in Table 1. The same load threshold was applied for all levels of the percentage of due dates set internally. Finally, none of the capacity adjustments in our experiments reduced the overall utilisation by more than 0.5 percentage points.

This section has outlined the Workload Control concept considered and how it will be modelled in the simulations. In the next section, we outline the characteristics of the pure job shop in which the performance of our Workload Control concept will be assessed.

3. Simulation model
The shop and job characteristics modelled in the simulations are first outlined in Section 3.1 before Section 3.2 summarises the experimental design and the measures used to evaluate performance.

3.1. Overview of modelled shop and job characteristics
A simulation model of a randomly routed job shop (Conway, Maxwell, and Miller 1967) – later referred to as a pure job shop (Melnyk and Ragatz 1989) – has been implemented in the Python© programming language using the SimPy© simulation module. We decided for a process model instead of an object-oriented model given its focus on activities, flow dependencies, and resources – all of which are characteristic of job shops. SimPy© was chosen since it is an open-source software under the MIT License that has been widely used in recent studies on Workload Control (e.g. Thürer, Stevenson, and Land 2016a; Gonzalez-R, Calle, and Andrade-Pineda forthcoming). We model a balanced shop since here load balancing is most important (Thürer et al. 2017). The shop contains six stations, where each station is a single, constant capacity resource. The routing length of jobs varies uniformly from one to six operations. All stations have an equal probability of being visited and a particular station is required at most once in the routing of a job. This means that we omit re-entrant loops to avoid additional complexities (see e.g. Thürer and Stevenson 2016b), which would distract from the focus of our study. 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 follows an exponential distribution with a mean of 0.648, which – based on the average number of stations in the routing of a job – deliberately results in a utilisation of 90%. While any individual job

<table>
<thead>
<tr>
<th>Norm</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>IMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRD</td>
<td>$\beta$</td>
<td>23.64</td>
<td>22.16</td>
<td>21.01</td>
<td>20.23</td>
<td>19.65</td>
<td>19.24</td>
<td>18.93</td>
<td>18.72</td>
<td>18.59</td>
<td>18.48</td>
<td>18.43</td>
</tr>
<tr>
<td></td>
<td>$\gamma$</td>
<td>20.46</td>
<td>19.19</td>
<td>18.24</td>
<td>17.58</td>
<td>17.1</td>
<td>16.79</td>
<td>16.54</td>
<td>16.38</td>
<td>16.28</td>
<td>16.21</td>
<td>16.16</td>
</tr>
<tr>
<td></td>
<td>$\gamma$</td>
<td>20.73</td>
<td>19.88</td>
<td>19.16</td>
<td>18.55</td>
<td>18.04</td>
<td>17.62</td>
<td>17.29</td>
<td>16.92</td>
<td>16.69</td>
<td>16.52</td>
<td>16.35</td>
</tr>
</tbody>
</table>
shop in practice will differ in many aspects from this stylised environment, it captures the typical job shop characteristics of high routing variability, processing time variability, and arrival variability.

3.2. Experimental design and performance measures

The experimental factors are (i) the five levels of the percentage of due dates set internally by the due date setting rule (100%, 75%, 50%, 25%, and 0%); (ii) the eleven levels of the workload norm for LUMS COR (from 4 to 14 time units); (iii) the two pool sequencing rules (PRD and MODCS); and (iv) the five levels of capacity adjustment reflected by \( \alpha \).

A full factorial design was used with 550 (5 \( \times \) 11 \( \times \) 2 \( \times \) 5) scenarios, where each scenario was replicated 100 times. Results were collected over 10,000 time units following a warm-up period of 3000 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 \( j \)).

To assess timing performance we measure the standard deviation of lateness. To assess throughput performance, we measure the mean lead time – i.e. the mean of the completion date minus the pool entry date across jobs. Please note that the fixed arrival rates in the simulation determine the total throughput that can be realised. Therefore, the throughput improvement capabilities of a production control method will manifest itself in shorter total throughput times, i.e. lead times, rather than increased throughput. As an instrumental performance variable, the average shop floor throughput time is also evaluated. While the lead 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. By Little’s law, the average shop floor throughput time is directly related to the average level of work-in-process on the shop floor.

4. Results

To obtain a first indication of the relative impact of the experimental factors, statistical analysis has been conducted by applying ANOVA. ANOVA is based on a block design, with the workload limit as the blocking factor, which is typically used to account for known sources of variation in an experiment. The results are summarised in Table 2. All main effects and most two-way and three-way interactions were shown to be statistically significant. Detailed performance results will be presented and discussed next in Section 4.1 – first for each control function in isolation, followed by the results for pairs of functions, and finally for the combination of all three functions together. Section 4.2 then discusses the relationship between the control functions and the throughput and timing performance objectives.

4.1. Performance assessment

In Figure 2, we depict the percentage tardy, mean tardiness, standard deviation of lateness, and lead time results against the throughput time results for the scenario where all due dates are set internally by the due date setting rule (Figure 2(a)) and for the scenario where all due dates are set externally, i.e. the due date setting rule is not active (Figure 2(b)). Only the results for PRD pool sequencing are presented in Figure 2, with the impact of the MODCS pool sequencing rule discussed in Section 4.2 below. Meanwhile, results for the intermediate settings of the percentage of due dates set internally are presented in Figure 3(a) (75%), Figure 3(b) (50%), and Figure 3(c) (25%). The results are presented in the form of performance curves, where the left-hand starting point of the curves represents the tightest workload norm of 4 time units. The workload norm is tightened step-wise by moving from right to left in each graph, with each data point representing one workload norm (from 14 to 4 time units). Tightening the norms (towards a norm of 4 time units) decreases the workload on the shop floor and, as a result, the throughput time on the shop floor. In addition, the results obtained when jobs are released immediately are also included. These results are given by the single point towards the right-hand side of each figure and represent the outcome when no order release control is applied.

In terms of the main effects, the following can be observed from the results:

- **Due Date Setting in Isolation:** This can be observed from the results for immediate release with no capacity adjustments (i.e. the single point towards the right-hand side) in Figures 2 and 3. From the figures, we can observe a strong reduction in the percentage tardy, mean tardiness, and the standard deviation of lateness when due dates are set internally.
- **Order Release in Isolation:** This can be observed from the curve for no adjustment and all due dates set externally by the customer (Figure 2(b)). We observe a reduction in the percentage tardy and mean tardiness when order release
control is applied if norms are set appropriately (i.e. neither too loosely nor too tightly). It can be concluded that this reduction is due to a reduction in the lead time, i.e. a throughput improvement, since the standard deviation of lateness is either maintained or it increases.

- **Capacity Adjustments in Isolation:** This can be observed from the results for immediate release (the single point towards the right-hand side) and all due dates set externally by the customer (Figure 2(b)). We observe a strong improvement in all performance measures when capacity adjustments can be made.

Due date setting and capacity adjustments both improve delivery performance; the former through improved timing and the latter through improved throughput. This can be explained by the strong correlation between high-load periods (when capacity adjustments are triggered) and tardiness (Land et al. 2015). In fact, if there is an overload, management may either increase the lead time allowance or increase capacity, i.e. both functions can improve delivery reliability at the moment when this is required by the high load. Order release has a weaker impact since its load balancing mechanism neglects the workload situation. While a larger set of available orders in the pool makes more effective load balancing more likely, this effect is weaker compared to capacity adjustments. The effectiveness of load balancing can be seen from the lead time reductions. But the main role of controlled order release is to shift work from the shop floor to the pre-shop pool, which results in shorter shop floor throughput times, better prioritisations, and a leaner shop floor.

### Table 2. ANOVA results.

<table>
<thead>
<tr>
<th>Source of variance</th>
<th>Sum of squares</th>
<th>Degree of freedom</th>
<th>Mean squares</th>
<th>F-ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lead Time</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norm</td>
<td>14,580.006</td>
<td>10</td>
<td>1458.001</td>
<td>1012.440</td>
<td>0.000</td>
</tr>
<tr>
<td>Percent due date internal (DD)</td>
<td>16,229.146</td>
<td>4</td>
<td>4057.287</td>
<td>2817.400</td>
<td>0.000</td>
</tr>
<tr>
<td>Pool sequencing rule (SR)</td>
<td>5886.250</td>
<td>1</td>
<td>5886.250</td>
<td>4087.440</td>
<td>0.000</td>
</tr>
<tr>
<td>Capacity adjustment (CA)</td>
<td>28,319.972</td>
<td>4</td>
<td>7079.993</td>
<td>4916.390</td>
<td>0.000</td>
</tr>
<tr>
<td>DD × SR</td>
<td>279.250</td>
<td>4</td>
<td>69.813</td>
<td>48.480</td>
<td>0.000</td>
</tr>
<tr>
<td>DD × CA</td>
<td>24,448</td>
<td>16</td>
<td>1.528</td>
<td>1.060</td>
<td>0.387</td>
</tr>
<tr>
<td>SR × CA</td>
<td>512.752</td>
<td>4</td>
<td>128.188</td>
<td>89.010</td>
<td>0.000</td>
</tr>
<tr>
<td>DD × SR × CA</td>
<td>3.206</td>
<td>16</td>
<td>0.200</td>
<td>0.140</td>
<td>1.000</td>
</tr>
<tr>
<td>Residual</td>
<td>79,118.034</td>
<td>54,940</td>
<td>1.440</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Percentage Tardy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norm</td>
<td>7.624</td>
<td>10</td>
<td>0.762</td>
<td>1452.790</td>
<td>0.000</td>
</tr>
<tr>
<td>Percent due date internal (DD)</td>
<td>57.242</td>
<td>4</td>
<td>14.310</td>
<td>27,268.230</td>
<td>0.000</td>
</tr>
<tr>
<td>Pool sequencing rule (SR)</td>
<td>2.950</td>
<td>1</td>
<td>2.950</td>
<td>5621.560</td>
<td>0.000</td>
</tr>
<tr>
<td>Capacity adjustment (CA)</td>
<td>6.770</td>
<td>4</td>
<td>1.693</td>
<td>3225.140</td>
<td>0.000</td>
</tr>
<tr>
<td>DD × SR</td>
<td>4.439</td>
<td>4</td>
<td>1.110</td>
<td>2114.460</td>
<td>0.000</td>
</tr>
<tr>
<td>DD × CA</td>
<td>2.769</td>
<td>16</td>
<td>0.173</td>
<td>329.810</td>
<td>0.000</td>
</tr>
<tr>
<td>SR × CA</td>
<td>1.291</td>
<td>4</td>
<td>0.323</td>
<td>614.880</td>
<td>0.000</td>
</tr>
<tr>
<td>DD × SR × CA</td>
<td>0.446</td>
<td>16</td>
<td>0.028</td>
<td>53.110</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual</td>
<td>28.833</td>
<td>54,940</td>
<td>0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean Tardiness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norm</td>
<td>7036.528</td>
<td>10</td>
<td>703.653</td>
<td>8864.450</td>
<td>0.000</td>
</tr>
<tr>
<td>Percent due date internal (DD)</td>
<td>2984.158</td>
<td>4</td>
<td>746.040</td>
<td>9398.430</td>
<td>0.000</td>
</tr>
<tr>
<td>Pool sequencing rule (SR)</td>
<td>1281.996</td>
<td>1</td>
<td>1281.996</td>
<td>16,150.290</td>
<td>0.000</td>
</tr>
<tr>
<td>Capacity adjustment (CA)</td>
<td>2432.767</td>
<td>4</td>
<td>608.192</td>
<td>7661.860</td>
<td>0.000</td>
</tr>
<tr>
<td>DD × SR</td>
<td>121.393</td>
<td>4</td>
<td>30.348</td>
<td>382.320</td>
<td>0.000</td>
</tr>
<tr>
<td>DD × CA</td>
<td>684.268</td>
<td>16</td>
<td>42.767</td>
<td>538.770</td>
<td>0.000</td>
</tr>
<tr>
<td>SR × CA</td>
<td>2.710</td>
<td>4</td>
<td>0.678</td>
<td>8.540</td>
<td>0.000</td>
</tr>
<tr>
<td>DD × SR × CA</td>
<td>75.917</td>
<td>16</td>
<td>4.745</td>
<td>59.770</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual</td>
<td>4361.091</td>
<td>54,940</td>
<td>0.079</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Standard Deviation of Lateness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norm</td>
<td>20,621.040</td>
<td>10</td>
<td>2062.104</td>
<td>8300.240</td>
<td>0.000</td>
</tr>
<tr>
<td>Percent due date internal (DD)</td>
<td>192,119.800</td>
<td>4</td>
<td>48,029.951</td>
<td>19,332.070</td>
<td>0.000</td>
</tr>
<tr>
<td>Pool sequencing rule (SR)</td>
<td>82,494.553</td>
<td>1</td>
<td>82,494.553</td>
<td>33,204.090</td>
<td>0.000</td>
</tr>
<tr>
<td>Capacity adjustment (CA)</td>
<td>92,636.006</td>
<td>4</td>
<td>23,159.001</td>
<td>9321.510</td>
<td>0.000</td>
</tr>
<tr>
<td>DD × SR</td>
<td>1254.560</td>
<td>4</td>
<td>313.640</td>
<td>126.240</td>
<td>0.000</td>
</tr>
<tr>
<td>DD × CA</td>
<td>3478.860</td>
<td>16</td>
<td>217.429</td>
<td>87.520</td>
<td>0.000</td>
</tr>
<tr>
<td>SR × CA</td>
<td>12,333.877</td>
<td>4</td>
<td>3083.469</td>
<td>1241.100</td>
<td>0.000</td>
</tr>
<tr>
<td>DD × SR × CA</td>
<td>106.462</td>
<td>16</td>
<td>6.654</td>
<td>2.680</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual</td>
<td>136,496.770</td>
<td>54,940</td>
<td>2.484</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
If we look at the performance impact of combining two of the three production control functions, the following can be observed:

- **Due Date Setting and Order Release**: This can be observed from the curves in Figures 2 and 3 for no capacity adjustment. We observe that all performance measures improve when a larger percentage of due dates can be determined.
internally, but here the influence on lead time performance is at its weakest. The order release function is able to enhance this lead time performance in combination with a strong reduction in shop floor throughput times. The impact of the order release function is stronger when not all due dates are set internally (i.e. Figure 2(b) and Figure 3(a–c)).

- **Due Date Setting and Capacity Adjustments:** This can be observed from the results for immediate release, i.e. the four single points towards the right-hand side in all figures, with the differences between points within a given figure demonstrating the influence of capacity adjustments and the differences between figures demonstrating the influence of due date setting. Both control functions impact on all performance measures and are highly complementary. Due date setting especially improves the standard deviation of lateness while the main effect of capacity adjustments is on reducing the lead time. Together, they strongly reduce the percentage tardy and the average tardiness. In the absence of any order release control, all improvements take place at high shop floor throughput times, i.e. at a high level of work-in-process.

- **Order Release and Capacity Adjustments:** This can be observed from the curves where all due dates are set externally by the customer (Figure 2(b)). Both functions support each other in reducing lead times. Order release allows this improvement to be combined with a strong reduction in shop floor throughput times, i.e. work-in-process.

Figure 3. Performance comparison for PRD pool sequencing: 75%, 50%, and 25% of DDs set internally.
However, the detrimental effect of the order release function on the standard deviation of lateness and on the mean tardiness limits the reduction in percentage tardy when the two functions are combined. To ensure the lowest percentage tardy when capacity adjustments are possible, order release should only be used to enable a relatively small reduction in work-in-process.

The combined effect of the three functions, i.e. Due Date Setting, Order Release, and Capacity Adjustments, can best be observed from Figures 2(a and b), while Figure 3 depicts the intermediate levels of due date setting (i.e. where 75%, 50%, and 25% of due dates are set internally). It can be concluded from the figures that the best performance is in fact realised by combining all three control functions. This can be seen from observing the lowest points across all curves. However, if both the due date setting and capacity adjustment functions can be executed to an intensive degree, order release can only be used to achieve a limited reduction in work-in-process, as represented by the shop floor throughput time. Strong reductions in work-in-process would be to the detriment of other performance measures. Hence, workload norms have to be calibrated in combination with the use of the other control functions.

Finally, similar conclusions on the impact of order release control (as under PRD pool sequencing) can be obtained for order release under MODCS sequencing. This can be observed from Figure 4(a and b), which provide the results for the same scenarios as in Figure 2, but under MODCS pool sequencing. However, compared to PRD pool sequencing, a much stronger reduction in the percentage of tardy jobs can be observed for MODCS. Meanwhile, the performance differences between the different levels of the capacity adjustment diminish. As an exception, the standard deviation of lateness does not improve, and this can be explained by the creation of shortest processing time (SPT) effects through the use of the MODCS sequencing rule.

4.2. Discussion of results: throughput improvements and timing performance

In terms of our two performance objectives – throughput improvement, which aims to reduce average lateness, and timing, which aims to reduce the dispersion of lateness – the following can be observed from the results:

- **Throughput Improvement**: Results confirm that order release and output control (capacity adjustments) have the strongest impact on lead times whereby order release mainly affects the share of the shop floor throughput time while capacity adjustments directly affect the total lead time. We can also observe a reduction in the lead time and shop floor throughput time for due date setting. This reduction can, however, be explained by SPT effects, which are created as part of the forward loading procedure since operations with short processing times find it easier to fit within the capacity norm (Thürer et al. 2013).

- **Timing Performance**: Results confirm that due date setting has the strongest impact on the standard deviation of lateness. But order release and output control (capacity adjustments) may also positively impact this performance measure. Order release may have a positive impact through its PRD pool sequencing rule, which considers the urgency of jobs; and output control improves performance since the periods of high load are also the periods during which orders are at the highest risk of becoming tardy.

Table 3 summarises our findings on the link between the three control functions and the two key performance objectives. Our results confirm Kingsman, Tatsiopoulos, and Hendry (1989) in the sense that input/output control has a bias towards load balancing, although it also realises good delivery performance in contexts where the urgency of orders and the size of the workload are highly correlated. We, therefore, agree with Kingsman, Tatsiopoulos, and Hendry (1989) that input/output

<table>
<thead>
<tr>
<th>Control Function</th>
<th>Performance objective</th>
<th>Throughput</th>
<th>Timing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Due date setting</td>
<td>Weak impact through the forward loading mechanism, which combined with the dispatching rule introduces SPT effects</td>
<td>Strong impact on timing performance</td>
<td></td>
</tr>
<tr>
<td>Order release</td>
<td>Strong impact on the shop floor throughput time, weaker impact on the total lead time</td>
<td>Weak impact through the pool sequencing rule</td>
<td></td>
</tr>
<tr>
<td>Output control (capacity adjustments)</td>
<td>Strong impact on the total lead time</td>
<td>Weak impact through the relationship between high-load periods and urgency</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Link between control functions and performance objectives.
control should be enhanced by a customer enquiry stage where due dates or delivery lead times (and prices) are quoted, especially in make-to-order contexts where due dates are negotiable. Furthermore, if due dates are fully under the control of the company then using an effective due date setting rule is likely to have a stronger impact on delivery performance than input/output control.

Figure 4. Performance comparison for MODCS pool sequencing: 100% of DDs set internally and all DDs set externally (0% of DDs set internally).

(a) 100% of DDs Set Internally
(b) All DDs Set Externally
5. Conclusions

Workload Control is a production planning and control concept that was developed for high-variety contexts. A unifying principle of Workload Control methods is input/output control whereby input control, such as in the form of an order release mechanism, is used to regulate the work that can enter the shop and/or shop floor; and output control, such as in the form of capacity adjustments, is used to regulate the outflow of work. It has however been argued in the literature that input/output control overemphasises throughput improvements to the detriment of the timing performance objective; and that input/output control should, therefore, be enhanced by a customer enquiry stage where due dates are quoted. Yet, although a broad literature on due date setting, order release, and output control exists, literature on the combined effect of the three control functions is limited.

In response to the above, this study has asked: What is the combined performance effect of due date setting, order release control, and output control? Using simulation, we have both confirmed and extended earlier arguments from the extant literature. First, we have confirmed earlier arguments on the positive performance effect of combining due date setting and order release and on order release and output control. Second, we have extended earlier arguments by showing that the best performance for all scenarios is realised via the combined use of all three control levels. Most importantly, we have shown that each control function can be related to a specific performance objective: output control and order release to throughput performance and due date setting to timing performance. Due date setting and output control, in particular, are shown to be highly supportive of each other as they address different performance objectives; therefore, their effect is complementary and additive. The contribution of order release to throughput improvement diminishes when capacity adjustments are in place. Without capacity adjustments, order release can take over part of this function to reduce the average lead time. In all settings, order release helps to realise a reduction in shop floor times, which can be important for exploiting the typical gains of a lean shop floor. The managerial implications of our findings will be discussed next before the paper closes with its limitations and proposed future research directions.

5.1. Managerial implications

Our results have shown that the best performance can be achieved by combining the three control functions, but a significant investment may be necessary in order to successfully implement all three functions together. Thus, a trade-off may have to be made between the degree to which a control function is implemented in line with the costs, additional complexity, and resulting performance improvement (Cransberg et al. 2016). Our study supports managers in striking the right balance between the three control functions. In practice, each shop is likely to require a different balance between throughput and timing improvements to eventually improve delivery reliability. The required performance improvement can be identified using, for example, the diagnosis framework outlined in Soepenberg, Land, and Gaalman (2008, 2012). Our study supports this framework by indicating the control function to use in order to best achieve this performance improvement. We saw that each of the three control functions has a major impact on a specific performance measure:

1. Due date setting on the standard deviation of lateness;
2. Order release (input control) on the shop floor throughput time or level of work-in-process; and
3. Capacity adjustments (output control) on the lead time.

Hence, due date setting has a strong effect on timing and order release and output control on improving throughput. A company should therefore first diagnose their shop to identify which of the two – poor average throughput times or poor timing – is the main cause of unsatisfactory delivery performance and then implement or improve the corresponding control function(s).

5.2. Limitations and future research

A shortcoming of our study is the limited environmental setting. While this is justified by the need to keep this study focused, future research is required to assess the impact of environmental factors such as the routing direction or level of processing time variability on performance. Similarly, we have only used one production control system, i.e. Workload Control. While this is justified by the job shop model applied in this study, future research could explore the impact of other production control systems on throughput and timing performance. This includes the development of control systems that integrate the two different performance objectives into one control decision. Finally, and as highlighted in the managerial implications, the investment in implementing a certain control function has to be balanced to some degree with the resulting performance benefits given any overall limit on the investment that can be made and based on the rate of diminishing returns. Future research is necessary to further explore this cost–benefit relationship to support managers in striking the right balance in practice.
Disclosure statement
No potential conflict of interest was reported by the authors.

Funding
This work was supported by National Natural Science Foundation of China [Grant Number 71750410694]; Guangdong Province Universities and Colleges Pearl River Scholar Funded Scheme 2017.

ORCID
Martin J. Land © http://orcid.org/0000-0003-1704-9874

References


