PERGAMON

International Journal of Project Management 20 (2002) 75–80

INTERNATIONAL JOURNAL OF PROJECT MANAGEMENT

www.elsevier.com/locate/ijproman

Project management applications of the theory of constraints beyond critical chain scheduling

H. Steyn *

Department of Engineering and Technology Management, University of Pretoria, Pretoria, 0002, South Africa

Abstract

The application of the Theory of Constraints (TOC) is an approach that can be used to develop a variety of management techniques. As a result of the multi-disciplinary nature of project management a variety of different applications within project management are possible. The application of the TOC approach to project scheduling led to the development of the "critical chain" technique that has been the subject of discussions in literature. This paper explains why TOC was initially applied only to project scheduling. A second application of TOC is to manage resources shared by a number of concurrent projects. The basic principles of this second application are discussed in this paper. In addition to the above-mentioned two applications the TOC approach can also be applied to other areas of project management such as project cost management and project risk management. © 2001 Elsevier Science Ltd and IPMA. All rights reserved.

Keywords: Theory of constraints; Critical chain; Project risk management; Project schedule; Project network; Resource scheduling; Project cost management; Multi project; Resource allocation; Human behaviour

1. Introduction

The Theory of Constraints (TOC) is an approach that is used to develop specific management techniques. It was first popularised by the novel, *The Goal* [1], that applied the principles to operations management. Since 1997 it has found application in two areas within project management. The first application is scheduling of a single project to reduce project duration and simplify project control. This is the main theme of the novel *Critical Chain* [2]. Only towards the end of this novel (Chapter 24) there is some indication of a further application to allocate resources that are shared by concurrent projects.

The assumptions and principles underlying the application to the scheduling of a single project have been investigated in an earlier paper [3] that, like most literature to date, covers only the application to individual projects. A few authors [4,5], however, do mention the application to multiple concurrent projects that is already being used in practice. The basic principles of this application are discussed in this paper. This paper further illustrates that the TOC approach can also be applied to areas such as project risk management and project cost management.

2. Scheduling a single project — critical chain scheduling

Traditionally time estimates for individual activities contain some provision for contingencies. Critical chain scheduling aggregates these provisions into a *project buffer*. As a result of aggregation the total contingency reserve and thus project duration can be reduced [3]. Because contingency reserves are removed from individual activities and aggregated in a project buffer, commitments regarding the completion date are only made at project level. All other people working on the project only make realistic estimates, communicate expectations on activity durations and attempt to meet the realistic estimates. This implies that all due dates on individual activities and sub-projects are eliminated [3]. Team members know that there is some allowance for contingencies (a buffer) but if they need to use some of it, they have to make this need visible and motivate it. This requires a mindset that is quite different from the one that normally prevails where all people working on activities are required to commit themselves to due dates [3].

Advocates of TOC often claim that in traditional project management provisions for contingencies are

* Tel.: +27-12-420-3647; fax: +27-12-362-5307.
E-mail address: herman.steyn@eng.up.ac.za

0263-7863/01/$20.00 © 2001 Elsevier Science Ltd and IPMA. All rights reserved.
PII: S0263-7863(00)00054-5

too large because people might tend to commit themselves only on time estimates that they could meet with a high level of certainty. Also, managers or co-ordinators at each level within an organisational hierarchy could build in their own reserves on top of the reserves built in by the people reporting to them. It is furthermore believed that there is a tendency to waste time as a result of reserves that are liberal. Should these assumptions be correct (the author supports this view), it would strengthen the case for critical chain scheduling. However, should the assumptions be incorrect the TOC way of scheduling will still reduce project duration because of the effect of aggregation [3].

In traditional project scheduling the critical path does not take resource availability into account and resource allocation is done as an additional step. The critical chain, however, takes resource availability into account to the extent that activities done by the same resource are scheduled in series. To prevent non-critical activities from delaying critical ones, *feeding buffers* are placed where non-critical paths feed into the critical chain. A feeding buffer contains most or all of the contingency reserve relating to the relevant non-critical path. Proper management of the feeding buffers prevents the critical chain from changing during project execution and leads to a rigorous project plan [3].

3. Reasons for the emphasis on scheduling

In applying the TOC concepts to project management Goldratt and other advocates of the TOC approach limited themselves initially to project time management and did not venture into other aspects of project management. Here is why: the TOC approach prescribes that the constraint of a system has to be identified and attention focussed only on the constraint until it is not a constraint any more. Project duration is considered the major constraint of projects in general. Three reasons for identifying project duration as the constraint are discussed below.

3.1. Positive cash flow obtained faster

Project costs often escalate as a result of extended duration. As the schedule of a project with a fixed scope increases, costs usually increase [4]. Furthermore, revenues are often reduced as a result of delays. Project duration and project budget are sometimes traded off with the assumption that reduced project duration would be possible only with an *increase* in project costs. The most common time–cost trade-off technique with this assumption is one that was developed in 1957 by the DuPont Company and Remington Rand Univac. This is often referred to as the Critical Path Method or CPM. The reason for this apparent discrepancy is that authors on techniques such as CPM often consider the project itself as an end on its own, while in reality the project often only exists to create another system or product. The objectives of many projects are to establish products, services or other outcomes that exist well beyond project closure. The *product* life cycle as opposed to the *project* life cycle has to be considered. The project life cycle only forms the first phase of the product life cycle. Although the product life cycle is described in systems engineering literature and in certain literature on project management such as BS6079 [6], it is neglected in much of the other literature on project management that consider hand-over, project closure and the lessons learned as the ultimate goal. Each system has a goal. If the goal of the project is to deliver a specific product or system, the resulting product or system should have some beneficial goal. In business the goal normally is to generate money through sales. The objective of the project should be to maximise the revenue of the resulting system or product. Total product life-cycle costs and total product life-cycle benefits (rather than the project budget) should be the main considerations. To plan and execute a project normally results in an initial negative cash flow. The objective of the rest of the product life cycle should be to obtain some benefit (a positive cash flow in the case of projects done for business reasons). If project risks are taken into account, the positive cash flow should generally far exceed the negative cash flow. Therefore, if a project were to be delayed, the overall effect on cash flow (or other positive outcome) could be expected to be impaired. If the *product* life cycle (as opposed to *project* life cycle) is considered, it often makes sense to reduce project duration in order to improve the internal rate of return even if it requires greater expenditure during project planning and execution. The optimisation of total product life-cycle benefit is the *first* motivation for the assumption that project duration is a very important constraint.

3.2. Contingency cost of delays

A *second* motivation for considering duration as an important constraint is that the contingency cost of project delays could be very high. For example, market share could be lost if a project is delayed. This is mentioned in the context of new product development in *Critical Chain* [2].

3.3. Preventing changes to stakeholder needs

A *third* reason is that it is believed that extended project duration not only leads to escalation of overhead costs, but also leads to scope changes because stakeholder needs could be expected to change over time. While scope changes are often not allowed, there are cases where scope changes during the project life cycle are permitted. When a project is completed early there is less time for the stakeholders' needs to change and thus certain scope changes could be eliminated. This is especially applicable to the development of products where changes in market needs or in technology could be addressed by subsequent product generations. (Recent examples include the development of mobile telephones and certain computer-related products). In industries where the project deliverable has an extended life (e.g. in the construction industry), expediting the project would also eliminate scope changes. However, in certain cases it might be argued that a project should not be rushed in an attempt to prevent such scope changes.

These three reasons provide the motivation for the initial focus of applying TOC principles to reduce project duration. Obviously the development of a scheduling technique for a single project had to be developed before a technique for multiple projects could be developed.

Three applications of TOC beyond the "critical chain" application to single project scheduling are discussed below, viz. application to multi-project environments, cost management and risk management.

4. Application of TOC principles to managing resources in multi-project environments

It is estimated that up to 90%, by value, of all projects are carried out in the multi-project context [7]. The problem of allocating resources to concurrent projects is not new: it is fundamentally the same as the problem of scheduling a "job shop" where multiple jobs are routed through a workshop with a number of dissimilar machines or resources such as lathes, milling machines, drill presses and heat treatment facilities. Job shop scheduling received ample attention in the 1960s and 1970s.

Elton and Roe [8] say that TOC works well when dealing with individual projects but falls short in explaining how companies could best manage a portfolio of projects. Senior managers, therefore, need to supplement the guidelines given by Goldratt [2] with other advice. They maintain that managers need to know how to juggle a portfolio of projects, and that the only advice that the book (*Critical Chain* [2]) offers on that subject is to make sure to allocate resources carefully across projects to minimise the constraints on shared resources. What managers need is a wider perspective of aggregate project planning such as is explained by Wheelwright and Clark [9]. The statement by Elton and Roe [8] "... to make sure to allocate resources carefully across projects..." is a fair description of what the novel, *Critical Chain* [2] offers on the topic of managing multiple projects. However, in addition to the simplified description of the approach to allocate resources to multiple concurrent projects described in the novel, a method has been developed to address resource allocation in the multi-project context. This material is still mentioned only superficially in literature in the public domain and is currently addressed only in proprietary course material [10]. The basic principles are discussed below.

The TOC approach involves five steps [1,2]:

1. identify the constraint(s) of the system;
2. decide how to exploit the constraint(s);
3. subordinate non-constraints to the decision(s) on exploiting the constraint(s);
4. elevate the constraint(s) (in other words: take steps to "widen the bottleneck"); and
5. by returning to Step 1 above, determine whether a new constraint has been uncovered, rendering the constraint under consideration a non-constraint or less critical.

For a single project Goldratt [2] defines the critical chain (the path that determines project duration) as the constraint. In a multi-project environment concurrent projects commonly depend on a pool of shared resources. A resource that is overloaded limits the number of projects that the organisation can execute and thus presents a further constraint. The objective is to maximize the number of projects that the organization can do concurrently. The concept of the critical chain removes resource contention within a single project but it does not address resource contention among projects. Therefore, another way had to be found to allocate scarce resources to multiple concurrent projects.

Applying the five-step TOC approach implies that the resource (or resources) constraining the capacity has to be identified (TOC Step 1). Then the sequence of work to be performed by such a *Capacity Constraining Resource* (CCR) has to be scheduled (TOC Step 2). The scheduling of other resources should be subordinated to this schedule (TOC Step 3). To accomplish Step 3 a *CCR Buffer* is inserted before each task performed by the CCR. (This type of buffer is sometimes also called a "drum", "strategic resource" or "bottleneck buffer"). Such buffers are illustrated in Fig. 1 where CCR indicates activities that are performed by the Capacity Constraining Resource. These buffers prevent the CCR from having to wait when a preceding activity has been delayed. In this way, non-constraints are subordinated to the constraint and the risk

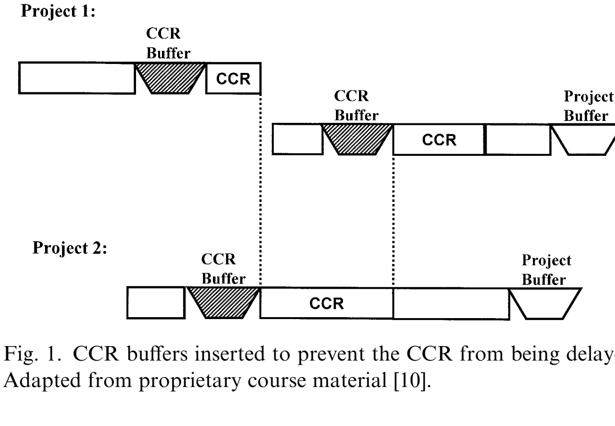

Fig. 1. CCR buffers inserted to prevent the CCR from being delayed. Adapted from proprietary course material [10].

of a resource, that is not the capacity-constraining one, delaying the capacity constraining one is addressed.

To make the best possible use of existing capacity before money is invested in additional capacity, money is invested to acquire additional resources (TOC Step 4) only after Steps 2 and 3 have been taken.

Adding additional resources to alleviate the pressure on the resource might result in a situation where the resource under consideration is not the constraint any longer. The manager should therefore go back to Step 1 to identify the current constraint (TOC Step 5). This ensures that the workload of all resources that are under pressure (and could thus limit the capacity of the organisation to handle projects) receive systematic attention at the right time.

To identify the constraint, managers should select the resource with the highest workload. More than one individual might believe that his/her workload is the highest and it might be difficult to identify which one actually has the highest load. This need not be a problem because the five steps imply an iterative process. Once Step 4 is reached, the capacity of the resource identified as the constraint is increased and when Step 5 is reached there is an opportunity to select another resource as the constraint. Therefore, if there is a resource with a workload that is higher than that of the one that is initially selected as the constraint, it will be addressed during a later iteration. One should therefore steer clear from extended arguments about which resource to select and just pick one that you feel is under pressure.

Software has been developed to facilitate resource allocation by means of the TOC approach and a number of different software systems for this purpose are commercially available. While maximising the number of projects that an organisation can handle, these programmes also maintain the principles of reducing project duration on each individual project.

Since any large project is comprised of a number of smaller sub-projects, a large project resembles a multi-project situation. For such a large project it is often practical to consider the sub-projects as separate projects each with a separate critical chain. Whenever sub-projects (a) are executed concurrently, (b) are scheduled to have separate critical chains, and (c) share common resources, they should be treated in the same way as multiple projects.

5. Application of TOC principles to project cost management

Turner [11] differentiates between planned project cost and budgeted project cost. When a work package is assigned to a team the planned cost (or realistic estimate) is the amount that is communicated to the team as the amount they should spend. "... every team knows that there is some contingency. But if they need to use it, they should come back and negotiate for it. They should not be given it automatically as they will use it and more". The principle that Turner [11] applies to cost management here is the same one utilised for schedule contingencies in critical chain project scheduling: the contingency or buffer is "owned" by the project manager and is not disaggregated and distributed to the various teams and/or individuals.

When provisions for schedule contingencies are aggregated, the required total contingency is reduced [3]. Like time estimates, cost estimates are also subjected to uncertainty and the principle of aggregation applies to cost estimates in the same way that it applies to time estimates. Suppose n activities each has a cost estimate with standard deviation σ and the person responsible for each activity builds in a contingency reserve. The total reserve will then be proportional to $n \cdot \sigma$. However, if the people responsible for the activities all make realistic estimates without provision for contingencies and only the project manager provides a contingency reserve, the required reserve at project level will be proportional to $(n)^{1/2} \cdot \sigma$ that is significantly less than $n \cdot \sigma$. Thus application of principle of aggregation of reserves results in a significant reduction of project cost.

In many cases, an employee (say a programmer or engineer) will not be rewarded for spending less than the amount allocated for a task. However, the employee is held responsible for technical problems that might result from his/her work. Therefore, once a figure that includes a contingency allowance has been approved at work package level, it is rational for the individual to spend the "available" money to reduce technical risks that he/she might be held accountable for to the lowest possible level. Therefore, one may find that the total approved amount is spent, regardless of the size of the provision for contingencies and regardless of whether or not an event that requires funds from the reserve occurs. When an unforeseen event requires additional funds, overspending is likely to occur. This explains the behaviour that a team "... will use it (the contingency reserve) and more" mentioned by Turner [11].

Budget reserves should, therefore, not be allocated to levels below the project level. Here, as in the case of project scheduling, a new paradigm is required: cost estimates made at work package level are merely indications of what would be realistic if nothing goes wrong, and are *not* commitments. Broadening the critical chain approach to cost management implies that the only commitment on project cost is made at project level by the project manager who "owns" the reserve (or cost buffer). At all sub-project levels a "no blame" culture for overspending should be developed to encourage people working on the project to quote realistic cost estimates without building in any reserve. If they believe that they would be blamed for cost overruns they could be expected to build in contingency reserves. This would defeat the objective of aggregating contingency reserves.

A further negative effect that is eliminated by this approach is the snowballing of costs when estimates are accumulated in a "bottom up" way to obtain an estimate for total project cost. Where managers or co-ordinators at each level within the organisation make commitments on costs (instead of estimating realistic costs without provision for contingencies) the co-ordinator at each level might be tempted to add his/her own precautionary measures on top of the estimates made by people reporting to them. In a structure with n levels, if each co-ordinator allows a 10% reserve for a specific activity, it leads to a total reserve of $(1.1)^n$ that is a contingency factor of more than 60% for this activity. This snowballing effect is obviously eliminated if only the project manager, and nobody lower down in the project structure, allows for contingency provisions.

When dealing with subcontractors it might be more practical to have each subcontractor build in his own reserve and make a firm commitment. However, the principle of aggregation would still apply and therefore TOC favours closer ties with subcontractors such as joint venture or strategic alliance agreements. The benefit of such closer collaboration would be to satisfy customer requirements at a lower total cost, thus obtaining a position of cost leadership. To obtain this the partners need to aggregate most of the reserve.

As discussed earlier in this paper, the TOC approach realises the need to optimise the product life cycle costs as opposed to project cost. The approach, therefore, emphasises a focus on project duration. Because of this focus, less emphasis would be placed on earned value analysis when the TOC approach is utilised. However, the author doubts that there could be any fundamental reason why earned value analysis should be incompatible with TOC. Where subcontractors are involved they still have to be paid for work performed (regardless whether or not the work forms part of the critical chain) and earned value measurements are still valuable.

6. Application to project risk management

Any project risk might be a constraint or could become a constraint. A mistake often made in practice is the following: project risks are identified and quantified during early stages of a project and management and engineering efforts are focused on reducing the highest risks identified. Risk events that are not initially listed under, for instance, the top three or five posing the highest risk, are neglected. This focus often results in a risk event, initially identified as not being critical, becoming the most important, but still being neglected.

If the highest risk event is identified, the focus should be on that event with the objective of eliminating the risk or, alternatively, reducing either the probability of its occurrence or its impact to a level where it would not be critical any more. The feedback loop implied by Step 5 of the TOC approach ensures that, by continually reducing the highest current risk, the overall risk is being reduced continually and systematically.

A Guide to the Project Management Body of Knowledge (PMBOK) [12] describes six elements of project risk management namely *Risk Management Planning, Risk Identification, Risk Assessment, Risk Quantification, Risk Response Planning* and *Risk Monitoring and Control*. This standard mentions that *Risk Identification* is an "iterative process" and that the iterations could be performed by different people. *Risk Monitoring and Control* includes "... identifying residual and emerging risks..." Except for indicating that risk identification is an iterative process, no explicit guidelines are given on how to prevent the problem of non-critical risks that become important at a later stage. Chapman and Ward [13] identify nine steps instead of the six identified in the PMBOK [12]. These nine steps are:

1. *define;*
2. *focus;*
3. *identify;*
4. *structure;*
5. *ownership;*
6. *estimate;*
7. *evaluate;*
8. *plan;* and
9. *manage*.

Chapman and Ward [13] explicitly indicate a feedback loop between *Manage* and *Define*. Also in the Australian/New Zealand Standard AS/NZS 4360 [14] the feedback loop to revisit risk identification is explicit. The principle of re-visiting the risk definition and risk quantification steps is therefore not new.

The TOC approach also suggests a feedback loop and goes one step further — it indicates *when* the feedback loop should be activated viz. immediately upon elevating the current constraint(s). In other words, as soon as significant progress has been made in eliminating the risk(s) on which you are focusing. A risk management model to ensure systematic reduction of risk is proposed in Fig. 2. It also ensures that risk events that are not initially listed under, for instance, the top three or five posing the highest risk will receive the required attention at the right stage.

7. Proposed further work

Due to the multi-disciplinary nature of project management it proves to be a fruitful area for diverse applications of the TOC approach. Except for illustrative

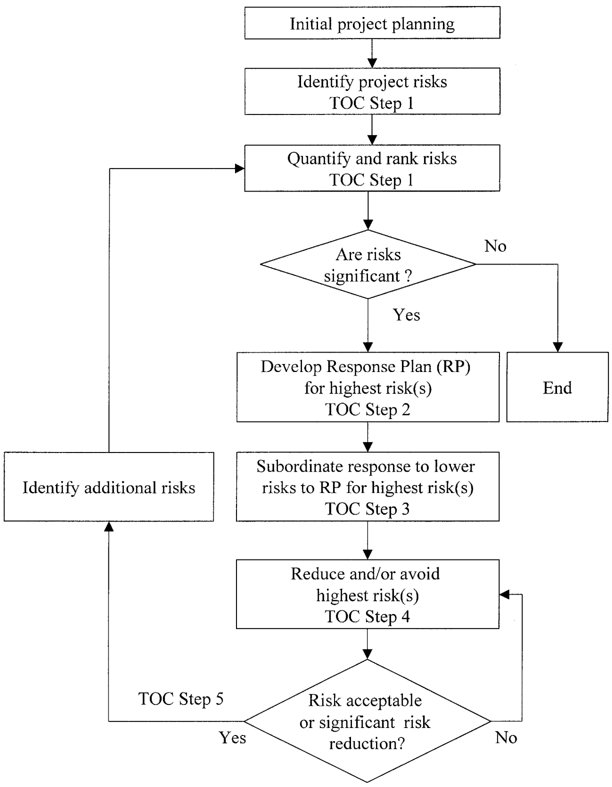

Fig. 2. Proposed risk management model for systematic risk reduction.

cases indicating that project duration could be reduced, literature still lacks information on practical results in the four areas mentioned in this paper, namely single project scheduling, multiple project scheduling, cost management and risk management. Results of cases where the TOC approach has been applied to these four areas need to be published. In addition, other areas of application should also be investigated. For example: Elton and Roe [8] correctly claim that TOC has not yet found significant application in the field of project selection. However, this does not imply that it is impossible to apply the TOC principles to this problem. The possibility of applying TOC to the problem of project selection could be investigated.

8. Conclusions

The TOC time management technique (critical chain scheduling) has been extended to allocate resources to multiple projects that share common resources. This application maximises the number of projects that an organisation can handle while maintaining the principles for reducing project duration on each individual project.

In addition to the contribution that the TOC approach makes to project scheduling it can also be applied to other areas of project management. Examples of new applications include project risk management and project cost management.

As a result of the principle of aggregation, application of the TOC approach to project cost management reduces project cost significantly.

When applied to project risk management the TOC approach provides a model that ensures systematic reduction of risk. This ensures that risks not initially identified as high receive the required attention at the right time.

References

[1] Goldratt E.M. The goal. Great Barrington, MA: The North River Press, 1st ed. 1984, 2nd ed. 1986, 2nd revised ed. 1992.
[2] Goldratt EM. Critical chain. Great Barrington, MA: The North River Press, 1997.
[3] Steyn H. An investigation into the fundamentals of critical chain project management. International Journal of Project Management, PM9996 (in press).
[4] Leach LP. Critical chain project management improves project performance. Project Management Journal 1999;30(2):39–51.
[5] Newbold RC. Project management in the fast lane — applying the theory of constraints. St. Lucie Press, 1998.
[6] British Standards Institute. BS 6079:1996 — Guide to Project Management. 1996. p 40.
[7] Turner JR. The handbook of project-based management. UK: McGraw-Hill, 1993.
[8] Elton J, Roe J. Bringing discipline to project management. Harvard Business Review 1998;March–April.
[9] Wheelwright SC, Clark KB. Creating project plans to focus product development. Harvard Business Review 1992;March–April.
[10] Avraham Y. Goldratt Institute, Introduction to the theory of constraints project management application, Version 7.3.2, New Haven CT, 1997.
[11] Turner JR. Controlling progress with planned cost or budgeted cost. International Journal of Project Management 2000;18(3):153–4.
[12] PMI Project Management Standards Program, A guide to the project management body of knowledge. Project Management Institute, Newtown Square, PA, USA (Exposure Draft 2000).
[13] Chapman C, Ward S. Project risk management — processes, techniques and insights. John Wiley and Sons, 1997.
[14] Standards Australia, AS/NZS 4360:1955. Risk management, Australia Homebush. 1995.