

Integrating material coordination and capacity load smoothing in multi-product multi-phase production systems

Citation for published version (APA):

Bertrand, J. W. M., & Ooijen, van, H. P. G. (1996). *Integrating material coordination and capacity load smoothing in multi-product multi-phase production systems*. (TU Eindhoven. Fac. TBDK, Vakgroep LBS : working paper series; Vol. 9309). Technische Universiteit Eindhoven.

Document status and date:

Published: 01/01/1996

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

**Integrating Material Coordination and
Capacity Load Smoothing in
Multi-Product Multi-Phase
Production Systems**

J.W.M. Bertrand
H.P.G. van Ooijen

Research Report TUE/BDK/LBS/93-09
January 1993

**Integrating Material Coordination and Capacity Load Smoothing in
Multi-Product Multi-Phase Production Systems**

J.W.M. Bertrand and H.P.G. van Ooijen

Research Report TUE/BDK/LBS/93-09

January 1993

Graduate School of Industrial Engineering and Management Science.

Eindhoven University of Technology

P.O.Box 513, Paviljoen F1

NL-5600 MB Eindhoven

The Netherlands

Phone: +31.40.472230

This paper should not be quoted or referred to without the prior written permission of the author

Integrating Material Coordination and Capacity Load Smoothing in Multi-Product Multi-Phase production Systems

J.W.M. Bertrand and H.P.G. van Ooijen

1. Introduction.

During the last decade much attention has been paid in research to the interaction between materials coordination and capacity loading. In particular in a multi-product multi-phase production system, where complex products are manufactured which each consist of many modules and components, and where similar modules and components are manufactured in specialized production departments, the interactions between the materials coordination decisions in the production chain, and the capacity loading decision for each production department are difficult to deal with. It is well known by now that the production throughput times of work orders in a complex production department, consisting of a number of work centres, and variable work order routings, are a function of the capacity load of the production department. A high capacity load will result in a large throughput time and low capacity load will result in small throughput times. This relationship between capacity load and throughput time has been studied by Bertrand 1981, 1983^a, 1983^b, Baker 1984, Kanet 1988, Bechte 1982, Karmarkar 1987, and Wein and Chevalier 1992.

On the one hand, this research has resulted in procedures to generate workload dependent work order due dates which largely improve the due date performance as compared to the performance of their workload independent counterparts (i.e. Eilon and Chowdhury 1976, Bertrand 1983^a, Baker and Bertrand 1981).

On the other hand the research has resulted in procedures to control the work order throughput times by controlling the workload in the department (i.e. Bertrand 1983^b, Bechte 1982). The decision function which controls the workload is known as the work order release function. It has been shown that the average work order throughput time in a complex production department can be controlled within narrow limits by releasing work orders to the department such that the workload is kept equal to a specific predetermined value. Under strict conditions closed network queuing theory can be used to determine the relationship between the workload (in terms of number of orders), average output rate (number of orders completed per unit of time) and

average throughput time (in time units per work order).

Since unreliable and variable throughput times often lead to a poor delivery performance, decision support systems based on these relationships have been developed to help the decision maker in practice to control the throughput times in their production department. Bertrand and Wortmann 1981 were among the first to report on the development and use of such a system and the remarkable performance improvement which was attained. Wiendahl (Wiendahl 1987) has developed a work order release system based on the principles presented by Bechte (Bechte 1982), and he also reports substantial improvements after the introduction of the system in many production departments (Wiendahl 1991).

Now it should be noticed that in both these approaches the work order throughput time is measured from time of release to the shopfloor, to the time of completion of the work order. Furthermore, it is assumed that always enough work orders are available for release to the shop floor, and that the time during which the orders wait for release is not part of the work order throughput time. In most situations in practice this is a quite unrealistic assumption, since work orders often are generated by a demand process which cannot be perfectly coordinated to the work order release opportunities. The coordination of demand for capacity and available capacity, for instance in the Rough Cut Capacity Check of an MRP II system, pertains to the averages in requirements and availability over, say, a month for a few months ahead, and does not pertain to the exact moments in time of work order arrivals and release opportunities. Therefore it is often only realistic to assume that the average capacity utilization over a longer time period is known beforehand, and to model the short term work order arrival process as a random process with a given arrival rate which is tuned to the available capacity.

The throughput time of work orders should be measured starting from the actual arrival time of the work order since this is the elapsed period of time which is experienced by the customer. It can easily be shown that under this condition in a situation where the FCFS sequencing rule is used (which is the case in many practical situations), the introduction of work order release, restricting the maximum number of orders on the shopfloor to a predetermined value, can only increase this average work order throughput time, as compared to the situation with immediate release. Furthermore simulation studies (Bertrand 1983^b, Ragatz 1985, Ragatz and Mabert 1988) show that also the due date performance is not positively affected by the introduction of work order release, except for the situations where very simple dispatching rules are used (Melnik et al. 1988). Recent research by Park 1987 and by Bobrowski and Park 1989 indicates a positive effect on total costs of the introduction of work order release in dual constrained job shops with

much labour flexibility, which effect deteriorates if the labour flexibility decreases.

From the theoretical studies cited above we may conclude that work order release, except for some special situations, hardly has any benefits, if work orders arrive according to a random process and throughput times are measured from the arrival time of the order. Now the question is if this is also true if we not consider the job shop in isolation, but extend the situation to also incorporate the order planning phase, the availability of materials for order release (the supply-side) and the criticality (required delivery performance) of the work orders for the down stream production phases (the demand side). Park and Salegna 1992 report good results by including the planning phase and manipulating the work order stream to smooth the load on the bottleneck resource. However they use a specific situation where all work orders start at the same work center which acts as the bottleneck. They also neglect the materials availability and the product criticality issues.

In this paper we concentrate on planning the work order release in relation to product criticality. If products have different criticality for the succeeding production phases the work orders for the less critical products are good candidates for postponed release in case of shop overload.

We position our research questions in an MRP environment; that is a situation where the shop produces items for which work orders are generated by a MRP system to replenish the inventories in the succeeding stock point, and each work order requires materials which are made available in the supply stock point by work orders placed by the MRP-system at earlier production phases.

Figure 1 shows this situation.

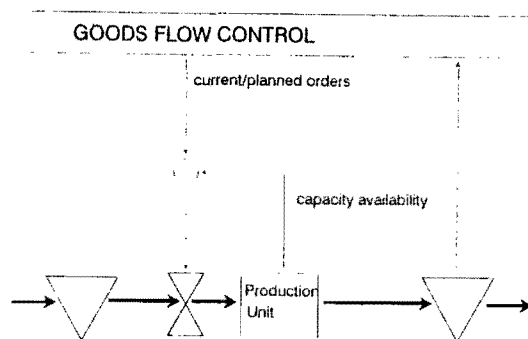


Figure 1. The environmental setting.

The organization of the remainder of this paper is as follows. In section 2 we present the control problem in its environment. In Section 3 we will present the system studied in detailed terms, followed in Section 4 by a presentation of the simulation study that has been carried out to

systematically investigate the research question for various environmental conditions. In Section 5 we present and discuss the results of this simulation study and in Section 6 we complete the paper with the conclusions.

2. The control environment.

MRP is nowadays a widely used system for material coordination in repetitive multi-product multi-phase production situations. A well known problem with the use of MRP in practice is the poor way in which the shop load in the production departments is taken into account. Generally the capacity availability of the resources in the departments is modelled in the rough cut capacity check which for each department results in a controlled average arrival rate of the work orders. Furthermore, the manufacturing economics of the departments are roughly modelled when setting the work order batch size rules for each department. Clearly batch sizes also have an impact on the workorder throughput times, as has been investigated by Solberg 1981, Bertrand 1985 and Karmarkar et al. 1984. Karmarkar also investigated rules for short term coordination of batch sizes, capacity load and throughput times (Karmarkar 1987). In this research we assume that work order batch size rules are given and that medium term capacity utilization of the department's resources are controlled to a predetermined level at the rough cut capacity check.

Due to the effect of batching, yield variations and demand variations downstream the manufacturing chain, we may assume that the work orders generated for the production department by the MRP system follow a random arrival process with a known arrival rate. We assume that for each product and material item, a work order lead time off set is used in the MRP system. This work order lead time represents the assumption made by the MRP system regarding the time that will elapse from the planned release of the work order to the delivery of the work order in the stock point. Furthermore we include the planning phase by assuming that up to a certain horizon planned orders are known before hand, and that over this given horizon, the work order due dates do not change (with this assumption we abstract from the much reported MRP system nervousness by assuming a certain frozen period in the planning horizon). Finally we assume that each work order requires components or materials which are always available in a stock point. Although this is a rather unrealistic assumption it is used to gain insight in the maximal effects of the use of a work order release function in production situations where work orders may have different criticality. The consequence of this assumption is that work orders for the department can always be advanced relative to their planned release time as given by

MRP, because the materials or components will already be available, even if no materials safety stock would be used. This in fact provides the work order release function with a possibility to manipulate the work order stream to smooth the load of the shop. If the work order release system decides to advance a work order then the materials required by the work order will be depleted from the stock point earlier than planned by MRP, and the work order will probably be earlier received in its stock point than planned by MRP. The MRP system will generate a reschedule-out message as soon as such an advanced work order is released. However since the reason for this early release is to smooth the capacity load, the materials planner should negate this message (standard MRP II does not support this capacity load smoothing function).

In this research we investigate the effects of using such a proactive work order release function, in case work orders may have different criticality, on the safety stock requirements for the receiving stock point. As a bench mark for comparison we use the standard MRP situation, that is a situation without work order release based on workload smoothing. In this latter situation we assume standard work order lead times that are equal to the average work order throughput time, and we assume that in the receiving stock points safety stocks are set to account for both uncertainty in demand and in variations in throughput times. In the standard MRP situations work orders are assumed to be released exactly at their planned release time, assuming that materials are available at that time.

The performance measures used are the service level of the receiving stock point on the one hand, and the total valued amount of materials in the receiving stock point on the other hand.

3. Work order release.

In many production situations there may be a number of components or products for which work orders are more critical (have a higher required delivery reliability) than work orders for the other components/products. There are a number of reasons why work orders may not have equal criticality. For instance consider two types of components; first components P_C (see Fig. 2) which are used in many products produced by the next production department, and second components which are only used in a few products produced in the next production department (P_{NC} in Fig. 2).

In this case a shortage of components P_C may cause a decrease of production or, even worse, a production stop in production department B. Besides the fact that capacity is lost, this may lead to loss of production and/or production stops in downstream production departments. If one does not want to use high stock levels this eventually may lead to a deterioration of the delivery performance to the market. Work orders for the components P_C are thus more critical than work orders for the components P_{NC} which are only used in a few products .

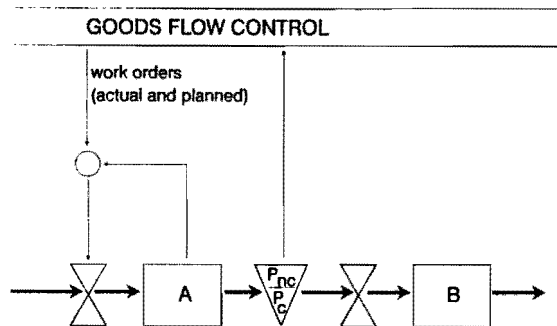


Figure 2. Selective work order release. P_C = critical products; P_{NC} = non critical products

An extreme example of different criticalities is the situation where part of the components are made to customer order, and the other components are made to stock. For components which are made to order no safety stocks are possible, only safety time may be allowed. Another example of unequal criticality is the case that components P_C have a low demand and a high risk of obsolescence, or are expensive. For these components one generally does not want to have high stock levels so work orders for these components are more critical than the work orders for other components that have a high demand or are quite cheap to keep in stock. Short and, above all, reliable throughput times for critical components are important for obtaining a high factory delivery performance at low costs.

Selective work order release.

From literature (Betrand and Wortmann 1981, Wiendahl 1987) it is well known that the use of a load based work order release function leads to shop throughput times that are more reliable than in case work orders are released immediately upon arrival. It is also well known that if we consider the total throughput time, that is the time that elapsed between the moment of arrival of the work order at the shop (and not the moment of release) and the finishing of that work order, the average of this total waiting time is larger than in the case of immediate release. From a number of simulations, the results of which can be found in Table 1, we may also conclude that

Utilization rate ↓		Shop tpt		Buffer tpt		Total lateness		Tardiness	
		avg	std	avg	std	avg	std	avg	std
85%	no WOR	33	37	0	0	0	22	7	17
	WOR	28	27	77	78	72	80	78	79
90%	no WOR	50	57	0	0	0	35	10	26
	WOR	42	40	90	89	82	93	92	90
95%	no WOR	101	113	0	0	0	69	21	51
	WOR	83	80	113	109	96	118	118	111

Table 1; Results of a simulation study for investigating the effects of the use of a work order release function on the reliability; work load norm set equal to the required minimum (see Whitt 1984) +10%; WOR means that a work order release function is used; the slack used in the case WOR=slack used in the case no WOR;
tpt=throughput time; avg=average; std=standard deviation.

if we use a load based work order release function, that then the reliability is less than in case of immediate release. Thus it makes no sense to apply load based work order release to all work orders that arrive at the shop. However, it is not clear whether this also applies to a selective use of work order release i.e. if we:

- release work orders for components that are critical immediately upon arrival at the shop (their planned release dates), independent of the load on the shop floor, so these orders will not be delayed in a buffer before the shop;
- use a release function based on a load limit for work orders for non critical components

This selective work order release function thus takes into account differences in criticality of the material at the down stream production phases. Work orders for critical components are always released upon arrival, whereas work orders for other components are only released if the load on the shop floor is less than the load limit. One may expect that in this case critical work orders have throughput times that are more reliable than in case no work order release function is used, since the load in the shop is more controlled. Moreover, if the number of critical work orders is not too high then even the average shop throughput time for these components will decrease if

the load limit is smaller than the average number of work orders on the shop floor in case no work order release function is used. For these components these effects lead to lower (safety) stock levels that are necessary for the delivery reliability in the next production phase. However the flow times for the non critical components will increase in average and variance, which will lead to higher safety stocks or lower delivery reliability for these components.

Proactive work order release.

As mentioned in the introduction we will also include the planning phase in our investigations. We will integrate the work order release mechanism and the planning phase by taking into account the planned work orders at work order release. Taking into account the planning phase means that we have knowledge of future work orders, so if the load on the shop floor is relatively low work orders can be advanced relative to their planned release times. If the load is relatively high non critical work orders are released later than their planned release dates. So the work order release mechanism now also can be used to shift peaks in the demand to later periods but also to shift work to earlier periods in order to fill up the gaps in the load.

4. The job-shop and experimental design.

The job shop we will use throughout this study consists of five work centers. Since real job shops often consists of more work centers, these five work centers must be seen as the five most utilized work centers. In job shops the throughput times are mainly determined by the waiting times at the work centers, which in turn are mainly determined by the utilization rate of the relevant work centre. For low utilization rates the waiting times will be small but as utilization rates approach 100%, the waiting times will go to infinity. Therefore, with respect to the throughput time, the most interesting work centers are those with a high utilization rate. In general this will be a subset of the set of all work centers and in our study we haven chosen a subset of size 5 (see also Conway et.al 1967).

The time between the arrival of two orders, or between two planned releases, for a product has a negative exponential distribution with a mean value of 1.1. Order routings are determined upon arrival. The routings are generated such that each work center has an equal probability of being selected as the first work center. After the first operation the probabilities of going to another work center or leaving the shop are set equal to 0.2, which results in an average routinglength of 5 operations per work order.

At each work center processing times are generated from a negative exponential probability density function with a mean value of 1 time unit. Set-up times and transportation times are considered to 0. The sequencing rule used at each work center is the First Come First Serve sequencing rule.

For the sake of common random numbers, each of the two product types has its own random number generators.

As work order release mechanism we use a the most simple (aggregate) mechanism, based on the total number of work orders on the shop floor: a work order may enter the shop floor if upon arrival the actual number of work orders on the shop floor is less than a certain, predetermined, load limit or as soon as a work order is finished and leaves the shop. If pulling forward is allowed then a work order will be advanced relative to its planned release moment if the load in the shop drops below the load limit and no work orders is waiting to be released. The load limit is set equal to the minimum required load limit (to get a throughput of 0.9 so all work orders can be worked up, see Whitt 1984) +10%, which for our situation leads to a load limit of 40 work orders.

It will be evident that a major role is played by the ratio of the arrival rates of work orders for critical products and work orders for non critical products. Therefore we will investigate the performance for the following values of the ratio of the arrival rate of critical orders and the arrival rate of all the orders: 0.10, 0.25 and 0.50.

The period length used is 40 units of time which is a little below the average throughput time in case no work order release mechanism is used (50). With regard to advancing work orders we assume in this study that at the start of each period only the (planned) orders within that period are known.

No distinction is made between critical and non-critical work orders when pulling planned orders forward; work orders are pulled forward in order of the planned release dates (FCFS).

Critical work orders are released immediate upon arrival and we expect that critical work orders will have more reliable throughput times than in case all work orders are released immediately upon arrival. However this will go at the costs of the performance for non critical work orders. Non-critical work orders are buffered if the work orders on the shop floor otherwise would exceed the norm. Therefore we expect these work orders to have less reliable throughput times than in case all work orders are released immediately upon arrival. To account for these changes in throughput time reliability one has to adapt the safety stock for both categories.

We assume that the planned work orders are known over a certain horizon, the planned order horizon, and that their due dates do not change over this horizon; thus over the planned order horizon the order due dates are frozen and the planned orders can be pulled forward.

The question now is whether there is a ratio for the valuation of both product types for which the total value of the stock will be less than in case no selective work order release rule is used. The lower the fraction of critical work orders the more reliable the throughput times of these orders will be since the less the work load on the shop floor will be above the load limit and thus the more stable the work load will be.

If work orders can be pulled forward from future periods then we expect that also the performance for the non-critical components will improve. In particular the tardiness performance will improve and there will be an increase of the earliness.

5. Discussion of the results.

We first performed a number of simulations to investigate the effect of selective work order release, thus planned orders are not pulled forward. The results from these simulations are given in Table 2. In these simulations work orders were not advanced relative to their due dates.

From this Table it can be observed that:

- The critical work orders indeed have a much better score on most of the performance measures used; only the average shop lateness and the average total lateness have become rather negative which indicates that these work order are finished too early. Most of the earliness can be explained by the fact that for the slack used to determine the lead time of the orders, the average slack for the situation without a work order release mechanism is used. It will be evident that by limiting the load on the shop floor the slack necessary for use on the shop floor will be less than in case the load is not limited and thus, if there is no buffer waiting time, the lead time will be less than in the unlimited situation.
- The better performance of the critical work orders goes at the cost of the performance of the non critical work orders: there total performance is much worse than in the situation without a work order release function.
- Compared to the situation with a work order release function, but without a difference in the criticality of the work orders, there is a slightly better overall score on almost all

util. rate =90%;		shop		buffer		lateness		lateness		tardiness	
		tpt		wait. time		shop		total			
		av	std	av	std	av	std	av	std	av	std
without work order release		50	57	0	0	0	35	0	35	10	26
ratio=50%	equal prio	42	40	78	77	-8	17	70	81	74	76
	overall	42	41	68	122	-8	17	60	125	67	120
	cr.orders	43	41	0	0	-8	17	-8	17	3	6
	ncr.orders	42	41	135	144	-8	17	127	147	132	143
ratio=25%	equal prio	41	40	59	71	-9	17	50	75	55	70
	overall	42	40	72	106	-8	17	64	109	70	104
	cr.orders	42	40	0	0	-8	17	-8	17	3	6
	ncr.orders	42	40	97	112	-8	17	88	116	93	111
ratio=10%	equal prio	41	40	61	69	-9	17	53	73	58	67
	overall	42	40	56	66	-8	17	48	70	53	65
	cr.orders	42	40	0	0	-8	17	-8	17	3	6
	ncr.orders	42	40	63	67	-8	17	54	71	59	66

Table 2; The results of the use of a selective work order release function for different ratios of the arrival rates of critical and non-critical orders in case work orders are not advanced;
av=average, std=standard deviation, cr=critical.

- the performance measures used for the situation with only 10% critical work orders.
- The performance on the overall tardiness is not better than or equal to the performance in the situation without a work order release function.
 - The influence of the number of critical work orders is mainly visible in the performance scores for the work orders for the non critical work orders. The smaller the number of critical work orders the better the performance scores for the non critical work orders. It is remarkable that there is hardly any influence of the fraction of critical work orders

util. rate =90%;		shop		buffer		lateness		lateness		tardiness	
		tpt		wait. time		shop		total			
		av	std	av	std	av	std	av	std	av	std
without work order release		50	57	0	0	0	35	0	35	10	26
ratio=50%	equal prio	43	41	73	77	-7	16	62	83	69	76
	overall	43	41	51	100	-7	16	40	105	51	98
	cr.orders	43	41	0	0	-7	16	-11	20	2	6
	ncr.orders	43	41	103	122	-7	16	91	129	100	121
ratio=25%	equal prio	42	40	51	70	-8	16	39	77	48	69
	overall	42	40	53	92	-8	16	41	98	51	90
	cr.orders	42	40	0	0	-8	16	-12	19	2	6
	ncr.orders	42	40	71	100	-8	16	58	107	68	99
ratio=10%	equal prio	42	40	49	59	-8	16	37	66	46	57
	overall	42	40	40	54	-8	16	28	61	38	52
	cr.orders	42	40	0	0	-8	16	-12	19	2	5
	ncr.orders	42	40	45	55	-8	16	32	63	42	54

Table 3; The results of the use of a selective work order release function for different ratios of the arrival rates of critical and non-critical orders in case work orders can be advanced;
av=average, std=standard deviation, cr=critical.

on the shop throughput time and the shop lateness, although, certainly if the number of critical work orders is 50%, the load limit a number of times will be exceeded by these work orders alone.

However the fact that non critical work orders only may enter the shop if the load is less than the load limit seems to be powerful enough to keep performance on the shop throughput time and the shop lateness more or less equal to the performance in case there is no difference in criticality.

A more practical setting is that at the beginning of a period the (planned) orders are known and that these work orders are released as soon as there is a release opportunity. Therefore we will not discuss Table 2 with respect to the safety stock requirements in the receiving stock point.

Next we investigated the effect of also using proactive work order release. In this situation (planned) work orders within the current period are advanced relative to their due dates. The results from this simulation study are summarized in Table 3. No prior is the situation with only a proactive work order release function, so all work orders have equal criticality.

Besides the fact that we may conclude that the observations from Table 2 also holds for Table 3, although they might be slightly stronger in a number of cases, we also observe that:

- the negative effects of the use of a selective work order release function for the non-critical work orders, are (slightly) decreased by also using proactive work order release;
- the shop performance remains more or less the same compared to the situation with only selective work order release;
- the performance on buffer waiting time, total lateness and tardiness, (average and standard deviation) has improved, compared to the situation with only selective work order release;

6. Safety stock requirements.

As we have seen differentiation is quite well possible if we use a selective work order release function. Now an important question is if this results in overall benefits, compared to the situation without work order release, for the safety stock or safety time required to deal with the delivery uncertainty. For critical work orders the performance regarding the standard deviation of the lateness is much better than in case no work order release mechanism is used. However for the non critical work orders this standard deviation has increased a lot. Thus it seems that nothing has really gained. However for critical work orders we generally require a higher delivery performance than for non critical work orders. Therefore there may be differences in delivery requirements such that the increase in safety stock for non critical components is offset by the decrease in safety stock for the critical components. In this section we investigate this question for the results obtained in the simulation experiments.

Suppose the critical and the non critical products both have the same inventory holding costs. In that case the possible benefits of the use of a selective work order release function must be found in an increase of the delivery performance for the critical products with the same, or lower, total inventory costs as in case no selective work order release function is used. Stated otherwise, it must result in a decrease of the total inventory with the same, or a higher, delivery performance as in case no selective work order release function is used. Let us therefore consider what happens to the inventory if we use a selective work order release function.

Suppose that without selective work order release, we have a safety stock, necessary to obtain the required delivery reliability, that for the critical products is equal to $k \times \sigma \times \mathfrak{d}_c$ and for the non critical products is equal to $\alpha \times k \times \sigma \times \mathfrak{d}_{nc}$, where σ is the standard deviation of the total lateness and \mathfrak{d} is the average demand. If the ratio of the demand for the critical and non critical products is β , then the total inventory holding costs are proportional to $(\beta \times k \times \sigma + (1-\beta) \times \alpha \times k \times \sigma) \times \mathfrak{d}$. Using a selective work order release function given a certain β , leads to an average lateness for the critical products and the non critical products of $l_{c(\beta)}$ resp. $l_{nc(\beta)}$ and to a standard deviation of the lateness of $\sigma_{c(\beta)}$ resp. $\sigma_{nc(\beta)}$. If we want to have the same delivery reliability as in the case without a work order release function we need to use the same k and $\alpha \times k$ as multipliers for the standard deviation of the lateness. However if we use the same lead time offset as in the case without a work order release function, the average lateness in general will be unequal to zero, so we have to account for this average lateness. This leads to a safety stock for the critical products of $\beta \times (k \times \sigma_{c(\beta)} + l_{c(\beta)}) \times \mathfrak{d}$ and a safety stock for the non critical products of $(1-\beta) \times (\alpha \times k \times \sigma_{nc(\beta)} + l_{nc(\beta)}) \times \mathfrak{d}$. So the total inventory holding costs are proportional to $\beta \times (k \times \sigma_{c(\beta)} + l_{c(\beta)}) + (1-\beta) \times (\alpha \times k \times \sigma_{nc(\beta)} + l_{nc(\beta)})$.

Now the question is if there are values α , β and k such that

$$\beta \times k \times \sigma + (1-\beta) \times \alpha \times k \times \sigma > \beta \times (k \times \sigma_{c(\beta)} + l_{c(\beta)}) + (1-\beta) \times (\alpha \times k \times \sigma_{nc(\beta)} + l_{nc(\beta)}) \quad (1)$$

In that case the use of a selective work order release function leads to lower total inventory holding costs.

With some simple calculations, using the data in Table 3, it turns out that only at very high values for k or very low values for α , depending on the value of β , the total inventory costs indeed can be decreased by using a selective work order release function. Data on this can be found in Table 4. This leads to the conclusion that if both the critical and the non critical products have the same inventory holding costs the practical situations where the use of a selective work order release function will lead to lower total inventory holding costs will be rare.

		fraction critical = 50%	fraction critical = 10%
$\alpha=0$	hor = 1	$k > 5.34$	$k > 16.7$
	hor = 2	$k > 2$	$k > -5.5$
$\alpha=0.5$	hor = 1	$k < -2.5$	$k < -2.5$
	hor = 2	$k < -0.23$	$k < 0.12$
$k=4$	hor = 1	$\alpha < -0.054$	$\alpha < -0.21$
	hor = 2	$\alpha < 0.025$	$\alpha < 0.0257$
$k=6$	hor = 1	$\alpha < 0.017$	$\alpha < -0.119$
	hor = 2	$\alpha < 0.034$	$\alpha < 0.0208$

Table 4. Some examples of combinations of k and α , for which equation (1) holds, for a number of situations.

Now suppose that the critical and the non critical products have different inventory holding cost and that for both kind of products we want to have the same delivery reliability, equal to the delivery reliability in the standard MRP situation.

If we denote the ratio of the inventory holding costs for the critical and the non critical products denote by r then the question is if there are realistic values for β , k and r such that:

$$(1+r) \times \beta \times k \times \sigma > r \times \beta \times (k\sigma_{c(\beta)} + l_{c(\beta)}) + (1-\beta) \times (k\sigma_{nc(\beta)} + l_{nc(\beta)}).$$

Using the data from Tables 2-4 we can conclude that this question can be answered (slightly) positively.

Example:

As we have seen in Tables 3, the standard deviation of the total lateness is 35 time units. If we assume that the lateness has a normal distribution then we need a total safety stock of $1.645 \times \sigma_{\text{lateness}} \times (\text{average demand})$ to obtain a delivery reliability of $\approx 95\%$. So we need a safety stock of $1.645 \times 35 \times 0.45 = 26$ units for both categories of products.

Now suppose we use the selective work order release mechanism and that the average fraction

of critical work orders is 50%. In that case, assuming that the lateness has still a normal distribution, we need a safety stock for the critical products of $1.645 \times 20 \times 0.5 \times 0.9 = 15$ units (0.9 is the average demand per unit of time) and a safety stock for the non critical products of $1.645 \times 129 \times 0.45 \approx 95$ units. However the average total lateness for the critical orders equals -11, so in general we already have a "safety" stock of $11 \times 0.45 \approx 5$ units. Therefore we only need $15 - 5 = 10$ units as real safety stock. For the non critical products the average total lateness is 91 so we need an extra "safety" stock of $91 \times 0.45 \approx 41$ units.

By using a selective work order release function, the safety stock for critical products decreases from 26 to 10, which is a decrease of 16 units. On the other hand the safety stock for non critical products increases from 32 to 136 units, which is an increase of 104 units. From this we can conclude that if the costs of inventory (also including the obsolescence risk, early finishing costs, the late delivery penalty etc.) for the critical products is at least $104/16 = 6.5$ times the inventory costs for the non critical products, the use of a selective work order mechanism leads to a better inventory performance (the same delivery performance at lower costs).

In case the number of critical work orders is only 10% we get the following figures:

Required safety stock for critical products: $1.645 \times 19 \times 0.10 \times 0.9 \approx 3$ units. Already available "safety" stock as a result of the early finishing: $12 \times 0.09 \approx 1$ unit.

Required safety stock for the non critical work orders: $1.645 \times 63 \times 0.81 \approx 84$ units. Extra required "safety" stock due to the late deliveries: $32 \times 0.81 \approx 26$ units.

In this case the required safety stock for the critical work orders decreases from $1.645 \times 35 \times 0.09 \approx 5$ units to 2 units which is a decrease of 3 units, whereas the safety stock for the non critical products increases from 47 to 110 which is an increase of 63 units. So in this case the costs of inventory for the critical products needs to be at least $63/3 = 21$ times the costs of inventory for the non critical products to lead to lower inventory costs using the selective work order release mechanism.

7. Conclusions.

In this study we have investigated whether the use of a selective and pro-active work order release mechanism, allowing to differentiate between critical and non critical work orders, leads to such a shift in the required safety stock that it is preferable to immediate release. The simulation

experiments conducted suggest that if inventory holding costs for both product types are the same only for very extreme differences in required delivery performance between the two product types selective work order release may be advantageous. However if the inventory holding costs for the critical products are much higher than for the non critical products, than selective work order release may lead to a reduction of the total (finished products) safety stock holding costs. Even if the number of critical products is about 10% and a delivery reliability is required of about 95%, the differences between the inventory holding costs need not to be that high for selective work order release to improve total inventory holding costs.

In case the early delivery of critical products is not used for reducing the safety stock but for reducing the lead time offset for critical products, the savings of the total safety stock costs will be somewhat less. However, in that case there will be a benefit at the customer side. Also if the lead time offset for the non critical products could be increased, the required safety stock for these products can be decreased due to a decrease of the average lateness, and again the use of selective work order release would become more attractive.

References.

- Baker, K.R., 1984, The effects of input control in a simple scheduling model. *Journal of Operations Management*, 4, 99-112.
- Baker, K.R., and Bertrand, J.W.M., 1981, An investigation of due date assignment rules with constrained tightness, *Journal of Operations Management*, Vol.1, No.3, February 1981.
- Bechte, W., 1982, Controlling manufacturing leadtime and work-in-process inventory by means of a load oriented order release, APICS Conference Proceedings.
- Bertrand, J.W.M., 1981, The effect of workload control on order flow times, *Operational Research '81*, North-Holland Publishing Company, 779-789.
- Bertrand, J.W.M., 1983^a, The effect of workload dependent due dates on job shop performance, *Management Science*, Vol.29, No. 7, 799-816.
- Bertrand, J.W.M., 1983^b, The use of work load information to control job lateness in controlled and uncontrolled release production systems, *Journal of Operations Management*, 3, 2, 79-92.
- Bertrand, J.W.M., 1985, Multiproduct optimal batch sizes with in-process inventories and multi work centres, *IIE Transactions*, 17, 157-163.

- Bertrand, J.W.M., and Wortmann, J.C., 1981, *Production Control and Information Systems for Component Manufacturing Shops* (Amsterdam: Elsevier).
- Bobrowski, P.M., and Park, P.S., 1989, Work release strategies in a dual resource constrained job shop, *Omega*, 17, 2, 177-188.
- Eilon, S., and Chowdhury, I.G., 1976, Due dates in job shop scheduling, *International Journal of Production Research*, 14, 223-237.
- Kanet, J.J., 1988, Load limited order release in job shop scheduling systems, *Journal of Operations Management*, Vol.7, No.3, 44-58.
- Karmarkar, U.S., 1987, Lot sizes, lead times and in-process inventories, *Management Science*, 33, 409-418.
- Karmarkar, U.S., Kekre, S., and Kekre, S., 1984, Multi item batching and minimization of queueing delays, *Working paper No.83-25*, William E. Simon Graduate School of Business Administration, University of Rochester, Rochester, New York.
- Kleinrock, L., 1975, *Queueing Systems* (New York: John Wiley and Sons).
- Melnyk, S.A., and Ragatz, G.L., 1988, Order review/release and its impact on the shop floor, *Production and Inventory Management*, 29, 13-17.
- Park, S., 1987, *The effects of input control in a dual constrained job shop*, dissertation.
- Park, S., and Salegna, G., 1992, Load smoothing with feedback in a bottleneck job shop, *paper of the TIMS/ORSA presentation Orlando 1992*.
- Ragatz, G.L., 1985, *An evaluation of order release mechanisms for job shops*, dissertation.
- Ragatz, G.L., and Mabert, V.A., 1988, An evaluation of order release mechanisms in a job shop environment, *Decision Sciences*, 19, 167-189.
- Solberg, J.J., 1981, Capacity planning with a stochastic workflow model, *AIIE Transactions*, 13, 116-122.
- Wein, L.M., and Chevalier, P.B., 1992, A broader view of the job shop scheduling problem, *Management Science*, Vol.38, No.7, 1018-1033.
- Whitt, W., 1984, Open and closed models for networks of queues, *AT&T Bell Laboratories Technical Journal*, Vol.63, No.9, 1911-1979.
- Wiendahl, H.-P., 1987, *Belastungsorientierte Fertigungssteuerung* (Munche: Carl Hanser Verlag).
- Wiendahl, H.-P.(hrsg.), 1991, *Anwendung der Belastungsorientierte Fertigungssteuerung* (Munche: Carl Hanser Verlag)