PERGAMON

International Journal of Project Management 19 (2000) 363–369

INTERNATIONAL JOURNAL OF PROJECT MANAGEMENT

www.elsevier.com/locate/ijproman

An investigation into the fundamentals of critical chain project scheduling

Herman Steyn *

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

Abstract

The theory of constraints (TOC) is a philosophy that is used to develop specific management techniques. The TOC technique for project time management is often referred to as the "critical chain" technique. There are indications that this technique is increasingly being used. This paper reviews the literature on this application of TOC. It also investigates the assumptions and principles underlying the approach. The technique is considered an innovation that would be useful to organisations capable of accepting a new paradigm. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Time management; Theory of constraints; Critical chain; Project schedule; Project network; Resource scheduling; Human behaviour; Buffer; Information systems; PMBOK

1. Introduction

A novel, *The Goal* by Goldratt [1], offers an introduction to the theory of constraints (TOC) approach to operations management. A decade and a half since the first print this book is still a best seller. During the period since the first publication of the novel TOC has found wide application in operations management. In 1997 Goldratt published another novel, *Critical Chain* [2], this time relating to project management.

Does TOC have the potential to contribute significantly to project management practices? The objective of this paper is to review the literature on the application of TOC to project time management and to investigate justification for the underlying assumptions.

2. A breakthrough or nothing new?

The introduction of TOC to project management met with some resistance: In a question-and-answer column Duncan [3] says that TOC borrows heavily from systems dynamics developed by Forrester in the 1950s and from statistical process control which dates back to World War II. He maintains that TOC project management presents some good ideas as new insights but that these ideas are not new. He also doubts whether it has much to offer if you are applying the PMBOK [4] concepts properly.

Drucker [5] says that a great deal of new technology is not new knowledge. Innovation is new perception. It is putting together things that no one has thought of putting together before, things that by themselves have been around a long time. The question that we should therefore answer is not whether the ideas underlying TOC Project Management are really new, but rather whether TOC Project Management puts things together in an innovative way that could be beneficial to the practitioner.

Elton and Roe [6] review *Critical Chain* [2] and admit that the approach brings discipline to project management. They also say that TOC works well when dealing with individual projects, but falls short in explaining how companies could best manage a portfolio of projects, so senior managers need to supplement the guidelines given by Goldratt [2] with other advice. Elton and Roe [6] are correct in stating that *Critical Chain* [2] does not offer much on the topic of scheduling multiple projects. However, a TOC technique for managing multiple projects exists but is beyond the scope of this paper.

Elton and Roe [6] 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. Even if TOC would never be applied to the field of project selection, it would still not mar any contribution that it makes in the areas of project time management.

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

PII: S0263-7863(00)00026-0

TOC is a management philosophy that is used to develop specific techniques. The development of such a technique for project scheduling is the main theme of Goldratt's novel [2]. The application of the philosophy to develop a technique has recently been summarised by Rand [7].

Rand [7], Barber et al. [8] and Patrick [9] all describe the application of the TOC philosophy to a single project while the application to both single and multiple projects is described by Leach [10] and Newbold [11]. It is claimed that the application of the TOC philosophy to project time management reduces project duration. The objective of this paper is to investigate this claim for a single project.

3. Project scheduling and human behaviour

Project human resource management is normally seen as a field of study quite separate from the tools and techniques of project time management. Behavioural issues in project estimation currently receive the attention of authors such as Busby and Payne [12] and certain aspects of human behaviour elaborated on in *Critical Chain* [2] were mentioned in 1995 by Holt [13]. However, the behavioural sciences and techniques such as PERT developed quite separately during the second half of the 20th century. Time management techniques therefore normally neglect human behaviour that could typically be expected during project planning and control. TOC project management, on the other hand, attempts to account for certain typical human behaviour patterns during project planning and execution.

4. Reducing project duration: assumptions and principles

Rand [7] and Barber et al. [8] go somewhat beyond the elementary principles explained in Goldratt's novel [2] but still fall short of describing and motivating the assumptions underlying the technique. Therefore, the basic principles and underlying assumptions are described and discussed below.

4.1. Assumption regarding human behaviour during project planning

The first assumption about human behaviour is that people do make considerable provision for contingencies when estimating activity duration. Advocates of TOC often use this assumption as a reason why the TOC approach should be implemented. As will be pointed out later in this paper this assumption is not critical because the TOC approach would still offer an advantage even if this assumption would not be valid. However, should the assumption be valid, it would strengthen the case for the TOC project time management technique.

Often there is little incentive to finish an activity ahead of schedule while not meeting a deadline normally reflects negatively on the individual. Rational people responsible for project activities therefore attempt to make commitments that they could meet with a high level of certainty. They are aware of the risks and even of the skewed probability distribution of activity duration assumed in stochastic PERT. Even people not consciously aware of this skewed distribution have an instinctive "feel" that the downside risks of delays are greater than the upside opportunities for early completion and could therefore be expected to build in contingency reserves that are rather liberal.

Furthermore, setting completion dates is often seen as a negotiation process. In negotiation it is common practice to make an opening bid that allows for cuts later on in the negotiation process. Should planners foresee an overall schedule cut, they could be expected to add additional reserve in order to protect their schedules from such a cut.

Also, managers at each level of the organisational hierarchy tend to add their own precautionary measures on top of the estimates of managers or co-ordinators reporting to them. Consider a structure with n levels. If the person responsible for an activity as well as the co-ordinator at each project level allows 10% for schedule contingency reserve, the provision at project level allowed for that activity would be $(1.1)^n$. In a structure with five levels, this would lead to a contingency factor of more than 60% for the activity.

That estimated activity durations do allow for contingencies is an axiom and is inherent to techniques such as PERT. Although the author is not aware of research justifying the assumption that the reserve is in practice more than necessary, this does not seem to be far-fetched. However, while many project managers might agree that people react in this way, others could certainly argue that people working on projects react in exactly the opposite way: a professional being proud of himself and his capabilities might be too optimistic when asked about task duration. Nicholas [14] says that "most people are overly optimistic and habitually underestimate the amount of time and cost it will take to do a job".

The assumption that estimates could be expected to be pessimistic is sometimes used as motivation to implement TOC. However, the main argument is not that there is in general too much contingency built into schedules but rather that it is built in at the wrong place: at activity level rather than at the project level.

4.2. Aggregation of contingency reserves

Aggregation of risk is a common principle that has been applied at least since the dawn of the insurance industry. When n owners of assets worth say $\$x$ each, do not have insurance on these assets, each one needs a contingency reserve to provide protection against the possible loss of his/her asset. Therefore a total reserve of $\$n \times x$ would obviously be required. Should they all, however, accept an offer of an agency that undertakes to insure all the assets against the risk of loss, such an agency would obviously require an aggregated reserve of less than $\$n \times x$ because the simultaneous loss of all assets would be highly unlikely. From the Central Limit Theorem it follows that, if a number of independent probability distributions are summated, the variance of the sum equals the sum of the variances of the individual distributions [15]. Therefore if n independent distributions with *equal* variance V are summated, it follows that:

$$V_\Sigma = n \cdot V$$

where V_Σ is the variance of the sum.

The standard deviation σ can be used as an indication of risk and since $\sigma^2 = V$ it follows that:

$$\sigma_\Sigma = (n)^{1/2} \times \sigma$$

where σ_Σ is the standard deviation of the sum.

Therefore:

$$\sigma_\Sigma < \mathrm{n} \times \sigma.$$

This illustrates the reduction in overall risk when risks are aggregated.

Because $(n)^{1/2}$ is significantly smaller than n, the effect of aggregation of independent risks is significant. The higher the number of risks that are being aggregated, the more marked the effect.

The TOC approach applies the principle of aggregation to project schedule risks: contingency reserves for individual activities are reduced so that activity durations are realistic but challenging. The provision for contingencies that are removed from the individual (lower-level) activity durations are replaced by a contingency reserve or "buffer" at project level. As a result of the effect of aggregation this buffer is smaller than the sum of the individual reserves that have been removed from low-level activities. Thus project duration is reduced. The higher the number of activities on the critical path of a project, the more project duration can be reduced.

In a project network, the "project buffer" resembles a dummy activity inserted just before the project end. However, unlike a dummy activity in activity-on-arrow diagramming a duration is associated with it. This is illustrated in Fig. 1 adapted from Goldratt [2].

The reduction in duration takes place even if all the initial individual reserves are realistic rather than pessimistic. Therefore, the assumption that the provision for contingencies is liberal is not a critical one. However,

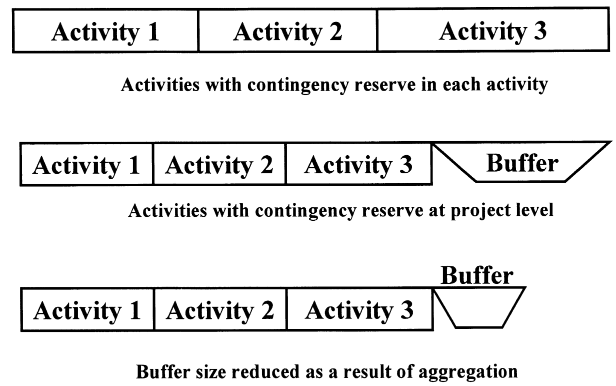


Fig. 1. Reduction of project duration as a result of aggregation.

the more provision is made for contingencies, the better the case for TOC project time management.

Aggregating project schedule contingency reserves is not common practice in traditional project management and applying the principle of aggregation could contribute significantly to the practice of project management.

4.3. *Implications of aggregating provisions for contingencies*

The most serious implication of aggregating reserves is that all due dates on individual activities and subprojects have to be eliminated. When TOC is implemented only the project manager makes a commitment on the project delivery date. At all levels below him/her workers only make estimates and communicate expectations. To encourage everybody working on projects to quote realistic durations, a corporate culture of "no blame" for not finishing on the estimated date (or for changing estimates in mid-stream), has to be developed. In other words everybody must recognise that, except for the project due date, the schedule indicates targets or expected durations rather than commitments. If they believe that they would be blamed for overdue delivery of outputs they could be expected to build in contingency reserves. This would defeat the objective of aggregating contingency reserves. Authors on project risk management such as Chapman and Ward [16] realise that one should distinguish among targets, expectations and commitments but do not apply this notion explicitly to shorten project duration by eliminating commitments at levels below project level.

A second implication is that, in the case of subcontractors, incentives for early completion of critical activities have to be negotiated.

4.4. *Assumption regarding human behaviour during project execution*

If project schedules in general allow for contingency reserves that are unrealistically high (as is often claimed by advocates of TOC Project Management), why then are so many projects not completed well within these extended schedules?

An extended project schedule could become a self-fulfilling prophecy during project execution: work could normally be expanded to fill the time available by improving the quality of the work even well beyond the requirement ("gold plating" or adding unnecessary "bells and whistles"). There is normally no incentive for a worker, e.g. a computer programmer or an engineer to complete his/her activity ahead of schedule. However, the worker would normally be held accountable for the quality of the deliverable. It is, therefore, from the viewpoint of the worker, sound risk management not to report early completion of the activity but rather to use the time "available" to reduce his/her risk of not meeting specifications. However, from the project manager's viewpoint, early completion of the project might be preferred to reduced technical risk on the specific activity.

Furthermore, individuals performing work on a project are quite often also tasked to do work on other concurrent projects. Multi-tasking or "jumping" between projects could result in a number of negative effects. The inefficiency of multi-tasking has not escaped the attention of a number of authors: Clark and Wheelwright [17], for example, believe that the optimum number of new product development projects assigned concurrently to a single engineer is two. In the question-and-answer column referred to earlier in this paper Duncan [3] calls the principle to keep resources "working on one activity [at a time] to avoid the overhead of context switching" a "good idea" that is "well documented in the project management literature". Another negative effect of multi-tasking is not mentioned by Duncan [3]: as a result of reserve built into the schedule, the pressure on the individual to commence work on the project under consideration at the planned start date is low. He/she is often tempted to work on another project instead. When contingency reserve is built into the schedule for an activity, the pessimism that prevailed during the planning stage seems to be replaced by optimism during the period following the start date. As the time available for the activity runs out, the pressure to perform the activity increases and only at a late stage does the activity receive the necessary attention. However, by then limited or no provision is available for contingencies. In this way the reserve that has been built in during the planning stage, is being "wasted" and should a risk event occur, it is likely that the activity would not be completed on time. Eliminating multi-tasking is in agreement with the TOC philosophy to focus on a constraint and subordinating non-constraints to the decisions on how to exploit the constraint.

TOC Project Management claims that it goes one step further than merely encouraging the application of the principle of avoiding multi-tasking as far as possible — it provides a tracking mechanism to tell the project manager when he/she should deviate from this rule. This is discussed in Section 7 later on in this paper.

The importance of the assumptions regarding human behaviour during project *execution* is similar to the importance of the assumptions regarding behaviour during *planning*. Should the assumptions on behaviour during project execution not be valid, it would not invalidate the technique because the principle of aggregation would still apply and project duration would still be reduced. Therefore the assumption regarding human behaviour during project execution is also not critical. However, if it would be valid, it would strengthen the case for the TOC time management technique.

5. Managing critical and non-critical activities

5.1. The concept of a critical chain

An elegant solution to take into account the limited capacity of a resource, the concept of the critical chain, was first published by Pittman [18] who gave credit for the concept to Goldratt. This concept is based on the obvious fact that the longest time required for project completion could be determined by a resource that, due to limited capacity, has to perform different activities sequentially. Although the critical path, is defined as "*... the series of activities which determines the earliest completion of the project...*" (PMBOK [4]), it is traditionally determined by precedence relationships only and resource limitations are taken care of only after the critical path has been defined. The critical chain, on the other hand, takes resource limitations into account and is composed of sections that are dependent on precedence relationships and other sections that are dependent on resource availability. The aggregated "project buffer" inserted at the end of the critical chain provides for contingencies on all critical activities.

Goldratt [2], Rand [7], Barber et al. [8], Leach [10], and Newbold [11] illustrate the concept of the critical chain. An illustration of the concept is also given in Fig. 2.

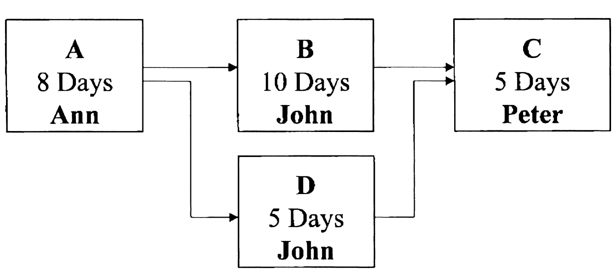
5.2. Scheduling non-critical activities

Traditionally, activities with float are sometimes scheduled to start as early as possible, seldom to start as late as possible and often, as dictated by resource availability, somewhere in between the earliest and latest start dates. Float is generally used to level the workload on resources. When the TOC technique is used, the concept of the "critical chain" already ensures that resources are not loaded beyond their capacity. While further resource levelling could be useful and is certainly preferred by functional managers, it implies an additional objective concerning non-constraints — an attempt to employ the minimum level of non-constraining resources and then keeping them busy for the duration of the project. The additional objective of keeping non-constraints busy is contrary to the TOC philosophy that states that non-constraints should be subordinated to the decisions on how to exploit the constraint(s). Therefore, if the TOC philosophy is embraced, there would be less emphasis on resource levelling.

If a non-critical activity is scheduled as late as possible there is obviously a considerable risk that it might delay the project should the activity take longer than planned. On the other hand, if an activity is scheduled to start as early as possible, changes in the scope of the activity (or changes to subsystems interfacing with the activity) might have an impact on the scope of work of the activity under consideration. This implies a risk of having to repeat the work.

The TOC Project Management approach is to schedule all non-critical activities as late as possible, but with buffers. The objective of these buffers (called "feeding buffers" because they are placed where non-critical paths feed into the critical chain) is to prevent delay of the execution of work on the critical chain when work on a non-critical path is delayed. The concept of feeding buffers is well illustrated in the literature [2,7,8,10,11] and is also shown in Fig. 3.

This approach provides advantages similar to those the "just in time" (JIT) approach offers in a production environment. In fact, it could be called "JIT-with-buffers". These advantages include flexibility in the event of unforeseen changes and improving project cash flow. Although project management techniques are used in JIT implementation [19], the author is not aware of application of the JIT principles to project scheduling before the advent of TOC.



Critical Path: A B C (23 Days)

Critical Chain: A B D C or A D B C (28 Days)

Fig. 2. The difference between the Critical Path and the Critical Chain.

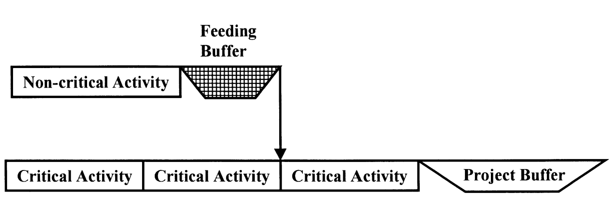


Fig. 3. Feeding buffer to prevent delay of a critical activity in the event of a delay of a non-critical activity.

5.3. Pursuing opportunities presented by good luck

Chapman and Ward [16] realise that the implications of good luck should be seized, but do not indicate how one could seize the opportunity when an activity is completed ahead of schedule.

Advocates of the TOC approach claim that in traditional project management any delay of a critical activity would delay the project, but the opportunity of one that is accelerated would generally not be seized. The *first* motivation for this claim is that a person working on a project might prefer to review his or her work rather than to report that it has been completed. This has been discussed earlier in this paper.

A *second* motivation is that, if a predecessor activity is completed earlier than planned, the resource (or resources) required to perform a successor activity might not be ready to start at the earlier date.

A *third* motivation is that the individual might have negotiated for an extended duration for the activity and might lose credit if he/she is now seen to have performed the activity in a much shorter period. He/she might feel that by reporting early completion, pressure might be invited on future estimates.

The solution to the problem of seizing opportunities to expedite a project is simple. For all *critical* activities, the responsible individual (as well as the functional manager responsible for providing this resource) receives advanced notifications of the estimated time when his/her activity should start. These notifications might typically be given a week or so before the expected start date and again, say, a day before the activity has to commence. This "early warning" or "count down" should ensure that, in the event of an activity being finished earlier than planned, the person responsible for the succeeding activity is prepared to take advantage of this opportunity. The responsible person (a) starts a critical activity as soon as he/she has received all necessary inputs, (b) works full-time on the activity and (c) passes the deliverables on to the next critical activity where the responsible person is ready to take advantage of any early completion by the predecessor. The concept of advanced warnings is referred to as "resource buffering" or "resource flagging" [10,11]. In addition to project buffers and feeding buffers, resource flagging presents a third type of buffer.

Expediting of activities on the critical chain would obviously lead to a reduction in the size of the feeding buffers. Coping with this is addressed in Section 7 below.

6. Defining the "optimal" schedule

Practitioners have to realise that, due to the unavailability of accurate data on activity durations, the development of a "perfect" project schedule is a myth. The input data into any model to simulate and optimise a schedule are mere estimates and are by no means accurate. With this uncertainty taken into account, there might be more than one scheduling solution that would be feasible and "good enough". Elaborate techniques require a large investment, especially in the form of high-level human resources. The construct that has to be optimised includes not only the schedule itself, but also the amount of effort to obtain the schedule (including contingency costs of other opportunities missed as a result of the effort). Chapman and Ward [16] phrase this idea in risk management terms: the level of investment to reduce a risk could in itself be a risk.

The advantage that the TOC scheduling technique offers is that it is simple and can be taught to project staff within a relatively short time. The approach therefore lends itself to be widely applied within organisations, provided that the paradigm shift could be made that (except for the project due date) schedules indicate targets and not commitments.

7. Control of progress

Project control is also influenced by the TOC time management technique. The risk of a delay is measured by the extent that the buffers have been consumed. Therefore, the project buffer and feeding buffers should constantly be monitored.

In traditional project management, the critical path changes as activities are completed ahead of or fall behind schedule. With the TOC technique, proper management of feeding buffers prevents the critical chain from changing during project execution. It is claimed that this provides a rigorous plan and simplifies project control.

To avoid multi-tasking, a resource is switched between activities only when a project or feeding buffer has been eroded to the extent that it poses a risk of delaying the project. In this way, the negative effect of switching between activities unnecessarily is minimised while the risks relating to delays are being monitored constantly and systematically.

8. Information systems to assist planning and control

The TOC approach requires that activities with float be scheduled "as late as possible with buffers". Commercial software such as MSProject 98^{TM} allows for non-critical activities to be scheduled to start *as early as possible* or *as late as possible* while other options (such as one initiating an activity to start on a specific date) are also common. The default option is normally "as early as possible". Resource levelling algorithms of software packages often work on the principle that activities are initially scheduled to start as early as possible and are only delayed afterwards if resource levelling is required. This cannot work when the "as late as possible" option is selected because tasks can obviously not be delayed any further with this option selected. To solve this problem new computer programmes have been developed. An example of such a programme is Prochain^{TM} (by ProChain Solutions Inc.) that is used as an add-on to MSProject 98^{TM}. Other programmes that facilitate the TOC scheduling technique include Concerto^{TM} (by Speed to Market) and Project Scheduler 8^{TM} (by Scitor). The development of these programmes indicates a market need that resulted from the increased use of the TOC approach to manage project schedule risks.

9. Suggested research

Authors such as Leach [10] quote illustrative cases where TOC scheduling has been applied and project duration reduced successfully. It is difficult to determine, for example, to what extent the TOC technique or the mere emphasis on network planning contributed to the success.

The TOC scheduling technique is based on logic reasoning and the assumptions regarding human behaviour discussed in this paper. Although the assumptions regarding human behaviour are not critical to the validity of the technique, scientifically acceptable evidence to prove the validity of these assumptions would be of value. Behaviour during planning could, for instance, be investigated by means of surveys. This could include interviews with and observations of individuals estimating activity durations and costs.

The practicality of project *execution* could also be evaluated. It has been mentioned earlier in this paper that, unlike the critical path, the critical chain does not change freely as activities are completed ahead of or fall behind schedule. Logic reasoning therefore seems to indicate that schedules generated by means of the TOC approach are rigorous and should stand up well against changes resulting from actual durations deviating from estimated ones. Although it would be difficult to make direct comparisons of the rigor of different techniques, case study research could provide evidence to substantiate the rigour of the TOC technique for project time management.

10. Conclusion

The application of the TOC principles to reduce project duration was by no means common prior to the advent of *Critical Chain* [2]. The TOC approach puts together concepts that have not been put together in the same way before and is therefore considered an innovation.

Logic reasoning indicates that the TOC scheduling technique should reduce project duration and that schedules generated in this way should be rigorous during execution.

The TOC approach changes the way we think about project scheduling. A major problem when implementing the TOC project management approach is that it implies a paradigm that is radically different from the prevailing one. It would, however, be unrealistic to expect improvement without change and change invariably leads to resistance. From proprietary course material on TOC project management [20], it is evident that the problem of resistance to change has already received much attention.

The approach has drawn the attention of companies involved in projects to the extent that software developers have launched new programmes to cater for this new need. It is not far-fetched to state that the approach has the potential to spark significant changes in project management.

References

[1] Goldratt EM. The goal. 2nd revised ed. Great Barrington (MA): The North River Press, 1992 (1st ed., 1984; 2nd ed., 1986).
[2] Goldratt EM. Critical chain. Great Barrington (MA): The North River Press, 1997.
[3] Duncan WR. Back to basics: charters, chains, and challenges. PM network, Project Management Institute, April 1999.
[4] Duncan WR, editor. A guide to the project management body of knowledge. Project Management Institute Standards Committee, 1996.
[5] Drucker PF. Innovation and entrepreneurship. UK: William Heinemann Ltd, 1985.
[6] Elton J, Roe J. Bringing discipline to project management. Harvard Business Review, March–April 1998.
[7] Rand GK. Critical chain: the theory of constraints applied to project management. International Journal of Project Management 2000;18(3):173–7.
[8] Barber P, Tomkins C, Graves A. Decentralised site management — a case study. International Journal of Project Management 1999;17(2):113–20.
[9] Patrick FS. Getting out from between Parkinson's rock and Murphy's hard place. PM network. Project Management Institute, April 1999.
[10] Leach LP. Critical chain project management improves project performance. Project Management Journal 1999;30(2):39–51.
[11] Newbold RC. Project management in the fast lane — applying the theory of constraints. St. Lucie Press, 1998.
[12] Busby JS, Payne K. Issues of organisational behaviour in effort estimation for development projects. International Journal of Project Management 1999;17(5):293–300.
[13] Holt JR. Why are our projects invariably late: how can we insure they are on time? Journal of Systems Improvement, International Society for Systems Improvement, University of Georgia, Winter/Spring, 1995.
[14] Nicholas JM. Managing business & engineering projects — concepts and implementation. NJ: Prentice Hall, 1990. p. 344.
[15] Moder JJ, Phillips CR. Project management with CPM and PERT. 2nd ed. New York: Van Nostrand Reinhold, 1970. p. 279.
[16] Chapman C, Ward S. Project risk management — processes, techniques and insights. John Wiley and Sons, 1997.
[17] Clark KB, Wheelwright SC. Managing new product development — text and cases. New York: The Free Press, 1993. p. 242
[18] Pittman PH. The theory of constraints applied to a single project environment. Journal of Systems Improvement, International Society for Systems Improvement, University of Georgia, Winter/Spring, 1995.
[19] Gelinas R. The just-in-time implementation project. International Journal of Project Management 1999;17(3):171–9.
[20] Avraham Y. Introduction to the theory of constraints project management application. Version 7.3.2. New Haven (CT): Goldratt Institute, 1997.

Herman Steyn, MBA, Ph.D., in his capacity as professor in the Department of Engineering and Technology Management at the University of Pretoria, is responsible for post-graduate education and research in Project Management. Prior to his appointment at the University in 1996, he held senior positions and acquired experience of more than two decades in project environments in more than one industry. He also consults to industry on project management and is a licensed instructor of the Avraham Y. Goldratt Institute on proprietary "Critical Chain" courses.