Performance prediction and diagnosis in two production departments

Paul P.M. Stoop
Department of Operations Planning and Control, Eindhoven University of Technology, The Netherlands

J. Will M. Bertrand
Department of Operations Planning and Control, Eindhoven University of Technology, The Netherlands

In complex, job-shop like production departments it is usually very difficult to predict the near-future logistic performance as well as to explain or diagnose objectively why and how a certain performance has been achieved. Presents a prediction and diagnosis method that has been developed and tested in two production departments. Describes how the method provides realistic logistic performance targets in the short term with respect to the throughput of a production department and order completion times. The method also quantitatively determines ex post the impact of occurred disturbances on the realized performance. In the pilot project the method provided a clearer insight into relationships between logistic key variables, gave decision support to the capacity allocation decision, and generated reliable performance targets for the short term. More importantly, the actual performance became more open to discussion due to the objective explanation of the achieved performance, which opens the way to performance improvements.

Introduction
Performance measurement is a popular subject today in the academic field as well as in practice. Modern organizations know that a good financial performance alone does not show that one is doing well overall. Although costs remain a major aspect of organizational performance, other aspects such as quality and timeliness may give organizations a competitive advantage. Due to changes in the market, the focus on the organizational performance has been changed over time. Especially the measuring of non-financial aspects has gained more attention (e.g. [1]).

In the literature about performance measurement different directions can be distinguished. First, there is a part of the literature that deals with the question of which performance measures should be used in a performance measurement system. The more general papers in this class discuss methods and criteria for developing the right measures (e.g. [2-4]), whereas the more specific literature enumerates selections of performance measures that are appropriate for a specific kind of organization or process (e.g. [5-7]).

The issue of implementing performance indicators is also addressed here, often illustrated with case studies (e.g. [8-9]). Second, all kinds of classifications of performance measures have been developed. Examples of these classifications are:

- financial versus non-financial performance indicators, e.g. [10];
- efficiency versus effectiveness performance measures, e.g. [11,12]; and
- input versus output performance indicators, e.g. [13].

An interesting point that is hardly mentioned in the performance measurement literature is that performance measurement actually comprises more than measurement alone. The goal of the measurements is to evaluate the achieved performance, which means that the actual performance is compared with performance targets. After a further analysis or diagnosis that explains how the actual performance has been established, one can start up appropriate actions for performance improvements. This process of measurement, evaluation, diagnosis, actions, measurement, etc., is regarded as a closed loop process continuously leading to performance improvements[14], and is comparable to the plan-do-check-act cycle[15]. Each part of this process of continuous improvement can and should be treated individually to reach the highest possible performance improvements; measurement alone does not automatically lead to improvements as is sometimes suggested in the literature (e.g. [16]). So, instead of speaking about performance measurement, we think it is better to speak about performance management[17].

In this paper we make the steps in the performance management process more explicit. We take the shopfloor with respect to the logistics performance as an example. At this level the actual performance is measured and reported frequently, but a thorough performance evaluation and diagnosis is seldom executed (and if these steps are executed then they are done so only in a qualitative rather than in a quantitative way). In our attempt to make the explanation more quantitative we will report in this paper on the implementation and use of a performance prediction and diagnosis system in two job-shops. The emphasis will not be placed on the system itself, but on the way people can use the system to achieve performance improvements.

In the next section of this paper we will describe the characteristics of the two production departments in which the prediction and diagnosis method has been implemented. The newly developed prediction model and the way the prediction and diagnosis method had been used will then be described. The penultimate section will discuss the results obtained by using the method and the final section presents conclusions and suggestions for further research.

Case descriptions
The two departments considered are production departments of a Dutch aircraft manufacturer. Both functional departments specialize in metal removing operations, such as turning, milling, drilling, and benchworking. The two departments specialize in small
products, such as special screws, pins, guards, and spacers. After a large reorganization the two departments were formed out of one large production department. Since then, both departments operate more or less autonomously, although there is still interaction about borrowing capacity and outsourcing works. The difference between the departments refers to the kind of products which have to be made. In the first department, which we will call department A, all newly developed products are being processed and all rush orders, whereas department B produces the recurring part of the products. Further, department A produces orders from department B in case of low utilization. The focus in department A is directed on flexibility, whereas department B concentrates on costs and efficiency.

In department A, there are 16 work centres. Department B consists of nine work centres. Average processing times are 3.4 and 2.5 hours respectively. Average order flow times are about 3.5 and 2.5 weeks for departments A and B respectively. In each department the work is done in two eight-hour shifts, five days a week. Overtime can be used to make up order backlogs. Each shift has about 12 operators and one shift leader. Most operators can perform two or more tasks, but the operator flexibility in department A is higher than in department B. The most important performance indicator for the shift leaders is the throughput of the department, measured in the total number of hours processed in a week. The feedback of the realized throughput is given weekly.

The goods flow of the total manufacturing process is planned and controlled by a materials requirement planning (MRP) system. Both production departments are part of the chain that supply products to the final assembly stage. Each production department has its own schedulers who have to control the progress of orders in the departments they are responsible for. The two departments considered here have two schedulers. Each of them manages a class of products. The most important performance measure for the schedulers is the delivery reliability. Splitting up orders, changing priorities, and negotiating with colleague schedulers are the main activities of the schedulers to keep the delivery reliability as high as possible.

Logistics management as well as operations management of the departments were interested in a method that could help with a more objective performance evaluation and diagnosis regarding the shops' throughput and order completion times. Although they felt that one could explain intuitively why a specific performance had been achieved, they recognized that this intuition lacked a quantitative basis. To their opinion, the outcome of the pilot project could be either a quantitative verification of what was already known in a qualitative way, or could give new insights in the processes at the shopfloor. The second reason for participating in the development of a performance evaluation and diagnosis method was that there was a need for objective short-term performance targets for performance evaluations. The strong fluctuations in the actual performance from week to week made medium-term performance targets (that were set fixed for a month ahead) unrealistic for each week.

The performance prediction and diagnosis method

In our proposed method of performance prediction and diagnosis a newly developed prediction model holds an important position. The prediction model consists of prediction rules to estimate a production department's throughput (in terms of hours processed) for the next week and to estimate order completion times. Queuing theory is taken as starting point for the prediction rules. Because the focus on the performance of production departments is usually short-term directed, the developed prediction rules take into account as much as possible the actual state of the department; they are state-dependent. This is based on the assumption that the short-term logistic performance of a production department to a large extent depends on actual situational factors of the department, such as the actual work in process and the actual capacity utilization rate. In this respect, the prediction model differs from existing state-independent or steady state (queuing) models such as the Queuing Network Analysier [18], which can only give the average performance over a relatively long period of time. For the two cases, a week is taken as prediction period.

This period of time corresponds with the current performance reporting period. The variables we take into account for the specific cases are depicted in Figure 1. The structure of the shop is modeled by the work centres in the shop and the number of machines per work centre. The actual state of the shop consists of all orders in the shop (work in process) with their characteristics: current position in the queues for the work centres, remaining routing and processing times, operation due dates, and indications for rush orders. Another variable is the work supply that is expected to be released by the schedulers to the production department in
the next period. Also, the available operator and machine capacity is an important factor which determines the logistic performance of the next period. This variable in particular may fluctuate seriously from week to week, which underpins the need for a state-dependent model for the short-term performance. Finally, two policies are considered to have a great influence on the performance: the order release policy and the order sequencing policy. The order release policy determines the work supply for the next week. In the two production departments there will be as much work released as had been processed in the preceding week (input-output control). The order sequence rule determines the sequence in which orders are processed at the work centres. In the two departments, operators first should process rush orders, and then orders in sequence of operation due date.

Because it is our main purpose to describe how people actually use the prediction and diagnosis method and what results can be obtained, no specific descriptions of the state-dependent prediction rules for the throughput and order completion times are given in this paper. For a description of these rules we refer to [14]. For verification of the assumptions underlying the prediction rules logistics management and operations management of the two production departments were consulted many times. They were also asked for suggestions and help to facilitate the process of implementation. The prediction rules were programmed in a software package which runs on stand-alone personal computers.

In Figure 2 a schematic overview of the performance prediction and diagnosis method is depicted. The method starts with making a prediction of the throughput for the next week. In the morning of each first day in the week the schedulers make a download of the actual state (work in process) and the realized throughput of the previous week of the department. These data are given to the shift leaders. The shift leaders read these downloaded data in the software program and enter the expected available machine and operator capacity for the coming week. Further, a linkage has to be made between operator and machine capacity (for example, in case an operator can operate two machines in a work centre at the same time).

Next, the shift leaders run the model to make the prediction. The model first determines which orders are expected to be released for the next week. Based on the expected work supply, the actual work in process and the expected available capacity, the model calculates the expected throughput and order completion times. An example of a throughput prediction has been given in the left-hand side of Table I. Analysing this predicted performance, shift leaders can decide to make another capacity allocation decision. In the example of Table I it seems wise to remove one operator (40 hours) from work centre 2, where the amount of available capacity is much higher than the expected throughput, to work centre 8, where the actual work in process is relatively high compared with the expected amount of allocated capacity. This new allocation decision can be entered in the model to make a new prediction. This what-if process can be continued until a shift leader is satisfied with the allocation with respect to the expected throughput. This predicted throughput now serves as a short-term performance target. The operations manager had been asked to use this performance target not as a target to judge individual shift leaders, but as an indication of the performance that could be achieved. The "target" should open the discussion between everybody who could influence the
performance about how the maximum performance could be reached. For example, it was asked which preparatory activities (for example, checking on time the availability of materials, drawings or production tools) should be initiated to make it possible to achieve the predicted throughput.

The next phase in the method is diagnosis or explanation of the latest week's performance. For that purpose shift leaders enter the actual allocated operator and machine capacity of the latest week. Then, the prediction model is run to make a so-called post-prediction of the logistic performance. This post-predicted performance is the performance that could have been achieved, knowing the actual available capacity and the actual work supply that has been registered during the week. The difference between post-predicted and actual performance will be subject to discussion and should be explained by other causes than differences between expected and actual available capacity and work supply. In the example of Table I, one usually looks first at the difference between the actual throughput and the post-predicted throughput of the whole department (301.2 versus 397 hours). In case of such large deviations one next focuses on individual work centres in order to determine at which work centre there is a serious deviation (in this example work centres 2 and 9 are the most problematic). Sometimes shift leaders can explain these differences, but often

Table I
Throughput predictions for a specific week for department B (all values in hours)

<table>
<thead>
<tr>
<th>Work centre</th>
<th>Queue length</th>
<th>Predicted throughput</th>
<th>Expected capacity</th>
<th>Actual throughput</th>
<th>Post-predicted throughput</th>
<th>Actual capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
<td>19.0</td>
<td>24</td>
<td>3.6</td>
<td>19.0</td>
<td>24</td>
</tr>
<tr>
<td>2</td>
<td>83.0</td>
<td>118.2</td>
<td>160</td>
<td>93.9</td>
<td>122.0</td>
<td>152</td>
</tr>
<tr>
<td>3</td>
<td>227.2</td>
<td>0.0</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>59.4</td>
<td>32.0</td>
<td>40</td>
<td>38.7</td>
<td>32.0</td>
<td>40</td>
</tr>
<tr>
<td>5</td>
<td>6.0</td>
<td>19.0</td>
<td>24</td>
<td>12.3</td>
<td>19.0</td>
<td>24</td>
</tr>
<tr>
<td>6</td>
<td>145.7</td>
<td>38.0</td>
<td>48</td>
<td>49.6</td>
<td>45.0</td>
<td>56</td>
</tr>
<tr>
<td>7</td>
<td>69.6</td>
<td>32.0</td>
<td>40</td>
<td>24.0</td>
<td>32.0</td>
<td>40</td>
</tr>
<tr>
<td>8</td>
<td>275.5</td>
<td>38.0</td>
<td>48</td>
<td>31.2</td>
<td>38.0</td>
<td>48</td>
</tr>
<tr>
<td>9</td>
<td>46.0</td>
<td>72.3</td>
<td>128</td>
<td>47.9</td>
<td>90.0</td>
<td>112</td>
</tr>
<tr>
<td>Total</td>
<td>912.4</td>
<td>368.5</td>
<td>512</td>
<td>301.2</td>
<td>397.0</td>
<td>496</td>
</tr>
</tbody>
</table>
they cannot. In the latter case they have to go
to the work centre and ask the operator who
is working there in order to obtain an expla-
nation. This explanation (e.g. continual diffi-
culties with setting up a machine) may be the
basis for performance improvements at that
work centre.

When the prediction and post-prediction
have been made, the results are fed back to
the schedulers. Because their primary task is
to control the flow of individual orders, they
have a look at the orders which are predicted
to be completed late. These are the orders
with a predicted completion date that is later
than the planned due date of the latest opera-
tion in the shop. The software model automat-
cally makes a list of orders that are expected
to arrive tardy. The schedulers have two pos-
sible actions to try to let the orders arrive on
time. First, they can decide to give an order a
priority status (possibly after some consulta-
tion with the assembly scheduler). The sec-
ond possible action is to ask the shift leaders
to allocate more capacity at the work centres
which the orders have to go through in their
remaining routing. Because shift leaders
have already optimized the throughput with
their capacity allocation, there may arise a
conflict between the maximum throughput
that can be achieved and the best possible
delivery reliability.

Results
To show the quality of the predictions of
order completion times, the prediction error
for all orders which were completed in the
departments over a period of ten subsequent
weeks have been determined by taking the
difference between the predicted completion
date and the real completion date. For each
department, the average as well as standard
deviations of the prediction errors are shown
in Figure 3. Three different kinds of predic-
tion errors are shown: the prediction error of
the:

- state-dependent prediction model;
- rule of thumb that the schedulers use (i.e.
  the sum of standard flow times per work
  centre); and
- planned due dates of the MRP system.

It can be observed that all predictions on the
average are too optimistic, but the reliability
(i.e. the standard deviation of the prediction
error) of both the state-dependent model and
the scheduler method is far better than the
reliability of the MRP “predictions”.

The prediction results of the throughput
predictions of department B are presented in
Table II. From this Table it can be observed
that there are work centres where it is very
difficult to predict the throughput (e.g. work
centres 8 and 9 which have a relatively large

Figure 3
Quality of completion date predictions

<table>
<thead>
<tr>
<th>Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>-10</td>
</tr>
<tr>
<td>-20</td>
</tr>
</tbody>
</table>

Department A

Department B

Model "Planner" MRP

Model "Planner" MRP

Prediction type

Average error Standard deviation error

Error = predicted completion date - actual completion date
standard deviation of the prediction error).
The shift leaders stated that they had been aware of this but up to now they had not been able to quantify these intuitive feelings. By using the state-dependent prediction model, the effects of differences between the expected available capacity and the actual allocated capacity on the throughput could now be made quantitative and objective.

Differences between expected and actual available capacity and work supply turned out to explain a large part of the observed performance deviations. The search for other factors that may cause differences between the post-predicted and actual performance is the next step in the performance diagnosis phase. The performance prediction and diagnosis method does not give support to this. Just like the analyses of differences between post-predicted and actual performance, this phase should be a co-operative search process where all who can influence the performance should take part. The qualitatively explanatory factors which are found by the participants of these discussions should be tried to quantify their real individual influence on the performance. Examples of these qualitative causes of throughput and completion date deviations in the cases mentioned by shift leaders and schedulers are: differences between calculated and actual processing times, rework (i.e. product repairs which is not registered), efficiency differences between work centres, and applying a sequencing rule that differs from the operation due date policy. Ideally, these factors should be included in the prediction model to get more reliable performance predictions on the one hand and more exhaustive explanations via the post-predictions on the other hand.

Conclusions and suggestions for further research

The presented method for performance prediction and diagnosis has proven to be a useful method for decision makers at the shopfloor level. The strength of the method is that, due to its formalized way of working, people are forced to think about the relationships between the expected performance, the actual performance, and the performance which could have been realized. The state-dependent prediction model calculates more realistic performance targets for the short term compared with state-independent ways to set performance targets which stimulates a proactive decision behaviour with respect to the resource allocation decision. The objectively determined post-predicted performance targets enable a useful, realistic and objective discussion in the performance diagnosis phase between people responsible for and causing the logistics performance of a production department. This communication leads to a better understanding of one another’s problems and can help to avoid future performance problems. Because the method can be applied in regularly held work meetings, the method will hardly cost extra time. Only the development of the prediction model can take a considerable amount of time, because the model should take into account the specific characteristics of the production department the model is developed for.

The execution of the subsequent steps in this method are not restricted to the type of departments we described in this paper. The subsequent steps of performance prediction, evaluation and diagnosis can be applied in every kind of department. The effect of using a state-dependent prediction model on the improvement in the prediction reliability in the short term, however, is expected to be the largest in job-shops because the predictability in this type of shops is the relatively low. More empirical research is suggested to test the performance prediction and diagnosis method in other types of departments, such as flow shops. Finally, simulation

<table>
<thead>
<tr>
<th>Work centre</th>
<th>Number of observations</th>
<th>Average error</th>
<th>Standard deviation error</th>
<th>Minimum error</th>
<th>Maximum error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18</td>
<td>7.9</td>
<td>9.2</td>
<td>-3.6</td>
<td>25.8</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
<td>-0.2</td>
<td>21.2</td>
<td>-48.0</td>
<td>53.0</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>9.1</td>
<td>13.0</td>
<td>-8.0</td>
<td>34.5</td>
</tr>
<tr>
<td>4</td>
<td>18</td>
<td>-3.1</td>
<td>14.0</td>
<td>-23.0</td>
<td>43.3</td>
</tr>
<tr>
<td>5</td>
<td>18</td>
<td>3.2</td>
<td>7.4</td>
<td>-13.0</td>
<td>26.0</td>
</tr>
<tr>
<td>6</td>
<td>18</td>
<td>3.6</td>
<td>15.3</td>
<td>-26.8</td>
<td>46.7</td>
</tr>
<tr>
<td>7</td>
<td>18</td>
<td>-2.6</td>
<td>7.8</td>
<td>-18.9</td>
<td>9.7</td>
</tr>
<tr>
<td>8</td>
<td>18</td>
<td>6.5</td>
<td>29.5</td>
<td>-57.0</td>
<td>73.9</td>
</tr>
<tr>
<td>9</td>
<td>18</td>
<td>15.7</td>
<td>20.2</td>
<td>-16.0</td>
<td>52.4</td>
</tr>
<tr>
<td>Total</td>
<td>18</td>
<td>39.2</td>
<td>41.2</td>
<td>-23.7</td>
<td>112.3</td>
</tr>
</tbody>
</table>
experiments can be used to develop the “best” state-dependent prediction rules for a specific situation.

References