Beyond *rounding up the usual suspects*: Towards effective quality management policies for production ramp-ups in high-tech supply chains

Henk Akkermans

ha@uvt.nl
Tilburg University
Economics and Business Administration,
Dept. of Information Systems and Management
PO Box 90153, 5000 LE Tilburg, The Netherlands

Abstract

This study investigates the issue of managing quality during production ramp-ups in high-tech supply chains. It combines an in-depth case study of one particular high-tech supply chain setting with insights from the recently-emerging literature on behavioral operations and synthesizes these two into a system dynamics simulation model.

Model analysis suggests that isolated and intuitively appealing quality management policies are likely to lead to suboptimal or even detrimental results. Of crucial importance is finding the balance between ramping up production rates sufficiently fast to capture short-lived market demand and avoiding to increasing production starts so high that workload levels in the supply chain move beyond the tipping point. This means that when workloads become too high, the entire supply chain can get bogged down in a vicious cycle of high workloads leading to low quality levels, which lead to high rework levels and hence to even higher workloads.

Especially promising policies to be used in combination are, firstly, moderate production ramp-up rates that turn out to generate more timely output than overly aggressive production ramp-ups. Secondly, policies that leverage the expertise that can be gained from analyzing defective units that cannot be repaired easily downstream, as these may yield especially valuable knowledge regarding hidden quality issues in upstream manufacturing operations. Other measures such as installing quality gates or superior error detection equipment are found to be ineffective or even detrimental for improving supply chain performance.

1. Introduction

In today's "high-clockspeed" industries (Fine 1998), such as the electronics industry, ever-shrinking product-life cycles lead tot the need for ever-shorter production ramp-ups. The shorter and steeper the production ramp-up, the greater the opportunity to capture the short-lived market demand while it is still there. The earlier your product is on the market, the less it will suffer from price erosion. Because of these short-life cycles, neither product designs nor manufacturing methods and tools are typically fully developed when the first production runs start. One thing is worse than not having your product out there early: having a defective product out early. Therefore, effective quality management during production ramp-up has become of paramount importance.

The current omnipresence of decentralized supply chains has complicated this challenge of quality management during ramp up further. This is not just because different stages of production may now take place on different continents, or at least in different companies. Decentralized supply chains also pose a specific challenge for quality management, which is the following. Typically, in high-tech electronics assembly, quite some of the errors that are generated upstream in the supply chain can only be fully detected downstream. This has two main reasons. Firstly, upstream one can only test the components, but further down the production process these components can be damaged during handling or transport, for instance when other components are added to the semi-finished product. Secondly, and this is most important, many of the errors that are generated upstream can only be detected when the product is fully assembled, as they only become visible through the interaction of the components with one another. And this makes it inherently impossible to detect such errors upstream. So, errors that are generated upstream can only be found downstream, but then the root cause of these errors has to be traced back to upstream operations. Obviously, in a fast-moving industry, such communication and joint problem-solving can be very problematic indeed.

Unfortunately, there is very little sound and well-tested guidance on how to proceed in this matter. There is a well-developed literature on quality management during product development processes, and in particular in IT system design, but it is unclear how this might translate into prescriptions for production during ramp-up. On the other hand, there is a wealth of practitioner literature on quality management in production, especially from the quality management/TQM/Kaizen/lean manufacturing movement, but here the implicit assumption is almost always a steady-state of production rates, whereas during production ramp-up there is never such a steady state.

So, what do companies do? Typically, as Captain Renault originally ordered his policemen to do in the movie classic Casablanca, they order to "round up the usual suspects" on the basis of common sense, implement such policies as well as they can and hope for the best. Of course, this leaves much to be desired. It is usually unclear to management who of these usual suspects are really guilty, e.g., what measures will work best and where and when. It is certainly not clear to what extent they reinforce each other or if they may even reduce each other's effectiveness. Unlike the infamous Captain, management does not go for the obvious measures because they don't care what the outcome will be, but rather because they cannot resort to a body of theory that goes beyond these obvious candidate measures.

This paper describes a study that employs simulation analysis to explore the beginnings of such a theory, to start moving beyond "rounding up the usual suspects". This study introduces and analyses a simulation model of production ramp-up in a high-tech supply chain on the basis of a field study in 1995 of a personal computer manufacturing company with a two-stage internal supply chain and external component suppliers. In this original field study, it appeared that what management wanted to do to improve quality during production-ramp up was often not effective. The present study takes this original setting as a start, but employs a simplified rendering of the more detailed simulation model that was developed in 1995 to guide the analysis back then. It finds that, with the present model, the fundamental efficacy issues with many of the obvious candidate policy measures persist.

This study presents policy analyses with the simulation model of different quality improvement measures. Some of these turn out to be effective, others turn out to have marginal effects or even downright negative effects on quality. Some of these outcomes are in line with managerial common sense. Most of them are not.

A first policy experiment was to explore the benefits of a less steep ramp-up production scenario. This was declared out of bounds by management from the start, given the marketing considerations sketched before. However, what management did not expect was that a less aggressive ramp-up scenario would lead to more product being available on the market *sooner*, rather than later.

A policy experiment whose outcomes did fit with managerial intuition was that high component quality would have a beneficial effect on both quantity and quality of factory output. What management also did expect was that the ability to maintain consistently high quality in upstream or downstream manufacturing operations would be beneficial. What management could not foresee was that consistently high upstream manufacturing quality yielded only marginal improvements in the analysis, whereas consistently high downstream manufacturing quality would lead to excellent results.

Another scenario that showed only marginal improvements was to install equipment with superior error detection quality. This was all the more frustrating since management was considering making such investments at the time. Significant contributions were expected by management from improving the quality and quantity from feedback from downstream repair work to upstream manufacturing. These were confirmed and could be implemented against significantly lower cost.

One policy that sounded especially attractive to management was the notion of a quality gate or stage gate. In the context of manufacturing, this meant that units could not move on to downstream production until a certain quality level had been reached in upstream production. Policy analysis suggested that setting a high quality gate would have disastrous results on supply chain productivity and output. Setting a modest quality gate might have some benefits.

Finally, this study looks at several combined policies, especially for those measures that would not require major restructurings in operations. Here a combination of a less aggressive ramp-up target in combination with better feedback from and more resources in downstream repair work appears especially promising.

In the concluding parts of this paper, the proposition is put forward that what management may really need is not a list of setting-independent usual suspects for improvement, but a deeper understanding of what really drives quality during production ram-up. Here a link is made to existing theories from behavioral psychology and operations management that focus on the important feedback effects between schedule pressures, workload and error rates. The higher the schedule pressure, the higher the workload (Oliva and Sterman 2001, Akkermans and Vos 2003). When workloads pass a certain threshold, process quality will drop so error rates will rise (an insight from psychology that goes back to Yerkes & Dobson 1908).

This leads to more rework, which further increases workload and limits output so that schedule pressure increases as well. An effective quality management policy during ramp-up ensures that the supply chain does not get trapped into such vicious cycles of every-greater workload and ever-lower quality and output rates.

2. Literature review

2.1. Production-ramp up as a "missing link"

There is relatively little academic literature that deals specifically with the issue of quality management during production ramp-ups and this is quite understandable. After all, also in practice production ramp-ups form something of a "missing link" between product-development and volume production. This is an area that is not well understood and that often leads to major problems in timely and cost-effective market introductions. (c.f. Mass & Berkson 1995, Akkermans 2001). So it should come as no surprise that there is no adequate body of theory to fall back upon either.

Why is quality management such an issue during production-ramp up in high-tech supply chains? Because quality is never at market standards when a product design is released by product development into production. On the one hand, this is an issue of industry clockspeed (Fine 1998). There is a rush to get to the market first and, therefore, also a rush to get into production ASAP (Iansiti & MacCormack 1997). On the other hand, design imperfections are inherent to the nature of high-tech products: it is practically impossible to figure out precisely how a design will turn out in terms of quality from the drawing board or CAD screen. For this, specific production expertise is need and also a lot of testing, so, production, and not just of a single prototype, but of actually some volumes.

What theory is then available? Definitely, the literature of the two functional areas that production ramp-up falls in between: product development and volume production. These are two strong and well-developed fields of expertise, both with their well-established sets of ideas on how to manage quality. We will select some of the key publications in these very broad fields to investigate what they have to say for this area.

Given the dominance of product development and high-volume manufacutring it is not surprising that most of the ideas that are tried out in practice by managers to deal with production ramp-up issues stem from either one of these areas. Unfortunately, these fields often contradict each other for production ramp-up policies. For instance, Boehm & Turner (2005) note that one barrier to the introduction of agile processes is the conflicts in development process that arise when agile processes are merged with standard industrial processes. And, at any rate, as Mass & Berkson noted in 1995, companies that use "the theme of the month" to get process improvements in this area rarely see much benefit.

So, we need more. Here comes into play a third source of literature that can be applied to this specific area. This is the relatively recent modeling literature on the interplay between behavioral psychology and operations management elements in how individuals and organizations deal with stress and market pressures. We will look at this literature as well, also because the simulation model that is introduced in Section 3 has some of the relationships suggested in this literature at its core.

Finally, there are some publications that deal specifically with the issue of quality management during production ramp-up in supply chains, although their insights are not easy to access, interpret and use in follow-up studies We will look at those as well.

2.2. Insights from lean production

What we call today "lean" or "six-sigma" or "lean-six-sigma" was started in the 1970s in Japan, notably Toyota. A central element in this approach to production management has always been quality, rather than quantity or speed (Ohno 1988, p.107). These ideas regarding quality come from some U.S. quality experts some decades earlier, notably Deming, Juran and Feigenbaum. The latter wrote down the core concept back in 1961: "The theme…is 'make them right the first time.' Emphasis is on defect prevention so that routine inspection will not be needed to as large an extent." (Feigenbaum 1961, quoted in Schonberger 1982, p.47.).

Originally, these ideas were targeted to improving ongoing production, not production during ramp-up. That is not to say that the Total Quality Management (TQM) concept or lean has not been applied to product development and production ramp-up, see e.g. Kennedy 2003).

What doing-it-right the first-time means for a production line was vividly described by Womack et al. in their study of Toyota assembly plants: "each worker along the line can pull a cord just above the work station to stop the line if any problem is found (…) but the line is almost never stopped, because problems are solved in advance (..) (Womack et al. 1990, p.80). And yet, "we observed almost no rework areas at all. (…) there were practically no buffers…and there were no parts warehouses at all." (p.80). So, very strict quality controls are assumed to lead to very low reject levels and hence to very low inventory levels, which makes it again easier to keep quality levels very high.

Lean manufacturing suggests a variety of techniques to achieve such consistently high quality levels. Quality circles are among them and also so-called "poka-yoke methods" or failsafe methods (c.f. Schonberger 1984, Hall 1987). The first techniques certainly assumes that quality is not perfect, but that continuous improvement is a natural state of affairs, the second one does assume that perfect quality is attainable, simply because it is possible to develop tools that make the occurance of error impossible. Both techniques, and most others, do assume a development over time. But neither of them treat the issue of production ramp-up as problematic in its own, which remains understandable from the roots of lean in the textile and automotive industry, where product-life cycles are usually fairly long and the production ramp-up period is small relative to the period of stable volume production. Similar comments can be made about the Six Sigma practitioner movement as well. This approach to lean, made famous by its adoption by GE, has as one of its One of the chief problems that it is narrowly designed to fix an existing process, not an inherently new setting (c.f. Morris 2006).

In summary: quality management is really fundamental to lean production operations, but historically the phase of production ramp-up has not been the focus of this literature. Its implicit assumption remains one of stable volume production, and quality improvement recommendations that fit an environment where design and operations quality are fairly high already. Managers who apply recommendations from the lean practitioner literature to production ramp-ups settings too lightheartedly will do so at their own peril.

2.3. Insights from agile product development

In research in product development, software product development is an important area of attention, for several reasons. However, management insights gained in the development of software tend to find their way into other, less extremely innovation-driven industries, for instance the car industry (Ward et al. 1995, Iansiti & McKormack 1997)

In software development, the issue of how to get to high levels of quality fast, very fast, has long been a crucial managerial concern, given the high degrees of innovation and shot life cycles of many software applications. Broadly speaking, there are two generic kinds of development models: sequential or waterfall models and what Molokken-Ostvold & Jorgensen (2005) call “flexible” development models.

- *Waterfall models* represent the “classical” (c.f. Brooke 1979) mode of software development. This means sequential processing of the generic phases: analysis, design, programming, testing. These are most suitable when technology, product features, and competitive conditions are predictable or evolve slowly [20].
- *Flexible development models* can be “incremental”, “evolutionary” or “agile”, but the bottom-line is always that several parts of the system are being developed in parallel, in small increments, and in close cooperation and with close communication between client and developers (Molokken-Ostvold & Jorgensen 2005). Interestingly, this is also where “lean” product development models should be located. Here one talks about set-based concurrent engineering, with many concepts being developed concurrently for each subsystem, that are evaluated against each other, so that the weak concepts can be eliminated, and insights from each other them can be combined in different ways in the final product (Kennedy 2003, p.122).

This distinction in product development is certainly not specific to software. In 1984, Takeuchi & Nonaka (1984) described these two different concepts of “the product development game” as the difference between playing a relay race or rugby on the basis of their work with high-tech electronics and automotive companies such as Fuji, Cano and Honda.

An important concept in product development is the notion of the *stage-gate*. A stage-gate model describes a work process from idea to delivered product. The model presents the stages of product development in a generic abstract project life cycle suitable for high-level management. Two stages are connected through a gate, and at every gate management can decide to stop the project or move to the next stage. This is at the managerial level. Importantly, such stage-gates can be applied in both types of development models, but with a crucial underlying difference, as Karlström & Runeson (2005) note. They state that at the level below the managerial decision-making process, at the technical level such a stage-gate model can relate to a waterfall development process or an agile development process. In the waterfall model, functionality is fixed: the project cannot move to the next stage until a certain level of completeness and quality has been reached here. So, time becomes flexible, and this is why waterfall models are especially associated with large project schedule overruns. In agile methods, time is fixed (“time-boxing” as it is also called). So, functionality and quality become variable.

In summary, the field of agile product development contains numerous suggestions on how quality improvement can be speeded up in a volatile setting. Parallel execution of tasks, intensive communication and feedback between stages are among the most directly relevant ones. However, we should not forget that in production ramp-ups, we ware NOT talking about designs, but about actual units of product. And when these units are defective, they cost money when they are not shipped to the customer and may cost even more money when they are shipped. Again, managers who apply recommendations from the agile product development literature to production ramp-ups settings too lightheartedly will do so at their own peril.

2.4. Insights from "behavioral operations"

Quality management during production ramp-up is clearly a complex phenomenon, with multiple interactions between different production stages, and multiple interactions between human behavior and cognition and the operations they influence. Although this phenomenon itself has not been investigated extensively from this perspective, there is a nascent literature that acknowledge this inherent complexity and takes both behavioral and operational considerations into account. Gino & Pisano (2006) label this emerging body of research as "behavioral operations", and they define it as "the study of attributes of human behavior and cognition that impact the design, management, and improvement of operating systems, and of the interaction between such attributes and operating systems and processes" (Gino & Pisano 2006, p.9.). Earlier on, Boudreau at al. (2003) also suggest a research agenda for the interface between operations management and human resource management, as "human responses to OM systems often explain variations and anomalies that would otherwise be treated as randomness or error variance in traditional operations research models".

In the context of the present article, what are important findings from this work in behavioral operations? Interestingly, of these use system dynamics methods and models, just like the present study, and most if not all of them trace back to John Sterman's seminal publication in weaving together operations and behavior from a dynamics perspective on how management decisions are made in the Beer Game (Sterman 1989). From this population, different studies present different key messages. Within the context of the present study, the following ones appear most relevant:

1. *Rising workloads lead to more errors*. Firstly, there is the notion that when workloads rise beyond a certain level, productivity will drops and error rates will increase. This finding has a long history in cognitive psychology, and goes back to neuropsychological research from the beginning of the 20th century (Yerkes & Dodson 1908). "This Yerkes–Dodson curve—originally derived from experiments testing how rats running through a maze responded to electric shocks of varying size—depicts a curvilinear relationship between stress and performance. Initially, stress improves performance, but, as it continues to increase, performance eventually declines". (Repenning 2003, p.312). This "law for all seasons" (Eigen 1994) has found several applications in system dynamics oriented research in behavioral operations. For instance, Rudoph & Repenning (2002) show how this "law" can explain the dynamics of disasters such as plane crashes and their organizational equivalents. Akkermans & Vos (2003) use this notion to explain why sales campaigns can have disastrous effects on the service supply chain when too many orders are pushed downstream.

2. *Higher schedule pressure leads to more errors*. One special manifestation of this relation between workload and error rate is that it is not really just the amount of work one has on one's plate, but also the awareness of how much really *should* be

done to get back in line with the company target for output. This negative effect of schedule pressure on the error rate has been empirically established by Oliva & Sterman (2001) in the service industry and found in the context of product development in high-tech electronics by Mass & Berkson (1995).

3. *Work pressure reduces detection rates*. One other indirect consequence of the Yerkes-Dobson law is that increasing work pressure leads to lower levels of error detection. One reason why is 1st order and obvious: if you are too busy, errors slip through. Ford & Sterman (2003) also point at another, more 2nd order effect: a collective urge amongst project mangers to hide rework that they know will have to be done sooner and later in order not to be the bringer of bad news. Hence, management forms a collective "Liars club", as Ford and Sterman put it.

4. *Error prevention is less rewarding than error solving*. This is also one of the reasons why a fire-fighting culture is so persistent in product development and production settings. As Repenning & Sterman (2001) point out, "nobody gets credit for solving problems that never happened." So, companies will have a persistent preference for investing in becoming better in problem solving than in error-prevention.

5. *Errors and workloads cascade down in multi-stage processes*. This particular finding is all the more detrimental for performance in a multi-stage setting such a supply chain because both errors and workloads will tend to cascade down in them. For instance, Repenning et al. (2001) found that persistent firefighting in NPD resulted from a positive feedback process whereby lack of attention to the early phases of the development process results in serious problems when projects reach their downstream phases. Those problems, in turn, consume scarce engineering resources that would otherwise be dedicated to the early stages of the next generation of products.

Ford & Sterman (1998) showed this in an earlier study of the dynamics of the product development process. They emphasized that, in product development, changes inherited by downstream phases from upstream phases corrupt downstream work, which must then be corrected. When inherited changes are discovered by a downstream phase they are returned for correction or improvement to the phase where the change was generated. Akkermans & Vos (2003) analyzed how this can lead to a cascading of workload and errors in a service supply chain. In a service supply chain context, this cascading of workloads was also investigated by Anderson & Morrice (2001) and Anderson et al. (2005).

6. *Effective (downstream) repairs lead to more (upstream) expertise*. The notion of the learning curve is well established in operations management and on example of a behavioral phenomenon that has become part and parcel of OM theory and practice. What the behavioral operations literature adds to this is specific notions of how and where people learn from operations and where and how this knowledge is best applied. Oliva & Sterman (2001) show how service staff increases its productivity through learning by doing in service encounters. In the context of processing legal cases, Akkermans & van Oorschot (2005) modeled how experienced staff would become more adept in accessing cases upstream in their service supply chain so that their subsequent processing might become easier.

2.5. Examples of an integrated behavioral operations approach to production ramp-ups

There are some examples of just the type of study that is presented in this article. Two of them are consulting cases. Mass & Berkson (1995) describe insights from what in fact was the precursor to the original field study that forms the basis for the current research. Their work is focused on managerial insights regarding the unintended side effects of setting overly aggressive commercialization targets for product development projects in high-tech electronics. Several of their findings will indeed resonate back in the current study. However, the model they base their findings upon is not in the public domain. Akkermans (2001) describes qualitative insights from a study in the high-tech electronics sector on issues related to premature release of designs into production driven by strong schedule pressures. Again, there is no description of the model employed here available, although the insights are in line with what the current study finds.

Anderson & Joglekar (1996) wrote a conference paper with a promising title on modeling the dynamics of technological ramp-ups, but this really focuses on the relation between process innovation and product innovation, not on quality management issues. Perhaps the most promising of all is Juergen & Milling (2006), albeit that this is presently only a conference paper without model documentation. This paper describes insights from production ramp-ups in the automotive sector, where the authors find that there is a trade-off between starting production early on, to maximize learning and early feedback into design, and waiting until production designs have accomplished a certain level of series production readiness. In the future, this paper may shed light on how effective production ramp-up policies may be different in a medium industry clockspeed setting such as the automotive sector.

3. Research approach and environment

This study is very much anchored in the real world. Its origins lie with a consulting study conducted by the author in 1995 for a major U.S. high-tech electronics firm. For the present article, a new and simplified simulation model of the field setting that was encountered there has been developed without losing the essential characteristics of the model. In this section, we first explain the research methodology. Then we introduce the empirical supply chain setting, the original management concerns and explain how the original field study was conducted. This section concludes with a brief explanation of the modeling and analysis that has taken place 11 years later, for the research that is presented in the current paper.

3.1. Research approach: Synthesizing research site findings with behavioral operations theory in a simulation model

In its methodology, the research reported in this article is similar to published work by Akkermans & Vos (2003) in *Production and Operations Management* and by Holweg et. al. (2005) in the *Journal of Operations Management.* Partly, because in all three papers a system dynamics model is presented that is based on a synthesis of field work, in the case of Akkermans & Vos a consulting study from the telecom industry, in the case of Holweg et al., an automotive research project sponsored by industry, and in the case of the current paper a consulting project with a high-tech electronics company.

Another source of similarity is that in all three papers this field work is used as input for follow-up analysis and synthesis, which generates a simulation model and insights that go beyond the findings from the original field studies. A final aspect of similarity is that the simulation model that is presented here was elicited through the use of causal loop diagrams and process mapping (Evans 1998, Sterman 2000).

One notable difference with these two earlier articles is that the present study deals with less “tangible” issues than amplification effects in a telecom service supply chain or the physical automotive supply chain and scheduling policies. As a result, the present paper needs to make a better case for its assertion that the “soft” causal links that were derived through group knowledge elicitation (Vennix 1996) in its field study are plausible and valid ones. Back in 1995, this would have been very difficult. Fortunately, the last decade has seen a number of publications that address similar issues, which we have labeled in the previous section as “behavioral operations”.

3.2. The research site

The client company in this case study was a major U.S. electronics manufacturer that had been very successful, enjoying considerable increases in market share in a market that was strongly growing. In the 5 years prior to this field study, its revenues growth was about 40%/year. The quality image of most of its products was also above industry average. Like all other firms in this industry, it was faced with shortening product life cycles, price pressure, very steep production ramp-up curves and a never-ending need to add new technological improvements to subsequent designs. On top of this, success was taking its toll as well. As a result of this fast growth, there were many problems with meeting delivery dates, getting adequate levels of quality and coordinating the mushroomed set of production units, systems and processes.

Prior to the field study the current research focuses on, the late Nathaniel Mass, then working with McKinsey & Co., had conducted two system dynamics modeling projects for this company to meet these growth challenges. Selected insights from both have become available in the public domain. One focused on the external supply chain with distribution channels. This case was written up by John Sterman in Sterman (2000), p. 743-755. The other study focused on product development issues and selected insights from this case were published in Mass & Berkson (1995).

Both studies had been extremely successful, and it was only natural that the area in between the two business domains covered so far would be investigated in a similar manner: the dynamics of the internal production supply chain itself. Together with Nat Mass, the author of the current paper conducted this system dynamics modeling study. In order to keep up with the tremendous demands that its highly dynamic industry in combination with the company’s fast growth placed on production quality, management was had put extensive quality assurance measures in place and was considering additional ones. In order to evaluate the effectiveness of the existing measures and the proposed additions, company management commissioned an exploratory system dynamics study that would focus on its production supply chain for personal computers (PCs), which is visualized in Figure 1.

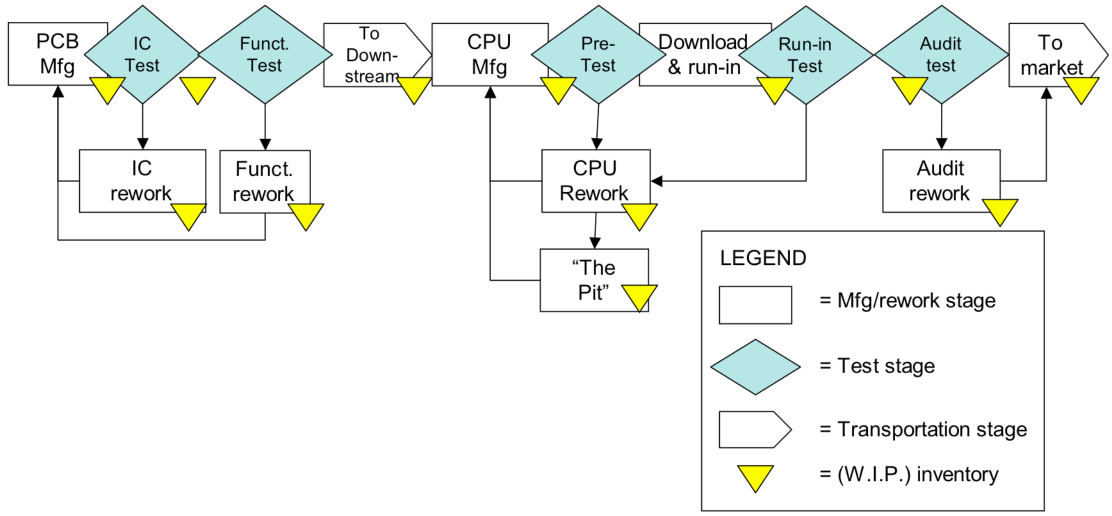

Figure 1: Schematic overview of process stages in the actual supply chain

Broadly speaking, this supply chain consisted of two main stages.

- In the first stage, "upstream manufacturing", a semi-finished product was created, a Printed Circuit Board (PCB) with mounted inter which was tested in two separate test stations, one for the Integrated Circuits or ICs themselves and the other for the functions they should collectively perform. Products that failed these tests would be repaired and fed back into the production stream
- In the second stage, "downstream manufacturing", which could be on a production line next door or on the other side of the globe, additional components were added to this semi-finished product in two main steps, interceded with various tests and audits, leading to three types of rework for defective units. Of particular concern is the rework process stage called internally "the Pit", because this is the place where all the defective units that could not be repaired immediately would pile up.

3.3. Management concerns in the field study

At the time, company management had already set in motion plans for improving quality during production ramp-up and was considering to set up more. However, it could only select these on basis of common sense. It could only "round up the usual suspects" and hope that the really effective ones were among those. Since it felt uncomfortable about this and had already experienced spectacular success with developing simulation models of its product development processes, it commissioned a study to investigate the effectiveness of different quality improvement measures. Table 1 summarizes managerial thinking in this area at the time. To some of the measures that were being considered management was fairly skeptical, to some simply opposed. An example of the former is the idea of putting experienced staff into the Pit rework area to generate production insights from their repair work on defective downstream units. At the time, nobody really knew for certain what that might give in production. An example of the latter is a less steep production ramp-up schedule. Even discussing this was fairly controversial at the company, since its success so far had been based on its ability to move faster than others in very fast-moving markets.

Quality Measure under consideration at the time	*Ex ante managerial inclination*	*Ex ante underlying reasoning*
Schedule less steep production ramp-up	Not to be pursued, unacceptable for Marketing	Cannot allow delays in getting product to the market
Ensure high component quality	Pursue, but may be difficult to implement	Better quality in means better quality out
Introduce fail-safe equipment and production methods	Positive, but very difficult to implement so more something for the long term	Doing-it-right-the-first time
Install superior error detection equipment	Definitely pursue	No defective materials should be allowed to move downstream
Increase pit repair work staff numbers and skill levels	Pursue possibly, or forget about pit stock, which is best	On the one hand: repairing these units may give insights into upstream quality issues. On the other: this is job that requires very skilled staff, who we can't really afford to miss elsewhere
Set up a quality gate before downstream production	Pursue, but unsure about location and level	No defective materials should be allowed to move downstream

Table 1: Quality measures under consideration with management in field setting

Other measures simply seemed sensible ideas, were really "the usual suspects" for any company that wants to work on improving its product quality. Fail-safe equipment, high component quality, superior error detection tools and quality gates all fall in this category. Today, one would call these measures part of the toolbox of "lean" manufacturing. Then, the label was "world class" or Total Quality Management or TQM, but the underlying reasoning was the same: no defective materials should go into the supply chain, and when defects would occur they should immediately be detected and corrected, not passed on further downstream.

3.4. Data collection and analysis at the research site

Data collection at the research site was conducted "from the ground up" (Glaser & Strauss 1967, Miles & Huberman 1984), and through a variety of techniques. First, a half-day brainstorm session was conducted by Nat Mass and the author with a small number of key experts and managers in the company. In this so-called Group Model-Building (Vennix 1996) brainstorm, the question was addressed: *"When something goes wrong in production, why is that and what happens as a result*?" This discussion was captured in causal loop diagrams (CLDs), in which the cause-and-effect relations of the relevant variables were visualized, as well as the feedback loops connecting them. Causal loop diagramming is a common tool utilized in system dynamics modeling to define flows, feedback loops and system structures (c.f. Vennix 1996, Sterman 2000). Causal maps are also used to synthesize findings in qualitative research (c.f. Miles & Huberman 1984). In this session, a number of key causal relationships were identified between outgoing product quality and various other characteristics of the production system, such as work pressure, operator skills, design fails or inspection quality.

On the basis of the results from this brainstorm session, the author developed — in interaction with company staff, Nat Mass and other consultants — a comprehensive system dynamics model of the various interactions in production operations. Key data were collected to parameterize and calibrate the simulation model. Then, a broad range of quality management measures was simulated to evaluate their effectiveness on a stand-alone basis and in combination. All this was done in a period of almost two months. The results of these experiments were presented in a second workshop for the same group.

The reception of the outcomes was not positive. As the scenario analysis in Section 5 will illustrate, managerial intuition was contradicted in many instances by the policy analyses. In fact, the messages that came out were so contradicted that the entire modeling study was dismissed on the grounds of incorrect formulation and technical errors in the model. Certainly, no immediate implementation of its recommendations was planned. And so, this author was left with the question: was the model just “wrong” or was the way it was presented not successful in convincing company management of the validity of its results?

3.5. Modeling and analysis one decade later

Over the course of the ten years that followed, there gradually became available studies that presented similar findings as the controversial outcomes of this original field study. These have been discussed in Section 2. This fed this author’s belief that, although the relationships that had been developed “from the ground up” (Strauss ref) with a small group of company representatives from one company in one moment in time were actually fairly generalizable to a broader range of settings.

I therefore developed a new model, based upon the original detailed simulation model, but this time stripped from parts that seemed unlikely to be changing its overall behavior (e.g., not two subsequent test units but just one combined unit, no separate audit rework site). I also made sure that the behavioral relationships contained in this model were consistent with the model formulations regarding the same relationships that had become available in the academic literature since then.

I compared the dynamic behavior of the new model broadly consistent with the original one, certainly in its management implications. I also conducted the “usual” test and experiments to assess the sensitivity of model behavior to changes in its key assumptions. I then revisited the original managerial issues, the “ususal suspects” that had been rounded up more than one decade earlier on, and used to model to investigate their “guilt” for this more generic model. The main findings of this modeling and analysis are presented in the subsequent sections.

4. Simulating production ramp-up dynamics in a high-tech supply chain

In this section, we present a formal model that integrates the characteristics of the dynamics of quality management during production-ramp ups. This model can capture the “cascading sequence of causes and effects that determine how well a company delivers against the key performance metrics of speed, quality, and cost” (Mass & Berkson 1995, p.20-21). In this section, we will limit our discussion of the model to the main goods flow structure and the key feedback loops that determine its behavior over time.

Technically speaking, this model is a complex interrelated set of non-linear differential equations. Because of space considerations, these are available from the author in a separate document, both in text format as in simulation software format. For ease of explanation, we will mostly use diagrams here that visualize the main interrelations between variables. In Figure 2, a stock and flow diagram notation is used (Sterman 2000), in which accumulations or state variables are symbolized by rectangles, so-called “levels” or “stocks”, and rate-of-change or “flow” variables are signified in the diagram as “pipes” with “valves.” In Figure 3, a causal loop diagram notation is used, in which only the causal links and loops are shown between variables, without distinguishing between different kinds of variables.

4.1. Main goods flow structure

The goods flow structure for the simulation model is shown in the stock and flow in Figure 2. This is a simplified version of the supply chain structure shown in Figure 1. In the case of a physical goods flow, the analogy with physical levels for the rectangles denoting stock levels, and with the pipes with valves connecting them, is obvious, but still the underlying mathematics are those of differential equations.

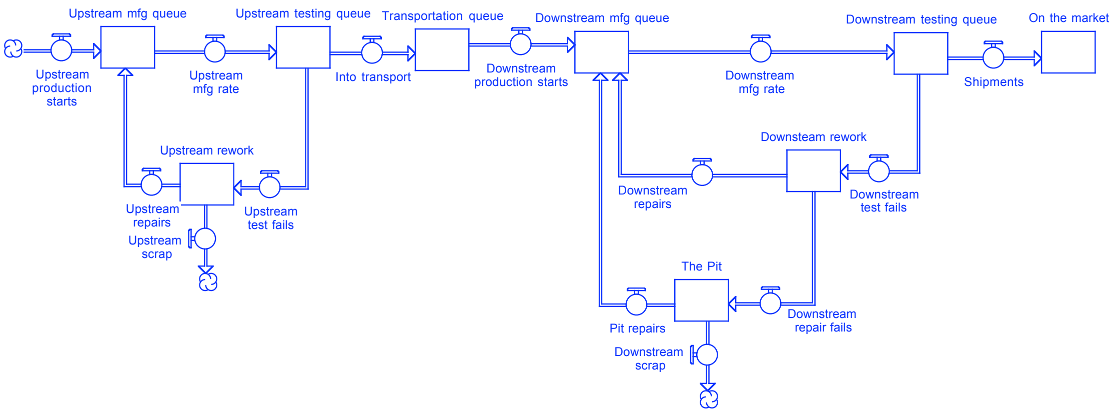

Figure 2: Overview of simplified supply chain structure in the simulation model

In the goods flow that is represented in the simulation model we will investigate, there still is an upstream and downstream part of the supply chain, with a manufacturing, a testing and a rework stage in both parts. In the downstream part of the chain, “the Pit” reappears, which is a second-level of rework, for those units that could not be repaired in the first-level rework stage. Components flowing in are not visualized, neither is shown what happens to scarp. The clouds at the beginning and end of valves denote system boundaries.

4.2. Main feedback control loops that drive dynamic behavior

A cornerstone of system dynamics modeling is the representation of feedback. During the group model-building workshop that was conducted in the original field study, four key feedback loops were identified that were felt to affect production performance considerably. All are shown in Figure 3. There are two reinforcing feedback loops identified in it, denoted R1 and R2, and two balancing feedback loops, labeled B1 and B2. We will discuss these in this order.

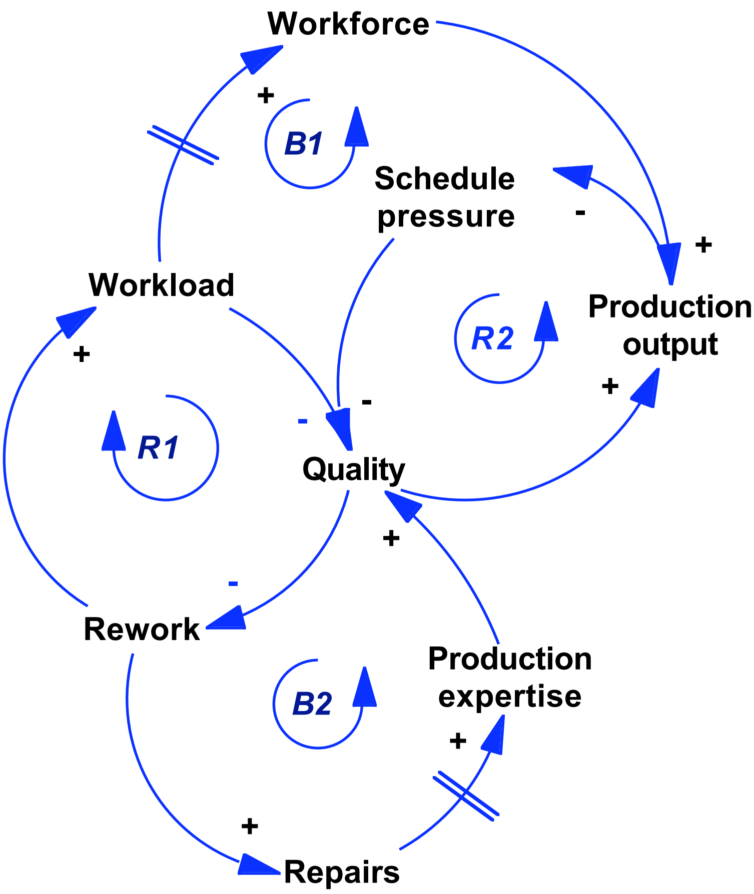

Figure 3: Key feedback loops in the model

R1: *The accumulating workload loop*: This reinforcing feedback loop operationalizes the following line of reasoning. People tend to make more mistakes when there is insufficient time to do a good job, i.e. when their workload is high. Unfortunately though, these mistakes will lead to an accumulation of defective units which, over time, will have to repaired. Repairs tend to require more - and higher skilled - work time than normal production. Thereby, workload will increase disproportionately, which will once again lead to more errors, etc.

In the model equations, the variable Workload is defined in a similar manner to Oliva & Sterman(2001) and Akkermans & Vos (2003) :

$$(Eq.1): Workload_i = \frac{Processing\_required_i}{Processing\_available_i} = \frac{WIP_i * Work\_content\_per\_unit_i}{Workforce_i * Worker\_Productivity_i}$$

(This formula applies to every one of the manufacturing, rework or testing stage, hence the subscript i for stage i.)

Workload affects quality in the following manner, which assumes a so-called “anchoring and adjustment process” (Einhorn and Hogarth 1981), common in behavioral psychology model formulations:

$$(Eq.2): Quality_i = Average\_quality_i * f_1(Workload_i) \\ * f_2(Schedule\_pressure_i) * f_3(Production\_Expertise)$$

Quality in a supply chain stage is equal to the current average quality in the stage times an effect on that quality of workload and of schedule pressure and production expertise, to which we will return shortly.

All these effects on quality are non-linear functions of the input variable. The relation between workload and quality is a version of the Yerkes-Dodson law, which posits an inverted U-shaped relationship between stress and performance on moderate to difficult tasks (Yerkes & Dobson 1908, Fisher 1986, Eigen 1994).

The values for this curve that are used in this model are shown in Figure 4. This graph shows that for normal workloads, between 0.75 and 1.25, there is no effect of workload on quality, but for very low workloads of less than 0.5 and especially for very high workloads of 1.5 and higher, the effect on quality becomes significant, up to a degradation of 25% for workloads of 2.5 and higher. In the model, it is conservatively assumed that this workload effects is only effective in the actual manufacturing stages, not in testing, rework and transport stages.

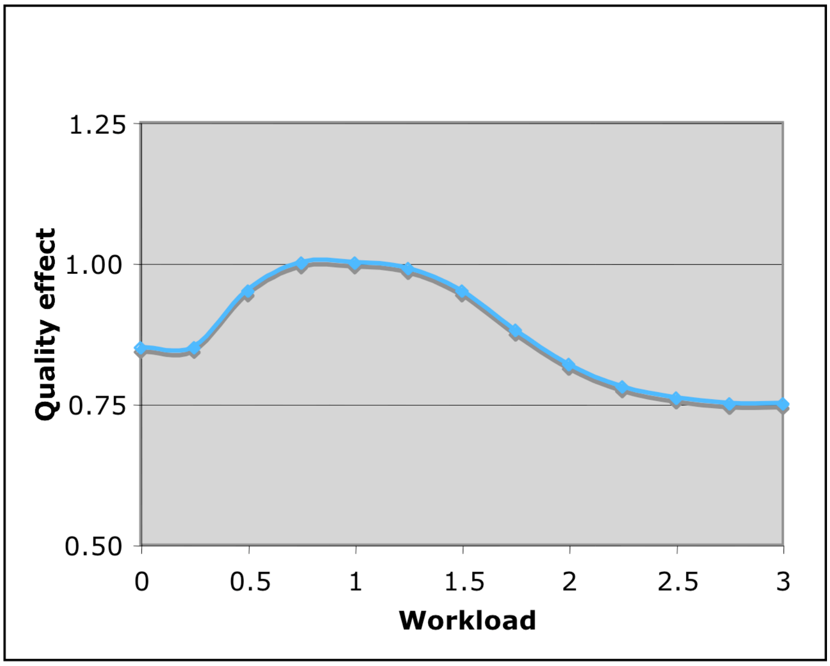

Figure 4: Relation between workload and effect on quality

Finally, it should be noted that workload also affects detection likelihood in the test centers in a similar manner, again via an anchor-and-adjustment type of multiplier effect on the “normal” detection likelihood, which is set at 95% in the base case version of the model.

R2. *The schedule pressure loop*. Quality is also negative affected by work pressure. This reinforcing loop states that for high levels of work pressure, employees will respond to work pressure by adjusting their behavior to meet throughput expectations. This effect was empirically observed by Mass & Berkson and by Oliva & Sterman (2001). The latter also empirically measured its effect. The feedback loop reads as follows: the higher the work pressure, the lower the quality, hence the lower the output, hence the higher the schedule pressure.

In the equations, Work pressure has as its input the % of market needs met, so:

$$(Eq.3): Perc\_MktNeeds\_Met_t = \frac{100\% * \sum_{T=0}^{T=t} Shipments_t}{\sum_{T=0}^{T=t} \Pr oduction\_starts_t}$$

In this model, production starts equal the target ramp-up of marketing.

The slope of the curve of the function that translates values of work pressure into an effect on quality is a monotonically increasing from 0.81 when 0% of the market needs are met up to 1.0 when market needs are fully met for 100%.

B1. The workforce adjustment loop. Two key balancing feedback loops are also shown in Figure 3. The first one is the workforce adjustment loop. In its general meaning, this is a fairly straightforward one: when workloads rise beyond normal levels, and after overtime has been exhausted as a buffer device, more employees will be hired, with some dealy. Conversely, when there is not enough work for the current workforce, people will be fired, again after a delay.

$$(Eq.4): Desired\_workforce_i = Experienced\_workforce_i$$
$$+In\exp erienced\_Workforce_i * f_5(Workload_i)$$

This formula again applies to every stage in the production process. Workforce consists of inexperienced and experienced employees, and only inexperienced employees are hired. Initially, there is a 50-50% distribution of "rookies" and "pros". f_5, the effect of workload on hiring, increases monotonically from 0.9 to 1.1. as workload increases from 0 to 3 and beyond.

The rate equation for the hiring of new staff takes this desired rate into account, after a delay of 4 weeks:

$$(Eq.5): Change\_in\_workforce_i = MAX(0,(Desired\_workforce_i$$
$$-In\exp erienced\_labor_i - Experienced\_labor_i)/Hiring\_delay)$$

B2. *The production expertise loop*. The second balancing loop is less straightforward. One good thing about repairs is that they tend to expose flaws in the production process. Thereby, the more repairs are made, the more process expertise is accumulated regarding the upstream production steps. In this way, more errors lead to more repairs, which will lead to higher skill levels, which will reduce errors over time.

This effect is different from a normal learning curve, which means that productivity, defined as the # units processed per employee per time unit, goes up. A learning curve is operationalized in the model through the flow of inexperienced employees to experienced staff, after a 2 month delay. Inexperienced employees have 80% of the normal productivity, experienced ones 120%.

As Equation 6 shows, production expertise is accumulated by every repair, but not all repairs carry equal weight, their relevance decreases over time and how effective this insights is fed back to production operations also varies.

$$(Eq.6): Production\_Expertise = \int_0^t \begin{array}{l}(Upstream\_repairs * Insightgainedperrepair\_Upstream \\ Downstream\_repairs * Insightgainedperrepair\_Downstream \\ + Pit\_repairs * InsightgainedperPitRepair) \\ * Quality\_of\_feedback \\ * Timeliness\_of\_feedback\end{array}$$

The level of insight gained per repair upstream is only 20% of insights gained from Pit repairs, which are much more complex and time-consuming as only the most difficult cases end up in this rework stage. Similarly, downstream repairs account 40% of the insight a pit repair generates. Quality of feedback is set arbitrarily at 20% in the base case (but subjected to sensitivity analysis in one of the policy scenarios later on), and Timeliness of feedback degrades nonlinearly from 1 to 0 as these repairs take place later on in the simulated 150 weeks of production.

4.3. Key performance metrics

To understand how this model behaves, we will have to pay special attention to most of the variables that appear in the four feedback loops discussed above. But to compare the performance of different quality management policies, we have to limit our attention to a small set. This set of key performance indicators or KPIs is used in the scenario comparison overview in Table 2 in Section 5. These KPIs are the following:

- *Average Output* (in 000units/week): this is the average rate of shipments over the simulation period (which is 150 weeks or 3 years).
- *Average WIP* (in 000 units): the sum of all inventory levels averaged over the entire simulation period
- *Average Workforce* (in full-time equivalents or fte): the size of the total workforce, inexperienced and experienced, over the entire simulation period
- *Average Quality* (in %): the quality level of outgoing products in the last internal supply chain stage, which is downstream testing, averaged over the entire simulation period. By quality level is meant here the percentage of products that have errors in them that may or may not show up immediately. For finished products, this translates to the likelihood that units will be in need of repair or scrap over their technical lifetime, say 3 years. Anyone who used a computer in the mid-nineties will acknowledge that a 25% likelihood of computer failure in 3 years time was not at all unrealistic. In the base case, our model shows an average quality of 73%. This was also in line with internal company estimates in the original field study.
- *Average Cycle time* (in weeks). This is calculated by dividing the total level of units that are in the supply chain by the flow rate of units going out of the system, which by Little's Law gives the average cycle time (Hopp & Spearman 2000, Sterman 2000). Or, in algebraic notation:

$$(Eq.7): Cycle\_time_t = \frac{\int_0^t Production\_starts_t - Shipments_t}{Shipments_t}$$

In the manner, we have a fair overview of both financial performance indicators, with shipment rate as a proxy for revenues and both headcount and WIP levels as indicators for cost, and key operations performance indicators such as quality and cycle time. This should enable a fair comparison of the pros and cons of the various suggested performance improvements.

4.4. Base case behavior

Before we start analyzing the performance impact of the various structural and control improvements contemplated by company management, let us first explore how this model behaves in the base case. We start with a steep production ramp-up curve, as shown in Figure 5. Production ramps up quickly, up to 10.000 units per week during the peak period. After 90 weeks, or almost two years, there is no longer a market need for additional units to be produced. By then, the next generation of PCs or laptops has long been introduced. All in all, in some 400.000 units have then been put into production.

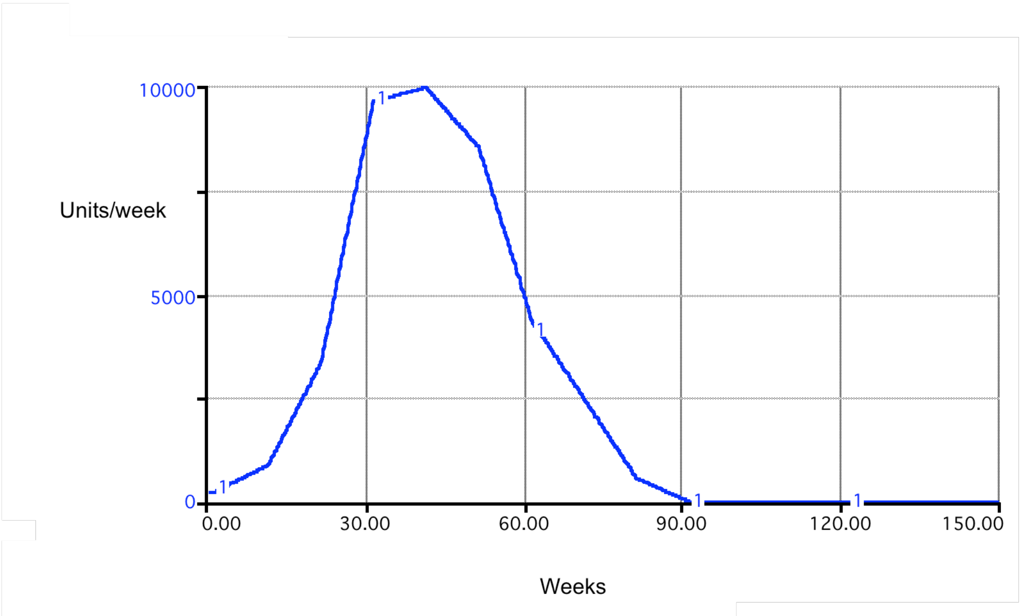

Figure 5: Production starts in base case

What does this inflow of new production starts result in? This is visualized in Figure 6. The peak in upstream productions starts returns here as the first data line, labeled with "1". One can observe how this inflow results in a delayed peak in upstream manufacturing (line 2). However, this peak reappears with a delay and considerably less prominent in the next stage downstream, which is visualized by the line "into transport", labeled with "3". [We will soon see that this is because a large portion of the units produced is found to be defective and ends up as rework.]. Downstream manufacturing can deal with the inflow from upstream quite well, it would seem from the behavior or "4", the downstream manufacturing rate, which is quite similar to line 4. And yet, the shipment rate of finished product into the market ('5") is much smoother than the downstream manufacturing rate. Again, we will see that this is because of defective units as a result of poor process quality in manufacturing.

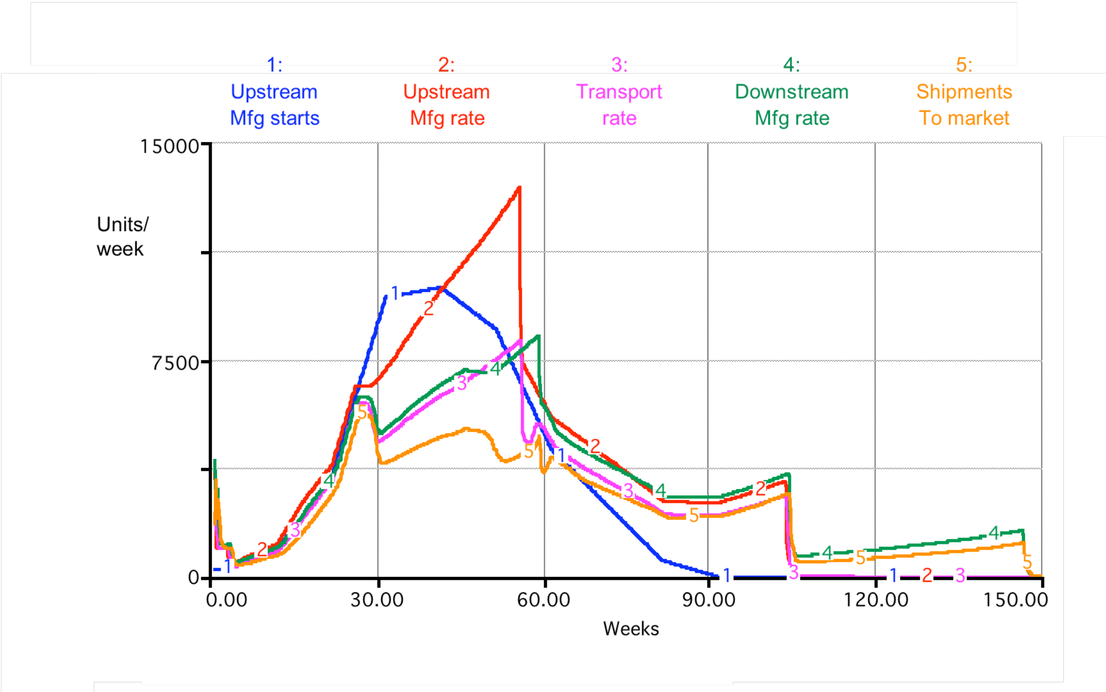

Figure 6: Key upstream and downstream product flow rates in the base case

The line in Figure 6 shows how the peaks that still exist in upstream process stages gradually disappear. Line "1", which visualizes the number of upstream test fails per week, peaks towards 5000 units after 1 year, which explains the bulk of the difference between upstream manufacturing and transportation to downstream in the preceding. Interestingly, it takes the upstream rework case a long time to repair these units. The peak in the upstream repair rate (line "2") is some 40 weeks later. Similar phenomena can be observed downstream as well. There is a peak in downstream test fails (Line "3") around 60 weeks that explains why so much less is shipped to customers than is leaving downstream production in this period. One difference with upstream is that here we have two rework stages. When downstream repair fails (Line "4"), the defective units automatically end up in the 2nd line rework stage called "the Pit", whereas they remain in the 1st line rework stage for the upstream part of the chain. Hence, downstream repairs peak in the pit (Line "5") and, again, after a very considerable time delay.

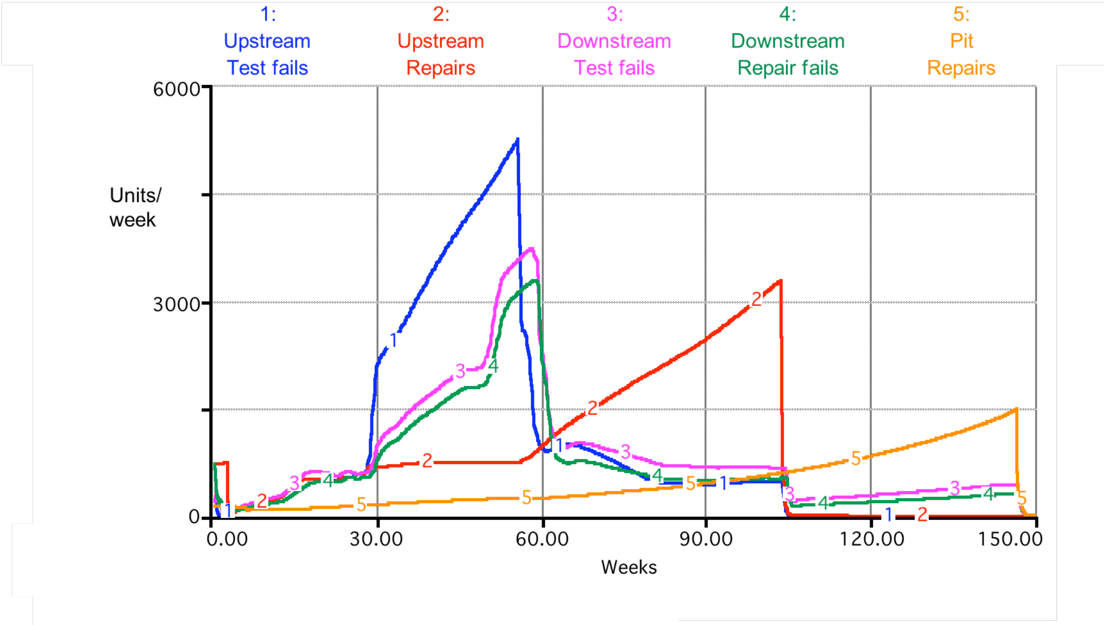

Figure 7: Output rates in test and rework stages in the base case

Why are there towards the end of the first year such high test percentages of units both upstream and downstream? That has everything to do with workloads. We already saw that higher workloads lead to lower process quality and hence to more errors detected and, thereby, to less output of product and instead more rework.

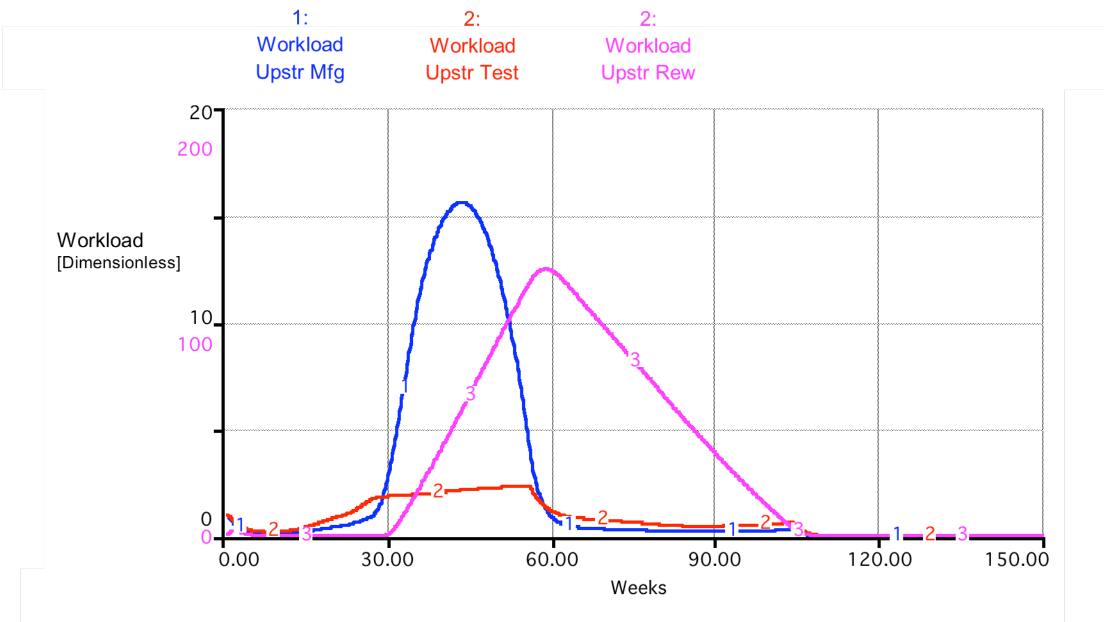

Figure 8: Upstream workloads

This becomes quite apparent from the graphs in Figure 8 and Figure 9 where the behavior of workload per stage is shown. In both upstream and downstream manufacturing, processing capacity is simply not able to deal with the ever-rising inflow of new units to be processed and so workloads (lines labeled “1” in both graphs) rise to very high levels. Workforce is increased, but just not fast enough to keep up with the rise in workload.

As we saw in Figure 6, the peak inflow in upstream manufacturing is higher than that in downstream manufacturing, and so the workload levels are similarly different: up to 15 in upstream manufacturing, yet near 4.8 in downstream manufacturing. Workloads in the test stages (lines labeled “2” in both graphs) are not so badly capacity constrained, since the work content for testing is considerably less than for manufacturing. What they are faced with is a recurrent inflow from reworked units that first failed testing. But in the testing stages, processing capacity is far more capacity constrained, since it takes much longer to figure out what is wrong with a unit and fix it than to “do it right the first time”. Workloads for rework stages (lines “3” and “4” in these graphs) have to be scaled much higher, since these peak at around 100.

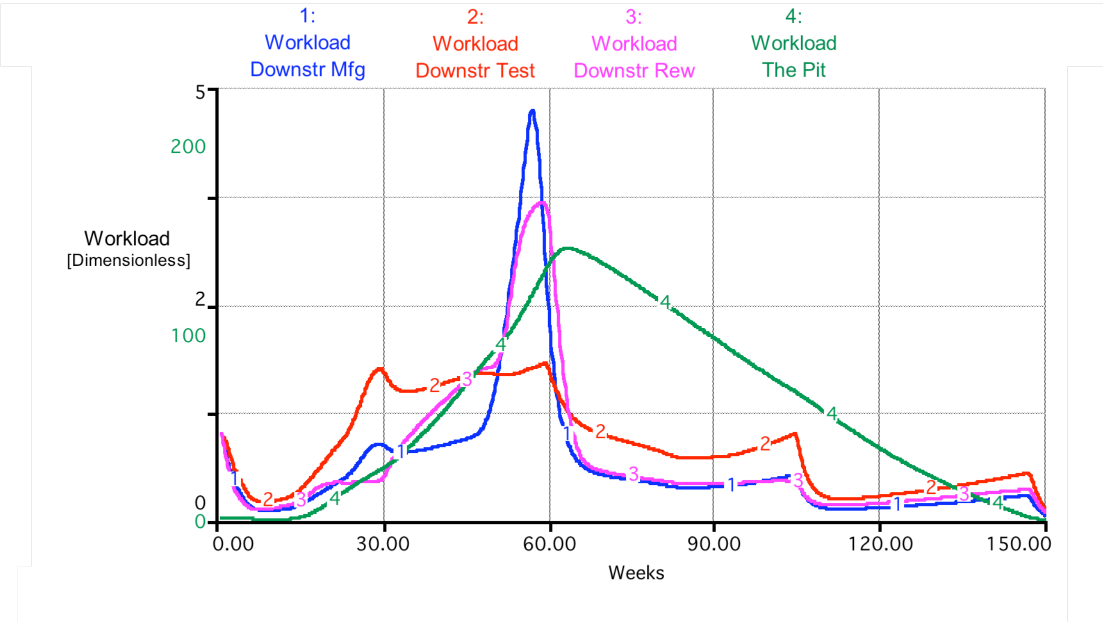

Figure 9: Downstream workloads

Almost by definition, the process quality in each of these stages must drop if workloads become too high. This fits with the graph in Figure 10 that shows quality in both upstream and downstream testing stage, just before the products go into transport. Again, one can witness the effects of the peak in production in the second half of the first year here, with upstream quality dropping to a low from week 30 onwards but recovering strongly after week 60, and with downstream quality dropping later, after week 50, but more steeply, and recovering somewhat later. The lower quality level in the third year is attributable to the high rework levels in the downstream part of the chain during that period. After all, no new units are released into upstream production after week 70, so in the upstream stages most of the negative workload effects have ebbed away by this time.

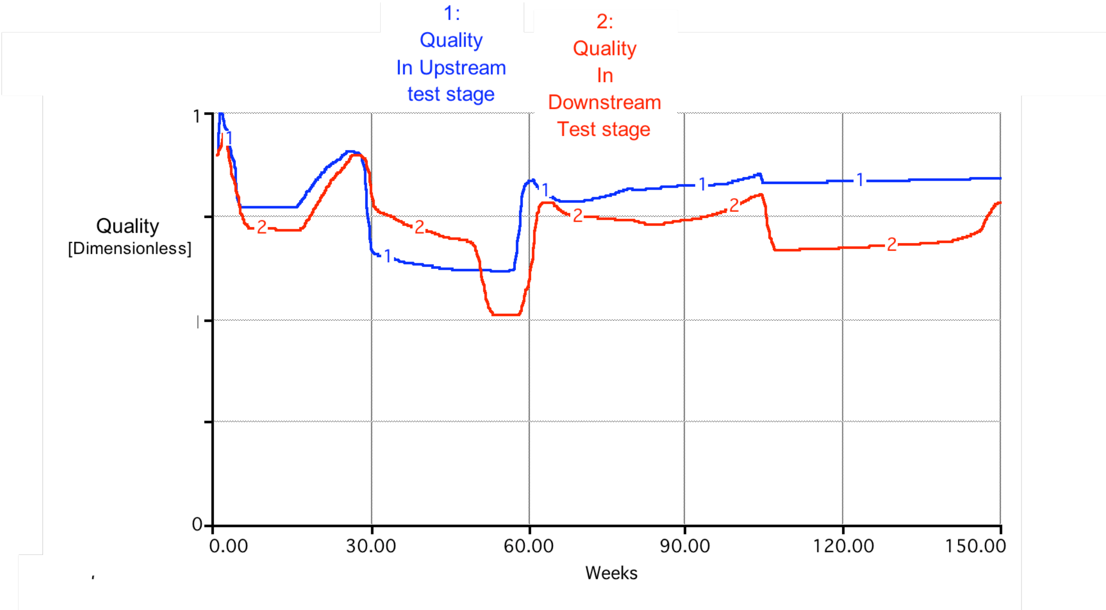

Figure 10: Quality in upstream and downstream test stage

Figure 10 shows us how this supply chain performs in terms of quality. How does this base case scenario do in terms of our other pre-defined performance criteria for output, cost, and time? Here we can look at Figure 11. We see total output to the market displaying the kind of S-curve growth pattern typical of product life cycles. The behavior of total inventory is proportional to the amount of product being pushed into the supply chain, with a significant delay in its peak level compared to the peak in production starts.

The workforce also responds with a delay to the increase in processing requirements, but does not really drop off again, partly because of inflexibilities in the labor force, and partly because rework requirements, which are very labor-intensive, remain considerable until very late in the simulation. Cycle time first grows in line with the growth in WIP, as Little's Law would make us expect, but there is a drop and then a renewed peak in the third year as a disproportionate percentage of the remaining units in the supply chain is rework, which is far more time consuming than regular processing. In comparison: regular manufacturing stands to testing, stands to regular rework and to pit rework as 6:1:8:25 (based on company data).

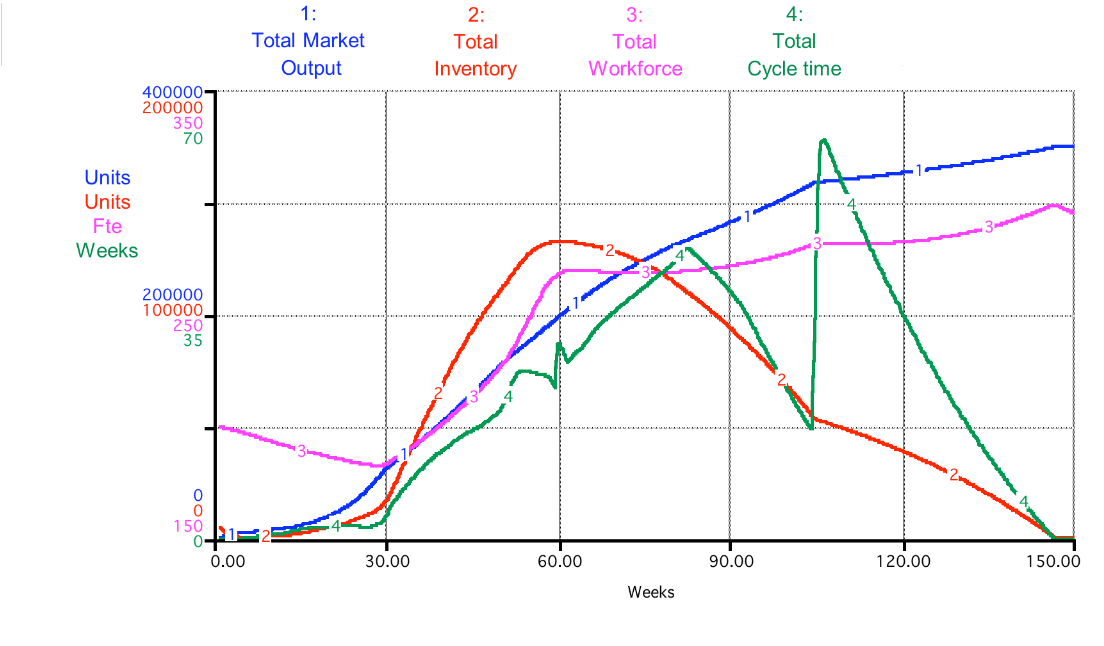

Figure 11: Behavior of key output performance criteria (excl. quality)

Average values for these performance indicators are presented in Table 2 below, where the performance outcomes of the various improvement scenarios that will be discussed next are compared against each other.

5. Scenario and policy analysis

In this section, we explore policies that aim to improve the performance of the supply chain on all levels, but with specific attention to quality, as this is clearly one of the key determining factors for all other performance dimensions (and the focus of this paper).

#	Scenario Name	Average Output (000u/wk)	Average WIP (000 units)	Average Workforce (fte)	Average Quality (%)	Average Cycle time (wks)
0	Base case	2.35	58.7	250	73%	43
1	Modest ramp-up	2.53	32.3	228	78%	28
2	High component quality	2.46	47.1	259	77%	37
3	High upstream quality	2.40	55.1	258	73%	44
4	High downstream quality	2.54	35.1	251	78%	21
5	High error detection	2.34	61.4	248	74%	44
6	Upstream info feedback	2.42	55.4	252	77%	41
7	Add Pit staff	2.37	41.9	276	77%	25
8	High quality gate (75%)	0.58	132.2	225	51%	169
9	Modest quality gate (65%)	2.36	58.7	249	74%	42
10	Combination of 6 & 7	2.45	38.8	276	82%	22
11	Combination of 5, 6, 7 & 9	2.45	39.9	275	83%	23
12	Combination of 1, 6 & 7	2.58	18.2	246	83%	11

Table 2: Summary of performance impact of different policy scenarios

5.1 Modest versus aggressive ramp-up

One of the quality management measures that was not under serious consideration was a more aggressive production ramp-up schedule. Company management that being first on the market was really a strategic imperative for the company, and that this could only be achieved by stressing the supply chain to the limit. However, also an earlier study with the same company had suggested that taking the raw edge off production ramp-ups might have very beneficial results for the product development area (Mass & Berkson 1995).

As Figure 12 shows, the notion of "going slow to go fast" also applied to the manufacturing supply chain. The line labeled "1" shows production starts under the "normal" fast ramp-up, the line labeled "2" shows how the same total product volume could be brought into the supply chain but with a more moderate growth rate.

What company management expected, was that a slower ramp-up of *inputs* to the supply chain, would also lead to a slower *output* growth of the chain. But what Figure 12 suggests, is that the opposite is the case. The line labeled "3", which is the output rate under the aggressive ramp-up regime, actually is slower in bringing finished product to the market than line "4", the output rate under the moderate ramp-up regime.

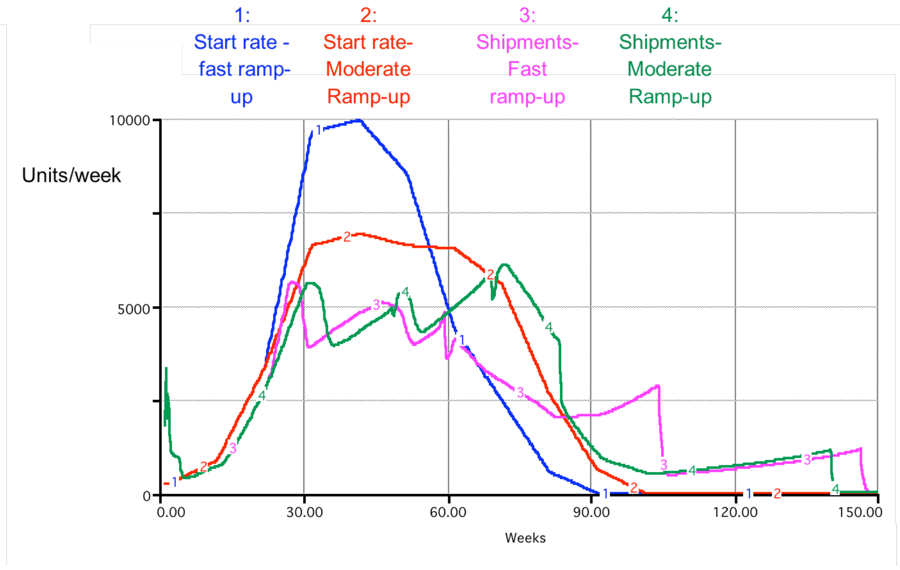

Figure 12: Comparison of in- & output rates for fast and moderate ramp-ups

Moreover, Scenario 1, as this is called in Table 2, considerably outperforms the base case with the fast ramp-up on every aspect of performance. Not surprisingly, this scenario was difficult to digest for company management during the original field study.

5.2 High component quality

Less controversial was the outcome for Scenario 2, the high component quality policy. In this scenario, both upstream and downstream component quality is increased from 90% to 99%. Getting such a policy implemented was the harder part of this scenario, obviously, as this would require significant coordination with component suppliers, if it was technically feasible at all.

5.3 Consistently high manufacturing quality

Another scenario that made intuitive sense to management as it resonated thoughts from lean or TQM as it was called at the time, was to redesign manufacturing operations such that quality of the output would become virtually independent of workload and worker skill-level, for instance through the use of fail-safe equipment. As this would require major redesign and employee training, this was thought to be a very costly policy and one for the longer term, but still one important enough to evaluate through the modeling effort.

What was not intuitively clear was where such manufacturing quality measures would pay off most: upstream, downstream or both. Scenarios 3 and 4 in Table 2 provide us with a preliminary answer: downstream quality improvements yield far more benefits than upstream ones. In both scenarios, quality in the actual manufacturing stage is fixed at 98%. Which is somewhat counter-intuitive, one might argue, as doing-it-right-the-first-time would imply high quality upstream. At least this was the finding from Akkermans & Vos (2003), where high service quality upstream was found to be crucial for high quality throughout the chain.

Model analysis suggests that this is because of workload effects. With high upstream quality your output becomes somewhat higher, because during the peak there is more output from upstream to downstream, but this also results in higher workloads downstream and therefore actually marginally lower quality delivered than in the base case (72.9% versus 73.2%).

5.4 Superior error detection

At the time of the original field study, company management was considering investing in better error detection equipment. The results from Scenario 6 suggests that this is not a high-leverage investment. There is some improvement, but only marginally so. Model analysis suggests that this is because in the base case, detection likelihood is already fairly high, at 95%, and workload effects are relatively limited in size, so there is relatively unused improvement potential to address, even with a detection likelihood of 99%.

5.5 Improve upstream info feedback & add experienced Pit staff

Company management puzzled about what to do with the defective units in "the pit", as it was evocatively called internally. On the one hand, this was just stuff taking up factory space. On the other hand, management realized that hidden somewhere in this pile, there could lay the answers to many of the thorniest quality issues. So, there was genuine interest in seeing what could be done in this area.

The modeling team looked at two related aspects of doing more with Pit inventory. Scenario 6 shows what would happen if far more attention would be placed on feeding back insights from repairs to upstream manufacturing. Since the upstream part of the supply chain was located at a different site, and given the extremely high work pressures, especially during the early stages of the ramp-up, this happened only sporadically and piecemeal. What if information feedback would really be addressed thoroughly?

In this scenario, the quality of info feedback is changed from 0.2 to 1.0, so a considerable increase indeed. But the pay-off might also be considerable, as the performance outcomes of this scenario in Table 2 suggest. On the KPIs of output rate and quality delivered, this policy shows up as being just as effective as the next policy alternative, Scenario 7, which suggests adding 10 experienced workers to the Pit staff, a fairly costly measure on a total of 200 fte. The information feedback scenario clearly outperforms the additional Pit resource policy of Scenario 7 in terms of workforce but is less effective in reducing WIP and cycle time. Why this difference occurs is fairly straightforward. More experienced staff in the not only leads to more and more timely feedback on quality issues upstream, but also to far more units repaired. And that means lower stock levels and, by definition, lower cycle times. A combination of both scenarios would even be more promising, as we will see in combination scenarios 10-12.

5.7 Quality gates, both high and modest

The discussion between modelers and company management on the effectiveness of quality gates, or "phase gates" as checks of product quality was an especially difficult ones. This was and continues to be a popular practice in many industries and is very much in line with lean/TQM thinking. In the words of Mass & Berkson (1995), "A quality gate is a hard screening and rework point beyond which products cannot pass until they are upgraded to at least the gating quality level". At least these authors found that during the early stages of product development, when quality levels are really low, so under 50%, such quality gates have benefits, but they too had found that "downstream quality gates add progressively less value and can eventually destroy value. (..) a quality gate positioned behind the prototype stage or in integrity testing is worse than no quality gate at all" (Mass & Berkson 1995, p.24).

Why is this? Because when the quality gate is set too high, workloads reach levels that go beyond the tipping point and upstream manufacturing gets caught in an almost never-ending recycling of low quality and huge rework rates, from which it cannot escape through superior insights from downstream repair work. In essence, it is simple: every week, new units are put into production, driven by the ramp-up schedule, but no units are leaving the production stage. So, work in progress and hence workloads keep piling up, far faster than the workforce can grow to accommodate it.

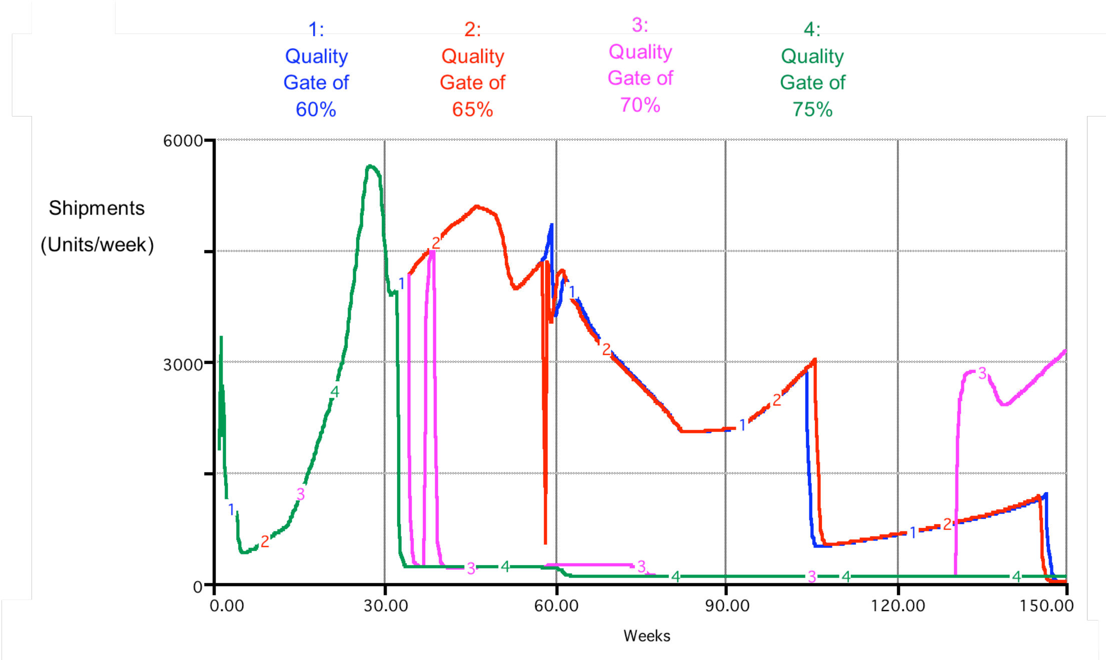

Figure 13: Impact of different quality gates on shipment rate

Figure 1 shows the impact of different gating quality levels. At 60%, line "1", there is effectively no quality gate since quality never drops below 64% in the base case. At 65% (the line labeled "2"), there is actually some improvement over the base case (Scenario 9 in Table 2). Although there is a brief drop of output around week 55, this helps somewhat in making downstream workload levels rise further, almost as a kind of Input-Output control (Wight 1970, Hopp & Spearman 2000). In this way, average output rates and quality levels are marginally higher than in the base case.

Line "3" shows the dramatic effects of increasing the gating quality level too high. At week 40, in the peak of the production ramp-up, the quality gates prevents product flows to downstream and it takes the supply chain until the end of the simulation to recover. When the gating quality is raised to 75%, as is illustrated by line "4", the supply chain never recovers.

Clearly, in these last two scenarios, the system has been moved beyond its tipping point, as Repenning et al. (2001) also found for product development projects in which fire-fighting became the rule. As these authors formulated this, "In models of infectious diseases, the tipping point represents the threshold of infectivity and susceptibility beyond which a disease becomes an epidemic" (Repenning et al., 2001) Similarly, in supply chains there exists a threshold for quality levels that, when crossed, causes a vicious cycle of low quality leading to high rework leading to high workloads leading to low quality to spread rapidly to the entire supply chain, in this case the upstream part of the chain.

5.10.Combined policies

Scenarios 10 tot 12 show how a combined policy can yield superior results. Scenario 10 shows that leveraging the Pit and the production knowledge that repairing its content can give is really a worthwhile endeavor, with an average quality level of 82% and an average cycle time of almost half the value for the base case. Scenario 11 shows that not every combination adds value. Better error detection and a modest quality gate don't significantly improve the performance of the supply chain. The last combination of policies, Scenario 12, adds to the Pit leverage policies the notion of the moderate ramp-up. This is clearly a very successful scenario, with superior performance on every aspect. Moreover, it does not require major investments (not even in workforce), but primarily a different managerial attitude towards marketing and knowledge management.

6. Propositions

What do these experiments tell us regarding the topic of this paper, effective quality management (QM) during production ramp-up in supply chains? Of course, several caveats need to be made regarding validity and generalisability. Nevertheless it is hoped that the combination of an in-depth case study, a fairly generic model and a closely aligned set of relevant causal relations from behavioral psychology provides sufficient triangulation opportunities to give credibility to our findings.

We must remain cautious with drawing far-reaching conclusions regarding specific policies on the basis of the specific outcomes of these policies in this specific model. After all, much will depend on specific parameter values and initial values of variables. Nevertheless, it seems appropriate to suggest the following four propositions:

Proposition 1. *[Single point solutions don't work]. Effective QM policies during production ramp-up will use smart combinations of different policies that reinforce each other. "Rounding up the usual suspects" is almost certainly a sub-optimal solution.*

Proposition 2: *[Keep away from the tipping point.] Effective QM policies during production ramp-up will ensure that workload levels in supply chain stages do not rise so high that the entire chain gets bogged down in a vicious cycle of high workloads, low quality levels, high rework levels and hence even higher workloads.*

Proposition 3: *[Go slow to go fast.] Effective QM policies during production ramp-up will avoid frontloading the supply chain with so much extra work that the net result is that product comes out slower, rather than faster, compared to a more moderate production ramp-up.*

Proposition 4: *[Get out of the pit]. Effective QM policies during production ramp-up will ensure timely downstream feedback on hidden upstream quality issues through*
(a) releasing enough low-quality WIP early
(b) adequately resourcing downstream repair, and
(c) comprehensive and timely feedback.

Conclusions

This study has looked at the highly dynamic issue of managing quality during production ramp-ups in high-tech supply chains. It has based its analysis on an in-depth case study of one particular high-tech supply chain setting. It has benefited from insights from the recently-emerging literature on behavioral operations, in particular on the interrelations of workloads, quality levels, schedule pressures, rework levels and production knowledge that appear to be at the core of this issue. It has combined the field data and the theory into a more general system dynamics simulation model of such as supply chain setting to explore the relative contribution of different policy options for effective quality management.

Fundamentally, what this study illustrates is that “rounding up the usual suspects” to select such policies is a perhaps understandable but very dangerous managerial attitude. This study agrees fully with Mass and Berkson (1995) when they state that: “Often, the problem is a failure to consider [the supply chain] as a dynamic whole. Managers try to improve performance by breaking the process apart. What’s needed instead is an integrated, holistic approach that captures [the high-tech supply chain] as a complex system. It must recognize and account for the cascading sequence of causes and effects that determine how well a company delivers against the key performance metrics of speed, quality, and cost.” (Mass & Berkson 1995, p. 20-21).

When behavioral psychology, operations management and system dynamics simulation join forces to address this challenge, it might just be the beginning of a beautiful friendship….

Acknowledgements

This paper is dedicated to the memory of Nat Mass, who originally made this research possible. Sincere thanks go to Kim van Oorschot (INSEAD) and Paulo Gonçalvez (University of Miami) for their helpful comments and suggestions.

References

Akkermans H.A. (2001) *Renga*: a systems approach to facilitating inter-organizational network development. *System Dynamics Review* 17: 179–193.

Akkermans, H.A., Vos, C.J.G.M. (2003): Amplification in Service Supply Chains: An Exploratory Case Study from the Telecom Industry. *Production and Operations Management 12(2)*, 204-223

Akkermans, H.A. & van Oorschot, K.E. (2005) Relevance Assumed: A Case Study of Balanced Scorecard Development Using System Dynamics. *Journal of the Operational Research Society* 56, 931-941.

Anderson, E.G. & Joglekar N.R. (1996) Modeling the Dynamics of Technological Ramp-up within Firms. *1996 International System Dynamics Conference*, Cambridge, Massachusetts, System Dynamics Society.

Anderson, E.G. & D.J. Morrice. 2000. A Simulation Game for Service-Oriented Supply Chain Management. *Production and Operations Management* 9 (1): 40-55.

Anderson E.G., Morrice D.J. & Lundeen G. (2005) The physics of capacity and backlog management in service and custom manufacturing supply chains. *System dynamics Review* 21(3), 217-247.

Boehm, B. & Turner, R. (2005). Management Challenges to Implementing Agile Processes in Traditional Development Organizations, *IEEE Software* 22 (5), 30-39.

Boudreau, J. Hopp, W., McClain, J.O. & Thomas, L.J. (2003) On the Interface Between Operations and Human Resources Management. *Management of Service Operations* 5(3), pp.

Eigen, K.H. (1994) “Yerkes-Dodson: A law for all seasons.” *Theory and Psychology*,

4: 525–547.
Einhorn, H., R. Hogarth. (1981) Behavioral decision theory: Process of judgment and choice. *Annual Review of Psychology* 32 53–88.
Evans, G.N., Naim, M.M. & Towill, D.R., (1998) Application of a simulation methodology to the redesign of a logistical control system. *International Journal of Production Economics* 56–57, 157–168.
Feigenbaum, A.V. (1961) *Total quality control: Engineering and management*. McGraw-Hill, New York.
Fine, C. H. (1998). *Clockspeed: Winning Industry Control in the Age of Temporary* Advantage. Reading, Mass., Perseus Books.
Ford, D. N., & Sterman, J. D. (1998). Dynamic Modeling of Product Development Processes. *System Dynamics Review*, **14**(1), 31-68.
Ford, D. & Sterman, D., (2003). The Liar's Club: Concealing Rework in Concurrent Development. *Concurrent Engineering*, **11**(3), 211-219.
Gino, F. & Pisano, G. (2006): Behavioral Operations. *Harvard Business School Working Paper*, Version February 3, 2006.
Glaser, B. & Strauss, A., (1967) *The Discovery of Grounded Theory: Strategies of a Qualitative Research.* Wiedenfeld and Nicholson, London.
Gonçalves P., Hines J. & Sterman J. (2005) The impact of endogenous demand on push-pull production systems. *System dynamics Review* 21(3), 187-216.
Hall, R.W.(1987) *Attaining manufacturing Excellence*. Dow-Jones Irwin, Homewood Ill.
Holweg M., Disney, S.M., Hines, P. & Naim, M.M., 2005. Towards responsive vehicle supply: a simulation-based investigation into automotive scheduling systems. *Journal of Operations Management* 23, 507-530.
Hopp, W.J. & Spearman, M.L. (2000*) Factory Physics*. Irwin McGraw-Hill, Boston.
Iansiti, M., & MacCormack, A., (1997) Developing Products on Internet Time, *Harvard Business Review*, September-October 1997, pp.108-117
Juerging, J. & Milling P.M. (2006) Manufacturing start-ups in the automobile industry. *Proceedings International System Dynamics Conference*, Nijmegen, July 2006. Accessible at www.systemdynamics.org
Karlström D. & Runeson, P. (2005): Combining Agile Methods with Stage-Gate Project Management. IEEE Software 22(3), 43-49.
Morris, B (2006) "Tearing up the Jack Welch playbook". *Fortune* June 11, 2006.
Kennedy, M.N. (2003) *Product development for the lean enterprise.* Thee Oaklea Press, Richmond VA.
Mass, N.J. & Berkson, B. (2005) Going slow to go fast. *McKinsey Quarterly* 1995(4), 19-29.
Miles, M.B., Huberman, A.M. (1984) *Qualitative Data Analysis: A Sourcebook of New Methods*. Sage, Beverly Hills,
Molokken-Ostvold, K. & Jorgensen, M. (2005): A Comparison of Software Project Overruns-Flexible versus Sequential Development Models. *IEEE Transactions on Software Engineering* 31(9) 754-766.
Ohno, T. (1988) *Toyota Production System. Beyond Large-Scale Production*. Productivity Press, New York. Originally in Japanese: *Toyoto seisan hôhshiki* (1979) Diamond, Inc, Tokyo.
Repenning, N. (2001). Understanding Fire Fighting in New Product Development, *Journal of Product Innovation Management*, **18**(5), 285-300.
Repenning, N. (2002). A Simulation-Based Approach to Understanding the Dynamics of Innovation Implementation. *Organization Science,* **13**(2), 109-127.

Rudolph, J.W. & Repenning, N.P. (2002) Disaster Dynamics: Understanding the Role of Quantity in Organizational Collapse. *Administrative Science Quarterly*, 47 (2002): 1–30

Schonberger, R.J. (1982) *Japanese manufacturing techniques. Nine hidden lessons in simplicity*. Productivity Press, New York.

Sterman, J.D. (1989) Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. *Management Science*. **35**, 321-339.

Sterman, J.D. (2000) "Business Dynamics: Systems Thinking and Modeling for a Complex World." Chicago, IL, Irwin-McGraw Hill.

Takeuchi, H. & Nonaka, I. (1984) The new product development game. *Harvard Business Review* Jan-Feb 1984, pp.137-146.

Vennix, J.A.M. (1996) : *Group Model-Building. Facilitating Team Learning Using System Dynamic*s. Wiley, Chichester.

Wight, O. (1970) "Input/output control: A real handle on lead time". *Production and Inventory Management Journal* 11(3), 9-31.

Womack, J.P., Jones, D.T. & Roos D. (1990) *The machine that changed the world*. Simon & Schuster, New York.

Yerkes R.M. & Dodson J.D. (1908) The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurological Psychology* 18: 459–482.