ELSEVIER

Journal of Operations Management 16 (1998) 63–75

JOURNAL OF OPERATIONS MANAGEMENT

Technical note

The effect of workload control (WLC) on performance in make-to-order companies

L.C. Hendry *, B.G. Kingsman, P. Cheung

Department of Management Science, The Management School, Lancaster University, Lancaster LA1 4YX, UK

Received 23 May 1996; accepted 24 February 1997

Abstract

Workload control (WLC) concepts are a new group of production planning and control methods designed to control queues in a job shop manufacturing environment. Their importance lies in the need to maintain this type of flexible manufacturing environment in make-to-order (MTO) companies, which manufacture different products for different customers. There have been several well-developed WLC concepts presented in the literature that address two major decision levels in MTO firms, the job entry level and the job release level. At the job entry level, customer enquiries are processed, and delivery dates (DDs) and prices are quoted to customers. At the job release level, decisions are made regarding which jobs should be released to the shop floor so that processing can commence. The effectiveness of WLC concepts at these two decision levels has been explored in the literature but is still inconclusive, especially in terms of the ability of the job entry level to address one of the most important objectives of WLC — the control of manufacturing lead times. This paper presents a simulation model designed to test the effectiveness of one of the most comprehensive WLC concepts presented in the literature. The model enables the effect of various control parameters within this WLC concept to be explored, including those used at the job entry level to control manufacturing lead times. Experimental results are presented, indicating that this particular concept can lead to lower manufacturing lead times when compared with an environment of no control, even when the same total workload is processed by the shop. © 1998 Elsevier Science B.V.

Keywords: Production planning; Workload control; Order review/release; Make-to-order; Job entry; Job shop; Lead time

1. Introduction

Workload control (WLC), sometimes referred to as 'order review/ release', has been described by Zäpfel and Missbauer (1993a) as one of the new production planning and control concepts available for practical applications. The main principle has been defined by Land and Gaalman (1994) as the control of the length of the queues in front of the workstations on the shop floor. If these queue lengths are kept short, waiting times and therefore overall manufacturing lead times will be controlled. There are three levels at which this control of queues can be attempted:

- Priority dispatching level: the day-to-day shop floor control level;
- Job release level: the short term production planning level;
- Job entry level: the medium term production planning level.

* Corresponding author.

0272-6963/98/$19.00
PII S0272-6963(97)00011-9

While the priority dispatching level has received most research interest, it is a relatively weak mechanism for the control of queues if used alone. Thus, WLC concepts use a stronger instrument, that of controlled job release. This means maintaining a 'pool' of unreleased jobs, which are only released if it would not cause the planned queues to exceed some predetermined norms. This, in turn, reduces the work-in-process and the task of priority dispatching is made easier. However, the job release stage can itself only be fully effective if the queue of jobs in the pool is also controlled. Otherwise, jobs may remain in the pool for too long, missing their promised DDs. Thus, comprehensive WLC concepts include a job entry stage at which the pool and the shop floor queues are both controlled.

The importance of the development of the WLC concept lies in its appropriateness to the make-to-order (MTO) sector of industry. This term, as defined by Kingsman et al. (1993), refers to firms that manufacture different items for particular customers. MTO firms do not have a standard product range but maintain a competitive advantage by selling their engineering capabilities. As has been argued in detail elsewhere (see for example Tatsiopoulos (1993); Zäpfel and Missbauer (1993a); Hendry and Kingsman (1989)), the diverse and unpredictable nature of their production is such that it is inappropriate for these companies to adopt many aspects of other new production planning and control concepts such as JIT and TOC. In particular, they do not have the type of repetitive manufacturing that enables dedicated facilities to be set up in a simplified shop floor layout. Thus, they cannot rely on the more visible 'situational management' on the shop floor, as described by Johnston (1995), leading to a decrease in the importance of higher level planning in many firms. Instead, they continue to operate as job shops, and the higher level control of shop floor queues, as offered by WLC, remains the most crucial part of their production planning and control systems.

The aim of WLC concepts is to improve several company performance measures including decreasing manufacturing lead times. This is a particularly important objective in MTO companies because they are often in a position of having to quote lead times to their customers before production commences; therefore, the length of the manufacturing lead time is visible to the customer. However, there has been some debate in the literature regarding whether the various WLC concepts can achieve their stated aims and, in particular, whether they can improve delivery performance measures.

The purpose of this paper is to extend this debate by presenting further research into the effectiveness of a WLC concept which has been developed at Lancaster University. It describes a simulation model which has been developed to investigate the effect of various aspects of this WLC concept on the performance of MTO companies, including the control of lead times. Some experimental results are presented that concentrate on looking at the effectiveness of using the job entry level of planning, as this level has not received much attention in the literature to date. Before describing the simulation model and the experiments, the paper first presents a discussion of previous research that has examined the effectiveness of WLC systems and gives a more detailed description of the particular WLC concept modelled in the simulation.

2. Previous research into the effectiveness of WLC

The majority of the research regarding WLC concepts presented in the literature has examined the job release stage in isolation of the job entry stage. A variety of different types of decision rules to determine which jobs to release have been tested along with their interaction with different priority dispatching rules (see for example Philipoom et al. (1993) and a review of this type of research as presented by Wisner (1995)). This work has been largely inconclusive in terms of the ability of a job release stage to control overall lead times. Although it is agreed that some mechanisms can reduce shop floor throughput times and WIP, it has been found by many authors that the time spent in the pool is extensive; therefore, the total lead time to produce a job is not reduced. For example, Melnyk and Ragatz (1989) show that some releasing rules can result in increased lead times when compared to an environment where all jobs are just released as soon as their materials are available. They suggest that the reason for this poor performance is that many of the rules studied only control input into the factory; they do not simultaneously control output in terms of capacity. In addition, there is no attempt to control the pool itself by the use of appropriate decision rules at the job entry level.

Studies that look at the job entry stage in isolation, or at both the job release level and the job entry level, are much more scarce but include research by Philipoom and Fry (1992), Melnyk et al. (1991); Zäpfel and Missbauer (1993b); Hendry and Wong (1994) and Salegna and Park (1996). Philipoom and Fry (1992) study the effect of rejecting some jobs at the order entry stage and show that by rejecting a small percentage of jobs, dramatic improvements in shop performance can be achieved. Melnyk et al. (1991) show that the effectiveness of the job release planning level is improved by the presence of a higher level planning stage which smooths the total workload by keeping it between predetermined minimum and maximum values. The higher level system and the job release stage were found to complement each other, in that the former reduces flow times while the latter reduces WIP. They also conclude that using control at both levels leads to a lesser need for complex priority dispatching rules. However, their results suggest that while mean tardiness is reduced by using both the job entry and the job release planning levels, the percentage of tardy jobs is lowest if a job entry level alone is used. These results are confirmed by Zäpfel and Missbauer (1993b) and Hendry and Wong (1994). However, Salegna and Park (1996) conclude that the percentage of tardy jobs can be reduced by a two-tier system. This is possible if the planning system used at the job entry level can pull jobs forward but is not able to delay jobs. The planning rules which they studied were fairly simple and were applied to a hypothetical bottleneck job shop. Thus, while their results are positive, they need to be confirmed for other job shop environments.

As in Salegna and Park (1996), most of the research described above has concentrated on fairly simple decision rules at the job release and job entry levels using simulation experiments that model an hypothetical job shop. There has been very little research into the effectiveness of more comprehensive decision rules. In their review of alternative WLC concepts, Land and Gaalman (1994) conclude that the three most comprehensive concepts are those proposed by Bertrand and Wortman (1981), Bechte (1988) and Tatsiopoulos (1983). The first of these concentrates on the job release stage and is not a full two-tier WLC system. The second addresses both the job entry and the job release levels, and has been tested using case study research. It was concluded that a comprehensive WLC system of this type enabled the factory studied to reduce lead times, reduce WIP, meet planned DDs, and guarantee a high work centre utilisation. However, as this research only addresses one factory, it is not clear whether these improvements could be achieved in other companies. A simplified version of the third concept, as initially proposed by Tatsiopoulos (1983), was the topic of the research by Hendry and Wong (1994) as mentioned above. However, as many of its features were ignored because of the nature of the simulation experiment used, a fuller test of this concept is still needed.

In summary, there is scope for much further research into the effect of the more comprehensive WLC concepts. In particular, further studies that consider the job entry level are needed to establish its independent effect and the interaction between this level of planning and the job release level. The method of research should preferably use more realistic simulation experiments than the hypothetical job shops studied in most of the research to date, so that more of the features of the concepts can be studied.

To this end, this paper presents a study of the independent effect of the job entry level of the WLC concept which was first proposed by Tatsiopoulos (1983) and has been extensively developed by researchers at Lancaster University (see for example Hendry and Kingsman (1991) and Hendry and Kingsman (1993) and Kingsman et al. (1993, 1994)). The method of analysis is by the use of a simulation developed to model real, rather than hypothetical job shops. The main features of this WLC concept are described in the next section.

3. The WLC concept developed at Lancaster University

The WLC concept developed at Lancaster University is part of a Decision Support System (DSS) that seeks to integrate the marketing and production functions of MTO companies. One of the most important features of this concept is the attempt of the DSS as a whole to offer advice to MTO company managers when they are bidding for orders. This is a key decision area in these companies because they operate in a very competitive environment and need to choose the terms of their bids carefully in order to attract a profitable portfolio of work. This means considering marketing criteria to determine the competitiveness of alternative bids, and production criteria to look at whether sufficient capacity could be made available to achieve the lead time that the customer requires or that the company wishes to offer. Addressing these issues has led to the development of many special features in this WLC concept. Different parts of the DSS have been described in a number of papers as the features have been developed (see for example Hendry and Kingsman (1991); Hendry and Kingsman, 1993; Kingsman et al. (1993, Kingsman et al., 1994)). As the simulation discussed in this paper concentrates on the job entry level, a brief description is given below (for further details see the papers listed above).

At the job entry level, the DSS has three main elements:

- GADDA (Generate Alternative Delivery Dates and Actions) that determines alternative ways of dealing with each enquiry while controlling the workload;
- a marketing element to assess past enquiries and determine the probability of future bids being successful if particular DDs and prices are quoted;
- EDMA (Enquiry Decision Maker Analyser) which is an interactive element in which the alternatives generated by GADDA are presented along with their associated probability of winning the bid as generated by the marketing element. The user either selects an alternative, generates others or decides to reject the job.

The simulation is concerned primarily with GADDA as this is the main WLC element at the job entry level. Some aspects of EDMA are also included in the simulation.

GADDA entails controlling backlogs of work that are called 'planned backlog lengths' (PBLs). There is a PBL for each work centre and for the shop as a whole, each of which is controlled over a limited time period known as the 'planning horizon'. All of the orders that are due to be commenced during this period must be completed by the end of this period; thus, the planned PBL for the end of the last week during the planning horizon is always zero. The planning horizon is effectively the maximum value of the PBL and the maximum manufacturing lead time (MLT) for all orders. This maximum MLT is too long for some of the smaller jobs accepted by most MTO firms; therefore, a separate maximum for small orders is used. Three important parameters to be determined by management before using GADDA are (1) the length of the planning horizon, (2) the maximum MLT for small orders, and (3) the threshold between large and small orders.

Apart from controlling the PBL and the total workload, GADDA also seeks to ensure that jobs are completed on or before their promised DDs. Each job is defined as a set of operations at different work centres, and each operation is given an individual Operation Completion Date (OCD). Planning is essential to ensure that there is sufficient capacity so that each OCD can be met and the final DD is realised at the appropriate time.

For most job enquiries, there will be a set of alternative DDs that could be quoted. For each one, GADDA calculates the capacity required to meet the associated OCDs and the DD at the same time as calculating the capacity required to maintain the PBL below its maximum for all relevant weeks/work centres. Thus, capacity is planned dynamically as each new enquiry is considered. Having determined the capacity required, the actual methods of altering capacity will vary from firm to firm. At the job entry level, they often include assigning overtime, reallocating operators, or subcontracting work.

To ensure that capacity is planned at the same time as planning the timing of jobs, input/output control, as defined by Wight (1970), is used at all levels of the DSS. Output control is exercised by planning future capacity levels as described above. At the job entry level, input control is exercised by determining which enquiries to bid for, which planned start date to use for each order, and when to assign processing priority to an order. In a MTO company, considering the possibility of rejecting enquiries is seen as essential to prevent the lead time syndrome, as described by Mather and Plossl (1978), in which lead times are not effectively managed but fluctuate wildly. The planned start date for each order should usually be within the planning horizon unless otherwise specified by the customer. Two alternative levels of processing priority are available: normal priority and hot priority. If the latter is chosen, then the planned queueing time for each job operation is less than if normal priority is chosen. This can only be effective if the number of jobs given hot priority is limited; so a maximum proportion of jobs to which this priority can be assigned is set before using GADDA.

Using input/output control in this way, GADDA generates alternative possible bids. The final decisions as to whether to bid, and if so, what price and DD to quote and what capacity levels to set, are made interactively through EDMA.

If the company decides to bid for an order, then this type of input/output control is exercised twice for that order, because in most MTO companies, the job entry level is split into two stages: (1) customer enquiry stage and (2) customer confirmation stage. The company makes a quote at the enquiry stage and then waits for the customer to either confirm the terms of the initial bid, or reject them, and then either request further negotiation or cancel the order. In some industries, the customer may take several months to confirm the quote. For example, this can occur when the customer is itself bidding for a larger job and is just subcontracting some of the work to the MTO company. Such a delay between these two stages of job entry may mean that those plans initially made at the customer enquiry stage will have to be adjusted at the confirmation stage.

While awaiting confirmation of a bid, further enquiries will be received; therefore it is necessary to have a policy regarding how to include outstanding bids within the planned workload. In GADDA, this is done by adding a proportion of the workload associated with each bid to the PBL for the appropriate work centres for appropriate weeks. The proportion is set as the total workload multiplied by the probability of winning that bid, as indicated by the marketing element of the DSS. At the customer confirmation stage, this proportion of the total workload is either deleted, if the job is rejected, or replaced by the total workload for the job, if it is accepted. In the latter instance, it cannot simply be assumed that this is possible, as this may lead to the PBLs exceeding their maximum or to work being completed after their DDs. It is necessary to recalculate capacity requirements.

For the purpose of simulation, some aspects of the methodology had to be simplified. In particular, the interactive element of EDMA could not be included, and there was insufficient marketing data from the company being studied to calculate the probability of winning each of the alternative bids to be considered. These simplifications are described in further detail in Section 4.5.

4. The simulation model

4.1. Aims and design of the simulation software

The main purpose of the simulation model as described in this paper is to test the effectiveness of the WLC concept developed at Lancaster University and to add to the general debate on whether WLC methods in general can improve the performance of MTO companies. However, the design of the model was heavily influenced by future objectives for the simulation code. First, it is designed for use alongside the DSS upon implementation as a tool to assist managers in choosing appropriate parameters to use in the DSS; for example, the best value of the planning horizon. Second, as some of the programming code required is exactly the same as that required for the DSS, the simulation was written so that the GADDA code will stand alone and could easily be removed for use in the DSS. Thus, the simulation was written in C++, making use of some simulation routines easily available to the researchers, rather than making use of a simulation package.

It is an ABC event-driven simulation, which has two main parts: (1) a shop floor simulator and (2) an order entry simulator. The former simulates the progress of orders through work centres on the shop floor. In particular, it deals with the start and finish time of each operation, management of queues and transportation between work centres. It can be used without the order entry simulator if required. At present, it is a representation of a particular factory as described in the next section. However, it was written as generically as possible, given time limitations, so that it can be adapted to simulate other factories with a minimum effort. The latter simulates the decision-making of Lancaster WLC concept. It is not tied to any particular factory but will make use of different input data according to the factory studied.

4.2. *The simulated factory*

The simulation presented here has aimed to be as realistic as possible by simulating a real factory rather than a hypothetical factory. The factory chosen is typical of many MTO factories, with a job shop layout. The basic characteristics of the shop floor as assumed in the simulation are as follows:

- There are 15 internal work centres, each processes one order at a time. Eleven of these carry out a unique process and cannot carry out work for each other. Therefore, they each have a separate queue of work. The remaining four work centres, referred to as '4WCs' hereafter, all carry out the same process with relatively large set up times between jobs. These work centres have a shared queue of work.
- When one of the 4WCs with a shared queue becomes free, an order, which uses the same work centre configuration as the order that has just been finished, is chosen from the queue of work whenever possible. This is essential to minimise the massive set up times for these machines. Otherwise the next job is chosen according to the same rules as used for the other 11 work centres, as described below.
- The queues of work are divided into two subgroups, one of hot priority orders and one of normal priority orders. When a work centre becomes free, orders are selected from its hot priority queue if it is not empty, otherwise an order is selected from its normal priority queue. In either case, it is selected using the dispatching rule chosen at the beginning of the simulation. At present the FIFO, SPT and LEAST SLACK rules are available.
- Capacity at the internal work centres can be altered by assigning overtime or by reallocating operators. There are no limitations on the frequency with which an operator can be reassigned.
- There are a total of 18 operators. The maximum number of operators that can be assigned to a work centre varies between 1 and 3 operators.
- Due to the fragile nature of the product and the processes required, there can be breakages of up to 20% during the course of one operation. Therefore, the quantity processed per order is increased appropriately at the outset so that, on the average, the right number of parts will be completed.

One particular part of the production process requires special equipment that is not available at the factory and is carried out at external work centres. The factory is a subsidiary of a larger company and as these external work centres are within the larger company, there is a commitment to undertake this part of the production for all jobs that require it. As it is impossible to be involved in the capacity planning of this factory, it is assumed that such work takes a fixed lead time of 24 h that is not dependent on order size.

Every order is different, so it was thought to be inappropriate to generate alternative orders for the simulation on the basis of certain typical characteristics. Instead, a sample of 85 orders was collected from the factory, representing about one year's work. Each time an enquiry is received in the simulation model, one of the orders from the sample is selected at random and the new enquiry is then assumed to have the characteristics and routing of that order. The characteristics of these orders can be summarised as follows:

- Number of operations per order: mean 9; median 11; minimum 3; maximum 13.
- Order size (in parts): mean 2391; minimum 6; maximum 13400
- Processing times:
  - at the 4WCs: mean 41 hours processing time per order; 24 hours set up time when required;
  - other internal work centres: mean 453 parts per hour; no set up times.
- Mean number of operations at the 4WCs: 1.28 per order
- Mean number of operations required at the external work centres: 1.15 per order.
- Total hours work per order excluding set ups: mean: 134; minimum 1; maximum 837.

4.3. *Validating the shop floor simulator*

The shop floor simulator was validated by comparing the performance of the simulated factory with the actual performance of the factory in terms of manufacturing lead times and the utilisation of machines. Initially, the match was poor. In particular, the work centre that is believed to be the company bottleneck was under utilised. In contrast, the work at another centre was building up and had not stabilised by the end of the simulation run. Thus, the work centre capacities were looked at again. It had been difficult to determine these initially as the company carries out a lot of last-minute worker reallocation, moving workers from slack to busy areas. As the simulation does not model worker reallocation at the day-to-day level, changes were not being modelled and average values were needed. The initial values were adapted several times until the factory managers agreed that realistic machine utilisations were being obtained by the simulation.

The final results showed a small discrepancy between simulated and actual lead times, the simulated values being slightly higher. However, it was agreed with the factory managers that this could not be further improved upon without allowing day-to-day capacity variations. Thus, at this stage, the validation of the shop floor simulator was completed. Further details regarding the validation process can be found in Hendry et al. (1993).

4.4. Calculation of the DD

For each enquiry, i, the DD is initially calculated using the formula:

$$DD_i = ST_i + TWK_i + NOP_i \times W_P$$

where:

DD_i = delivery date for job i;
ST_i = planned start time for job i;
TWK_i = Total work content of job i;
NOP_i = number of operations in the routing of job i;
W_p = waiting time per operation for a job with priority p (p could be normal or hot priority).

The job is then added to the workload for the appropriate weeks to check whether this DD is realistic. If it is not, then either the capacity is altered, or the DD may be delayed. Thus, all DDs set by GADDA are realistic values.

4.5. Simplifications of the WLC concept

One aspect of the DSS described in Section 3 that is not included in the simulation model is the marketing element. This was not possible because the factory being studied could not provide the required data so that the probability of winning certain types of order for given delivery dates could be assessed. This data is not routinely collected in manufacturing organisations; therefore, it is difficult to determine even a theoretical probability distribution. Instead, it was assumed that all of the orders would have the same probability of success. This probability is one of the input parameters for the simulation. In all of the experiments run to date, it has been set at 1. This simplification seemed to be realistic since the set of orders used in the simulation were real orders. However, it means that the ability of the DSS to assist in competitive bidding cannot be assessed by the current simulation model.

A second aspect of the DSS that is not replicated exactly in the simulation model is the use of interactive decision-making. For example, it is not possible to allow an outside decision maker to choose whether or not to bid for a job and if so, which DD to set. Allowing this type of interaction would increase the time required to run each simulation experiment substantially, and this was not thought to be worthwhile. Therefore, the simulation makes some assumptions about the behaviour of all decision makers and also allows the effect of different types of decision maker to be examined.

The main assumption about all decision makers is that they will follow the main principles of the WLC concept and will only accept jobs if the desired input/output control can be achieved. For example, if after looking at the possibility of adjusting capacity, a DD before the end of the planning horizon cannot be found, then the enquiry is rejected. Thus, the choice of parameters, such as the planning horizon, used in a run of the simulation, will affect the total number of orders that are processed by the shop.

The main variation in decision maker characteristics modelled in the simulation to date affects the way in which the DD is chosen. When selecting a DD, the simulation always stops as soon as a feasible DD has been found; so the order in which alternative

DDs are tried may be crucial. Two possible orders representing two types of decision maker have been modelled.

- The cautious decision maker will always start by trying to assign a DD which gives the job normal priority and for which the planned start date is always as soon as possible. If capacity cannot be adjusted to allow this, then the options of trying to assign hot priority and/or later start dates are examined.
- The decision maker tries to rush as many orders through the system as possible. This person always assigns hot priority to orders as long as the proportion of jobs that have hot priority is kept to below the maximum percentage allowed, rather than preserving processing priority for cases where it is essential.

Allowing this type of variation allows the simulation experiments to determine the sensitivity of the DSS to the type of user.

Other variations are also possible, such as the way in which capacity is adjusted, whether to try reallocating operators first or assign overtime first, if these are the available methods.

4.6. Varying the input data

All of the data used in the simulation is input via a data file which can be changed easily at the beginning of each simulation run. This makes it possible to experiment with both the characteristics of the factory and the parameters used within GADDA. Thus, the simulation is a powerful tool for examining the effectiveness of GADDA under various experimental conditions, and numerous different simulation experiments could be carried out. A subset of these experiments are presented below, demonstrating some of the abilities of the simulation model, while also adding to the general debate in the literature regarding whether WLC can control lead times.

5. The simulation experiments

These experiments address the following main research questions.

(1) Can WLC improve DD performance by reducing actual MLTs and reducing discrepancies between planned and simulated MLTs?

(2) Is the DD performance of the WLC concept affected by the influence of the cautious or rush order decision maker as described in Section 4.5?

(3) How does the choice of parameters for GADDA, i.e., of the length of the planning horizon and the maximum MLT for small orders, affect the DD performance of WLC?

Although the simulation model allows many of the input parameters to be varied for experimental purposes, only a few of these were varied in order to address these particular questions. The main parameters that were fixed are listed in Section 5.1 and all those which were allowed to vary are listed in Section 5.2. The performance measures used to assess the above research questions are listed in Section 5.3.

5.1. Fixed experimental factors

Data collected from the simulated factory and preliminary runs of the simulation led to the following choices for the main fixed experimental factors:

- the priority dispatching rule is SPT;
- inter-arrival rate of enquiries follow a negative exponential distribution with a mean of 3.88 real days;
- mean delay for customer confirmation = 8.2 real days;
- proportion of bids that are successful = 100%;
- small order threshold = 300 processing h;
- normal priority waiting time per operation = 1 working day;
- hot priority waiting time per operation = 0.5 working days;
- maximum proportion of jobs that can be assigned hot priority = 15%;
- maximum overtime per operator per day = 3 h;
- mean h per shift = 7.4 h;
- number of shifts per day = 2;
- where possible, capacity is adjusted by reallocating operators, overtime is only assigned if this is not possible.

5.2. Variable experimental factors

The values for the experimental factors that were allowed to vary were chosen after some preliminary simulation runs. These factors can be classified into three types that examine three effects of the use of the WLC concept studied.

(1) The backlog effect: two parameters affect the size of the PBL in GADDA, as described in Section 3. Their alternative values used in the experiments were: (a) the planning horizon, 13 weeks or 19 weeks; and (b) small orders maximum MLT, 5 weeks or 9 weeks.

(2) The DD effect: the effect of different criteria for selecting DDs was examined by allowing two options: (a) always trying to assign normal priority first, and (b) always trying to assign hot priority first.

(3) The methodology effect: using the WLC concept described in Section 3 or exercising no control over workloads.

This led to a total of 9 experiments, one without the methodology in use and 8 experiments derived from the combinations of the four possible parameters from category 1 above and the two possibilities from category 2.

For each experiment, the simulation was run for 500 weeks. The first 25 weeks worth of results were deleted to remove any start-up effects. Each experiment was replicated 10 times. The results presented in Section 6 are all in terms of average performance over those 10 replications.

5.3. Performance measures

Although the main performance measures of interest relate to the manufacturing lead times and are therefore measures of DD performance, it is obviously possible to achieve short lead times at the expense of other performance measures. Therefore, it is necessary to examine a variety of measures so that any tradeoffs are highlighted. The total number presented in Section 6 is 12, and they can be grouped into the following categories:

(1) measures of DD performance: mean and standard deviation of the planned MLT and the simulated MLT;

(2) measures of workload: number of orders, mean order size, mean weekly workload, mean weekly WIP, average queuing time;

(3) measures of capacity adjustment and utilisation: total extra overtime, total reallocation of operators required and percentage capacity utilisation.

6. Experimental results

The results of the simulation experiments are shown in Tables 1–4. In the first two rows of Table 1, ANOVA results are presented, looking at the significance, at the 5% level, of varying the parameters used in the methodology upon two of the most important performance measures for these experiments. It can be seen that the backlog effect, that is, the length of the planning horizon and the size of the maximum MLT for small orders, has a significant effect upon both of the performance measures — the simulated MLT and the total workload. This emphasises the importance of selecting these parameters carefully in order to gain the required MLT control. The choice of method of assigning the DD does not have a significant effect on either of the measures. This suggests a degree of stability within the GADDA methodology which is not influenced significantly by the two types of decision maker being analysed.

In the last row of Table 1, ANOVA results to investigate the effect of using the methodology are presented. These were calculated by comparing the experiment with no control with an experiment using GADDA with loose control, a planning horizon of 19 weeks and a maximum MLT for small orders of 9 weeks. The results indicate that in these two experiments, there is no significant difference between the workloads, but the simulated MLT is different. It is shown later, in Table 2, that the difference is that the GADDA methodology in this experiment leads to a lower simulated MLT. Thus, these results indicate that this method of WLC can achieve the desired reduction in MLTs while still processing the same total amount of work.

Table 2 presents the detailed delivery performance measures. It can be seen that in all cases, the simulated lead times achieved with GADDA are lower than those achieved when no WLC control is exercised. The smallest MLT achieved by GADDA

Table 1
Summary of ANOVA results

Experimental factor	Performance measure	
	Simulated MLT	Total workload
Backlog effect	Significant	Significant
DD effect	Not significant	Not significant
Methodology effect	Significant	Not significant

Table 2
Delivery performance measures

DD effect	Backlog effect {planning horizon, small-order MLT maximum}	Mean MLT (days)		Standard deviation of MLT (days)	
		Planned	Simulated	Planned	Simulated
Normal first	{19,9}	41.0	46.2	25.2	29.1
	{19,5}	39.1	42.9	27.4	29.0
	{13,9}	36.6	34.1	19.6	20.1
	{13,5}	33.7	30.0	21.2	20.6
Hot first	{19,9}	40.0	45.1	24.6	27.7
	{19,5}	38.8	42.8	27.6	29.0
	{13,9}	35.9	33.7	19.5	19.8
	{13,5}	33.4	29.9	21.3	20.7
Methodology effect: no WLC			50.7		37.3

Table 3
Workload performance measures

DD effect	Backlog effect {planning horizon, small-order MLT max. }	No. of orders	Mean order size (h)	Mean weekly workload (h)	Mean weekly WIP (h)	Average queueing time (h)
Normal first	{19,9}	868	150	288.0	311	54.3
	{19,5}	692	143	220.0	188	46.7
	{13,9}	816	121	217.6	130	27.8
	{13,5}	625	105	145.5	65	21.5
Hot first	{19,9}	863	147	279.9	283	52.0
	{19,5}	667	142	212.4	177	46.0
	{13,9}	811	119	213.0	122	27.2
	{13,5}	613	104	141.3	63	21.2
Methodology effect: no WLC		871	150	290.8	332	63.0

Table 4
Capacity adjustment measures

DD effect	Backlog effect {planning horizon, small-order MLT max.}	Mean extra overtime per week (h)	Mean man reallocation per week (h)	Capacity utilisation (%)
Normal first	{19,9}	5.8	3.8	45.5
	{19,5}	3.0	2.2	34.9
	{13,9}	1.1	1.2	34.6
	{13,5}	0.3	0.4	23.1
Hot first	{19,9}	5.3	2.8	43.8
	{19,5}	2.7	1.8	33.7
	{13,9}	0.9	0.9	33.9
	{13,5}	0.3	0.3	22.5
Methodology effect: no WLC		0	0	46.8

is linked to the highest degree of workload control, that is, a planning horizon of 13 weeks and a small order MLT maximum of 5 weeks. Using GADDA also leads to a reduction in the standard deviation of the simulated MLTs. The differences between planned and simulated MLTs are all less than one week. The smallest difference of 2.2 days occurs when there is tight control over the total workload with the planning horizon set at 13 weeks, and loose control of the small-order MLT maximum that is set at 9 weeks. This does not mean that this level of control leads to better predictability in general, but that the planned waiting times used to calculate the DDs were more appropriate for these control parameters. Further experimentation with different waiting times could improve the predictability for each level of workload control investigated.

Table 3 illustrates the effect that GADDA and its chosen parameters have upon the workload in the system. As expected, the tighter the planning horizon and the maximum MLT for small orders, the fewer the number of orders and the smaller the weekly workload. This is because GADDA is choosing not to accept jobs when there is insufficient capacity. The mean order size is also greatly reduced when the tighter planning horizon of 13 weeks is used. It is partly this reduction in workload which leads to the lower simulated MLTs discussed above. However, in comparing the use of GADDA with a system of no WLC, note the similar workloads of 288 for the first experiment using GADDA listed at the top of the table, and 290 for the experiment without any WLC at the bottom of the table. Although the workloads are similar, the experiment using GADDA gives a lower MLT (46.2 h compared with 50.7); a lower average queueing time (54.3 h compared to 63) and a lower mean WIP (311 h compared to 332). Therefore, it is clearly not just the reduction in workload that leads to better performance.

Table 4 looks at the last group of performance measures, measures of capacity adjustment and utilisation. It can be seen that GADDA makes good use of these facilities and that extra overtime is planned where workloads are at their highest. This occurs when the number of jobs accepted and the percentage of capacity utilisation are at their largest, and indicates that more output control is needed when the input control is relatively slack.

The main conclusions that can be drawn from these experiments can be summarised in terms of the research questions posed at the beginning of Section 5.

(1) WLC, as demonstrated using GADDA, can lead to lower actual MLTs even when the same total quantity of work is processed through the factory. The difference between the planned and simulated MLTs is less than one week for all of the experiments. Although reasonable in most MTO manufacturing environments, this could be improved further by careful choice of control parameters as discussed under (3) below.

(2) The performance of GADDA is not significantly affected by the manner in which the decision maker chooses the DDs for orders.

(3) The choice of parameters for GADDA allow the user of the system to manage MLTs at a level that is acceptable to the company. By using the simulation before GADDA is implemented, management can gauge which values of the planning horizon and the maximum small order MLT will lead to the desired competitive average MLT values. Any tradeoffs between the length of the MLTs and other measures such as the number of orders accepted can also be examined. Thus, management can use the simulation to make important strategic choices such as whether to minimise MLTs or to maximise the number of orders processed. In addition, the reliability of the planned MLTs can be altered by using the simulation to adjust the value of the waiting time W in the DD formula, as presented in Section 4.4, to match the level of WLC control being used. The simulation experiments suggest that the values used should be increased for a planning horizon of 19 weeks and decreased for a planning horizon of 13 weeks.

7. Conclusion

This paper has presented a simulation model which is a more powerful tool for examining the effect of WLC than those used by most of the simulations presented in the literature to date. It allows one of the more comprehensive WLC concepts, as first presented by Tatsiopoulos (1983) and extensively developed by Hendry and Kingsman (1991), Hendry and Kingsman (1993) and Kingsman et al. (1993, 1994), to be examined in more detail than has been achieved previously.

To demonstrate the use of the simulation and to add to the general debate on the effectiveness of WLC, a total of 9 simulation experiments have been presented. These show that WLC, used at the job entry level alone, can lead to shorter MLTs while still processing a similar total workload. The mean desired length of the MLTs is a strategic choice for MTO companies and the experimental results also demonstrated that it can be controlled using the particular WLC concept studied by choosing appropriate control parameters. In fact, the simulation model has been designed so that it can also be used prior to the implementation of this WLC concept to determine appropriate control parameters, so that using the system leads to the desired company performance.

The simulation modelled one particular MTO factory, but has been set up so that it can be altered to model alternative factories with relative ease. While the factory chosen is typical of many MTO firms, further research is needed to repeat these experiments while modelling other factories so that the generality of the results can be confirmed.

Further simulation experiments using this particular factory are also possible to look at other aspects of the performance of WLC, such as the interaction between the job entry level and the priority dispatching level. In addition, the simulation model needs to be developed further so that the interaction with the third planning level, job release, can also be studied.

The results of this research suggest that the WLC concept is indeed an effective management tool for MTO companies who wish to control aspects of performance such as MLTs. To be implemented effectively, this concept requires greater integration between the marketing and operations functions of the MTO firm. It is only when the operations managers play an active role in negotiating lead times that capacity can be effectively planned in the medium term and realistic tenders can be offered to customers. Thus, this research underlines the need for more functional integration in manufacturing companies and shows that WLC is a concept that deserves further research and greater implementation in practice than has been witnessed in companies to date.

Acknowledgements

The development of the simulation model was sponsored by the ACME Directorate of SERC (as was), with collaborative support provided by Pilkington Optronics. Much of the programming of the simulation code was carried out by two research assistants, Elaine Wilson and George Denholm.

References

Bechte, W., 1988. Theory and practice of load oriented manufacturing control. Int. J. Production Res. 26, 375–395.

Bertrand, J.W.M., Wortman, J.C., 1981. Production control and information systems for component manufacturing shops. Elsevier.

Hendry, L.C., Kingsman, B.G., 1989. Production planning systems and their applicability to make-to-order companies. European J. Oper. Res. 40, 1–15.

Hendry, L.C., Kingsman, B.G., 1991. Job release: part of an hierarchical system to manage manufacturing lead times in make-to-order companies. J. Oper. Res. Soc. 42, 871–883.

Hendry, L.C., Kingsman, B.G., 1993. Customer enquiry management: part of a hierarchical system to control lead times in make-to-order companies. J. Oper. Res. Soc. 44, 61–70.

Hendry, L.C., Wilson, E., Kingsman, B.G., Worden, L., Mercer, A., 1993. Integrating production planning in make-to-order companies. Proc. Int. Conf. on Managing Integrated Manufacturing. 1, 427–436.

Hendry, L.C., Wong, S.K., 1994. Alternative order release mechanisms: a comparison by simulation. Int. J. Production Res. 32 (12), 2827–2842.

Johnston, R.B., 1995. Making manufacturing practices tacit: A case study of computer-aided production management and lean production. J. Oper. Res. Soc. 46, 1174–1183.

Kingsman, B.G., Worden, L., Hendry, L.C., Mercer, A., Wilson, E., 1993. Integrating marketing and production planning in make-to-order companies. Int. J. Production Economics 30-31, 53–66.

Kingsman, B.G., Hendry, L.C., Mercer, A., de Souza, A., 1994. Responding to customer enquiries in make-to-order companies: problems and solutions. Int. J. Production Economics 46-47, 219–231.

Land, M., Gaalman, G., 1994. Workload control concepts in job shops: a critical assessment. Pre-prints of the Eighth International Working Seminar on Production Economics. Igls, Austria, 3, 201–226.

Mather, H., Plossl, G., 1978. Priority fixation versus throughput planning. Production and Inventory Management 19, 27–51.

Melnyk, S.A., Ragatz, G.L., 1989. Order review/release: research issues and perspectives. Int. J. Production Res. 27, 1081–1096.

Melnyk, S.A., Ragatz, G.L., Frendall, L., 1991. Load smoothing by the planning and order review/release systems: a simulation experiment. J. Oper. Management 10, 512–523.

Philipoom, P.R., Fry, T.D., 1992. Capacity-based order review/release strategies to improve shop performance. Int. J. Production Res. 30, 2559–2572.

Philipoom, P.R., Malhotra, M.K., Jensen, J.B., 1993. An evaluation of capacity sensitive order review and release procedures in job shops. Decision Sci. 24, 1109–1134.

Salegna, G.J., Park, P.S., 1996. Workload smoothing in a bottleneck job shop. Int. J. Oper. Production Management 16 (1), 91–110.

Tatsiopoulos, I.P., 1983. A microcomputer-based interactive system for managing production and marketing in a small component manufacturing firm using a hierarchical backlog control and lead time management methodology. PhD thesis, Lancaster University, UK.

Tatsiopoulos, I.P., 1993. Simplified production management software for the small manufacturing firm. Production Planning and Control 4, 17–26.

Wight, O., 1970. Input/Output control: a real handle on lead time. Production and Inventory Management, 3rd Qtr., pp. 9–31.

Wisner, J.D., 1995. A review of the order release policy research. Int. J. Oper. Production Management 15, 25–40.

Zäpfel, G., Missbauer, H., 1993a. New concepts for production planning and control. European J. Oper. Res. 67, 297–320.

Zäpfel, G., Missbauer, H., 1993b. Production planning and control systems including load-oriented order release — problems and research prospectives. Int. J. Production Economics 30-31, 107–122.