



Norwegian University of
Science and Technology

Production Planning and Control in Make-to-Order Manufacturing

Selection Criteria

Ruben Eide Strøm

Global Manufacturing Management

Submission date: June 2018

Supervisor: Erlend Alfnes, MTP

Norwegian University of Science and Technology
Department of Mechanical and Industrial Engineering

Abstract

Purpose: The purpose of this thesis is to identify appropriate Production Planning and Control (PPC) methods in a Make to Order (MTO) manufacturing environment. This is done through two steps. First, the MTO environment needs to be defined. This narrows down the field of appropriate PPC methods, as well as highlights the characteristics of the environment in which the methods are to be applied. By looking at the variety in systems that has undergone review, these characteristics should be defined under two categories: complexity and dynamism. Complexity represents numeral values that are known, such as Bill of Material (BOM), number of routings, product mix (number of products), and number of converging or diverging routings. Dynamism represents numeral values that vary over time in a significant way, such as demand fluctuation, processing time variation, and rate of innovation (new product introduction). Using these environmental characteristics, PPC methods will be compared by how they perform in the different scenarios. Key PPC methods as suggested by Stevenson et al. (2005) are under consideration: MRP (ERP), Kanban, CONWIP, DBR (TOC), POLCA and WLC. How these methods perform in various scenarios relating to the environmental characteristics, and how the methods address the important performance issues found in the MTO industry relating to their competitiveness, will be compared in order to provide a guide as to which PPC method to choose.

Design: A literature review will be conducted, identifying key literature that holds validity, both in terms of PPC methods and MTO environmental characteristics. In addition the Key Performance Indexes found in the MTO industry will be used as a base mark for PPC performance given the various manufacturing environments described.

Findings: The MTO environment were found to include a number of definitions. High-Variety/Low-Volume production are for the most part some kind of MTO manufacturing. In addition, Engineer-to-Order (ETO) share a number of traits when compared to MTO, as the production part of the Customer Order Decoupling Point (CODP) is shared. Engineering and reengineering becomes part of the environmental factors (complexity and dynamism) that impact the applicability of a PPC method.

Of the six key PPC methods in the MTO industry defined by Stevenson et al. (2005), four showed the most promise for optimal performance in the MTO environment identified. Kanban does not perform well in MTO other than as a support for a manufacturing line, and MRP uses pure push logic, commonly found in any MTO company that has yet to review

their PPC. In certain scenarios push does perform better, at least when compared to the identified PPC methods. If complexity and dynamism reach very high levels, the variable inputs needed to build any of the other PPC systems are impossible to define. If the complexity and dynamism lie within acceptable levels, all four could be viable.

WLC is the system that shows the most promise. This is because two other PPC methods share many traits with WLC, namely CONWIP and DBR. This lowers the barrier of implementation, and allows for a gradual maturing of the system, which can start out as a loose CONWIP/DBR, and over time increase the number of deciding inputs to become a WLC. POLCA is a stricter system than WLC (especially if a simple, loose CONWIP/DBR is used as a base), but does show great promise. Of the two systems, WLC handle dynamism better, whereas POLCA handle complexity better.

Research limitations: There is severe limitations when studying PPC methods without a case study. Because the parameters of each manufacturing environment receives varying degree of attention in the reviewed literature, a case study would help pinpoint crucial environmental factors. It would also help in showing the actual functionality of the studied methods.

Value: The thesis will hold value for companies that fall under the MTO category defined in the thesis, given that they need to select a PPC system to improve production performance, and address any performance issues they may have.

Keywords: Production planning and control, Kanban, CONWIP, WLC, POLCA, DBR, TOC, MTO, ETO, HVLV, customization

Contents

Abstract	i
1. Introduction	1
1.1 Research Questions.....	2
1.2 Scope	2
1.3 Objectives	3
2. Methodology	4
2.1 Literature Review	4
2.2 Limitations.....	6
3. Literature Review	7
3.1 Environmental Properties of MTO production.....	7
3.1.1 Criteria for determining applicability of PPC methods	10
3.1.2 Summary of Characteristics in MTO Manufacturing	11
3.2 Production Planning and Control	11
3.2.1 Kanban	12
3.2.1.1 Summary	16
3.2.2 CONWIP.....	17
3.2.3 Drum-Buffer-Rope.....	21
3.2.4 Paired-Cell Overlapping Loops of Cards with Authorization	24
3.2.5 Workload Control	27
3.3 Sequencing rules.....	31
3.4 Summary.....	33
4. Discussion	35
4.1 Defining the MTO industry	35
4.1.1 Criteria for determining applicability	35
4.2 Production Planning and Control	36
4.2.1 Kanban	36

4.2.2 CONWIP	37
4.2.3 DBR	38
4.2.4 POLCA	39
4.2.5 WLC.....	40
4.2.6 PPC Summary	42
4.3 Pre-Shop Sequencing.....	47
4.4 Summary.....	47
5. Conclusion.....	49
5.1 Further Work	50
Reference list.....	51
Table 1 Initial Search Keywords	5
Table 2: Environmental Factors (Birkie and Trucco, 2016a), in (Strøm, 2017)	8
Table 3: POLCA release list (Riezebos, 2010)	25
Table 4: Release and dispatching rules (Thürer et al., 2012b)	30
Table 5: Literature of PPC methods in the MTO industry	43
Figure 1: Lead Time WLC (Oosterman et al., 2000)	28
Figure 2: Workload contribution of job j across time	30

1. Introduction

“Production Planning and Control (PPC) Systems are crucial tools for meeting increasingly high customer demands and expectations in the present highly competitive manufacturing climate.” (Stevenson et al., 2005)

When identifying how Value Stream Mapping applies to High-Variety/Low-Volume (HVLV), the most common improvement areas found was process improvement through the use of Single-Minute Exchange of Die (SMED), standardization of the workplace (5S), error-proofing, and similar, simple universally applicable techniques (Irani, 2011). More complex improvement areas were layout changes and Production Planning and Control (Strøm, 2017). PPC methods applicable were the improvement area that had the most differentiation based on the environmental factors (Birkie and Trucco, 2016a), and therefore is a topic of interest for manufacturers that fall under the category HVLV.

Production Planning and Control (PPC) systems purpose is to improve performance in key areas. These include, but are not limited to, reducing Work in Progress (WIP), minimize Shop Floor Throughput Times (SFTT) and lead times, lower stockholding costs, improve responsiveness to change in demand, and improving Delivery Date (DD) adherence (Stevenson et al., 2005). Addressing these performance goals, PPC systems functionality includes planning material requirements, demand management, capacity planning, and the scheduling and sequencing of jobs. This thesis will focus on scheduling and sequencing by reviewing the functionality of Kanban, CONWIP, POLCA, DBR and WLC in context of a High-Variety, Low-Volume and Manufacture- or Engineer-to-order (MTO/ETO). There are PPC methods that are more complex and E-based, such as ERP and APS systems (or an integration of the two).

When addressing Production Planning and Control (PPC), the main factors controlling how an order flows through the value stream are: Order acceptance, sequencing, release and shop-floor dispatching (Thürer et al., 2016b). This thesis will focus on the internal value stream, which will include sequencing, release and shop-floor dispatching. Five PPC methods are reviewed: Kanban, CONWIP, DBR, POLCA and WLC. These methods determine the “when?”: when is an order released, and how is it controlled throughout the shop-floor. Pre-shop sequencing rules determine the “which?”: which order is selected for release. Shop-floor sequencing determines the “how?”: determining the flow of work, sequencing controls lead time and consequently when an order is finished.

1.1 Research Questions

RQ1: Which Production Planning and Control Methods are applicable in MTO manufacturing?

In order to understand how PPC methods differ, and which are appropriate for which scenario, there is a need to identify the PPC methods that are of further interest. This is done by both identifying methods that do function in MTO manufacturing, as well as identifying the desired performance and improvement goals for a MTO company.

RQ2: What are the selection criteria for applicable PPC methods?

There are a number of methods available. This question aims to find which methods to choose based on the characteristics of a manufacturing environment.

1.2 Scope

When attempting to describe applicability of PPC systems in a HVLV environment, there are two main scopes which has to be defined. First: What PPC systems are evaluated, and why? Second: What are the characteristics of the environment described?

Evaluating which PPC methods to evaluate is difficult due to the fact that there is an increasing number of solutions. On one hand in academics, on the other hand in the professional world, where software producers attempt to lay claim to various markets by claiming that they provide universally appropriate solutions (Stevenson et al., 2005).

Evaluating HVLV environments, there is a need for specific description. The emergence of PPC systems were designed for a Make-to-Stock (MTS) environment. The MTS environment is more predictable and with a higher degree of repetition. Not only is MTS easier to control compared to MTO or ETO, solutions are more universally applicable. Furthermore, in both MTO and ETO, different characteristics affect applicable solutions in a more severe way. It is possible to provide simple products that needs the customization that MTO and ETO offer.

Stevenson et al. (2005) define the MTO industry as the «highly customized» industry. This is further described by Adrodegari et al. (2015), who describe characteristics of the ETO strategy. Regarding “*product customization*”, the ETO characteristics customization characteristics is described as “*high customization, deep and unique bill of materials*”. Stevenson et al. (2005) include product design under the MTO banner, but is described as “*product design can be included in the MTO definition and considered as a remaining*

process to plan and control, such as for Engineer-To-Order (ETO) companies". This thesis will differentiate MTO and ETO for situation if/when different approaches fit better.

1.3 Objectives

The objective of this thesis is to provide guidelines regarding applicability of PPC in MTO manufacturing. This will be achieved by reviewing existing literature, focusing on PPC used in MTO and the reasoning for the presented conclusions. By connecting practices, both traditional and adapted, to the characteristics of the production environment, the goal is to present best-in-class options for a Value Stream Manager to utilize in order to achieve effective PPC with waste reduced to its necessary minimum. The results will be presented in the form of a table highlighting original practice and necessary adaptations dependent on defining environmental factors in the production environment. In order to do so, these objectives need to be met:

1. Describe the problem and scope of the thesis
2. Review of the relevant literature regarding PPC principles in a MTO environment
3. Identify practical solutions as well as the prerequisite environmental factors required for them to be relevant
4. Discuss the contingency factors when selecting a PPC system

2. Methodology

Findings include previous findings (Strøm, 2017). These are used as a basis to build upon when identifying selection criteria and functionality of PPC methods in HVLV manufacturing. The methodology used throughout this thesis is a literature review. This provides a qualitative approach to the problem, answering questions phrased with “why” and “how”, RQ 1 and 2 (Rajasekar et al., 2006). “What” is a question that can be answered in both a qualitative and a quantitative manner. In this thesis the question is answered qualitative, through analyzing and comparing papers relevant to the topic of PPC in MTO.

The thesis is structured through four steps:

1. Define the problem statement
2. Conduct literature review
3. Establish guidelines for applying PPC solutions based on environmental factors in the manufacturing environment
4. Discussion and conclusion

2.1 Literature Review

Using databases Web of Science and Scopus, a literature search was conducted. Combining the search word “lean” with different variations of high variety, the functions of the search sites was used to limit the search to the correct field of science. After it was established with a certainty that the result indeed was providing the literature from the correct field, the abstracts and keywords were skimmed through in order to establish relevant papers. This provided the current collection of papers used in this paper. Because both Agile Manufacturing (AM) and Quick Response Manufacturing (QRM) share many similarities with lean manufacturing, when investigating the applicability of lean manufacturing practices, AM and QRM were included in the search (Stump and Badurdeen, 2012, Gosling and Naim, 2009). This ensured that when either topic was covered in the literature, if a lean practice was used it would be discovered, adding to the number of sources when assessing applicability of lean practices. Lean, agile or QRM are prerequisites for an article or conference paper to be reviewed. The lack of any of mentioned terms in the keywords or abstract cause discarded. Table 1 shows the search phrasing and the date in which the search was performed. In addition papers suggested by researchers with affiliation to NTNU were used.

Table 1 Initial Search Keywords

Keywords 1	Keywords 2	Keywords 3
lean OR agile OR QRM	AND eto OR mto OR "high variety" OR "high mix"	
lean OR agile OR QRM	AND eto OR mto OR "high variety" OR "high mix"	
lean OR agile OR QRM	AND "engineer to order" OR "make to order"	
lean OR agile OR QRM	AND one-of-a-kind OR "project manufacturing" OR "customized manufacturing"	
lean OR agile OR QRM	AND "engineer to order" OR "make to order" OR eto OR mto OR "high variety" OR "high mix" OR "one of a kind"	
lean OR agile OR QRM	AND "engineer to order" OR "make to order" OR eto OR mto OR "high variety" OR "high mix" OR "one of a kind"	AND tools OR techniques OR practices
lean OR agile OR QRM	AND "engineer to order" OR "make to order" OR eto OR mto OR "high variety" OR "high mix" OR "one of a kind"	AND tools OR techniques OR practices OR SMED OR VSM OR POLCA OR CONWIP
lean OR agile OR QRM	AND "engineer to order" OR "make to order" OR eto OR mto OR "high variety" OR "one of a kind"	

This ensured a solid body of literature. Because PPC is an important part of VSM, these searches provided relevant literature when assessing PPC in MTO. Using EndNote, the abstract was once again reassessed. Relevant papers were opened using Google Scholar, Science Direct, Emerald Insight and Ieee Xplore and reviewed to ensure that they were indeed relevant. By reading the papers they were labeled by content. Through this process, the references of the reviewed literature were assessed, and through a process called the “snowball effect”, additional papers were added to the body of literature, going through the same process as described previously in this chapter.

After a body of literature regarding PPC in HVLV, MTO, ETO and other relevant environments had been reviewed, additional searches was conducted, and relevant papers were “cherry picked” based on a short analysis of the title and abstract. Due to an increased

number of hits in Scopus rather than Web of Science, final additions to the literature were conducted in Scopus only. Relevant papers were found using Google Scholar.

For the additional searches conducted, PPC searches in Scopus were made with the search words: Kanban, CONWIP, DBR, POLCA, WLC, MTO, ETO, HVLV, as well as synonyms of and the full definitions without abbreviations.

2.2 Limitations

Production Planning and Control is as will be described later an element of manufacturing management that can be solved in many ways. Which of the methods that perform the best is still up for debate, and one of the main problems is the lack of case studies (for POLCA), and the contradiction between case-studies and simulation (WLC). Therefore, the lack of a case-study in this thesis is a major limitation, as it fails to address the most pressing issue in the field.

3. Literature Review

The following chapter will define an environment that fall under the category MTO, and how four different PPC methods apply dependent on the characteristics that commonly occur in the defined MTO environment. The basis comes from an investigation of High-Variety/Low-Volume (HVLV) manufacturing (Strøm, 2017), where MTO and ETO were amongst the definitions of possible environments that could undergo the HVLV definition. This was too vague, and a proper definition for the environment is needed in order to properly categorize suitable PPC methods.

The chosen PPC methods are based on previous literature. Stevenson et al. (2005) highlight both the nature of the MTO industry, as well as reviewing key concepts within the industry. Here, Material Requirements Planning (MRP), Theory of Constraints (ToC), Workload Control (WLC), Kanban, Constant Work-in-Process (CONWIP), and Paired Cell Overlapping Loops of Cards with Authorization (POLCA) are highlighted as key concepts within the PPC literature. The overall conclusion however, does not define the proper suitability of each PPC methods given the environment they are implemented in. Therefore there is a need to investigate what defines these methods, what they share in common, and how they differ.

3.1 Environmental Properties of MTO production

Before addressing appropriate PPC methods, the environment in which they are to be analyzed needs to be identified. Business size (Stevenson et al., 2005) state that applicability of PPC methods in the MTO sector include applicability for Small and Medium sized Enterprises (SMEs). Defining the size of a company are done by assessing revenue, balance, and number of employees. In the MTO industry there are different key characteristics compared to MTS, such as the importance of accurate and competitive delivery date quotations at the customer enquiry level (Stevenson et al., 2005).

In addition to dividing HVLV manufacturing between MTO and ETO, there is a distinction of the degree of repeatability of an order: Repeat Business Customizers (RBC) and Versatile Manufacturing Companies (VMC) (Amaro et al., 1999). RPCs provide customized products continuously over a period of time (the length of a contract), whereas VMCs provide a high variety of products with variable demand that are produced in small batches and little to no repetition. Even with this distinction, complexity within each category is not uniform (Stevenson et al., 2005). This underlines the statement that there is little to no “one size fits all” solution in the HVLV environments presented in this thesis. The main operational issues

in the MTO industry is capacity planning, order handling (acceptance/rejection), and due date adherence (Soman et al., 2004).

Characteristics of the Make-to-order industry

MTO production is production where, when an order is placed, the design is ready. After the order is accepted, what remains is manufacturing and assembly (Stevenson et al., 2005). This allows a company to have a modular design. The final product is a combination of existing designs assembled together. Thus, there is no design included in necessary processes from order to final product. MTO manufacturing can be defined by having all manufacturing processes after the customer order is accepted (Stevenson et al., 2005). Gravel and Price (1988) describe a job-shop environment, where production is MTO, showing that MTO can be found in both flow- and job shops.

Characteristics of the Engineer-to-order industry

(Adrodegari et al., 2015), (Birkie and Trucco, 2016a)

Difficulties in planning and controlling production in an ETO environment can be traced back to uncertainty. The specifics of the product produced, the volume, and the mix is uncertain. This makes planning and control more complex. Customer specific, highly customized products produced in low volumes, often one-of-a-kind, paired with labor intensive processes often requiring highly skilled labor further increase the complexity of planning and control (Powell et al., 2014).

Birkie and Trucco (2016a) highlight a number of variables that apply when implementing lean manufacturing in ETO. These are described as environmental factors, and are divided into complexity and dynamism categories, both internally and externally, shown in table 2. In addition to these factors, some additional are added. Both processing time and process time variation were found to be contributing factors when analyzing VSM in HVLV manufacturing (Strøm, 2017).

Table 2: Environmental Factors (Birkie and Trucco, 2016a), in (Strøm, 2017)

	Internal	External
Complexity	1. Product diversity and novelty 2. Process interdependencies 3. Variety of decisions 4. Necessary competence 5. Inconsistency of goals 6. Short life cycle Additional: 7. Processing time (Lane, 2007)	1. Diversity of inputs 2. Diversity of customer segments 3. Suppliers and subcontractors involved 4. Regulatory requirements 5. Technological requirements
Dynamism 	1. Internal performance issues 2. Rate of innovation 3. Changes in modes of production Additional: 4. Process time variation (Lane, 2007, Bohnen et al., 2013, Thüerer et al., 2016a, Stump and Badurdeen, 2012, Darlington et al., 2015)	1. Change in customer demographics 2. R&D expenditure changes 3. Demand unpredictability and instability 4. Suppliers & subcontractors performance predictability 5. Predictability of competitors' actions/pressure 6. Changes in regulatory requirements

Some companies who compete in the ETO environment have large projects where both the shop floor and the workforce is limited to work on one project at a time. As such, there is no pre-shop pool of waiting jobs. However, there are a pre-shop pool of waiting jobs. The scheduling of these jobs can be optimized using correct PPC methods. Is Card-Based PPC a possible solution for this particular problem?

There are many planning processes in an ETO project. All stages contain inputs that are influenced by the accuracy of the PPC method chosen, and the quality and execution of the PPC system implemented.

Groping ETO and MTO together

An argument for grouping MTO and ETO together, is that, from a material flow perspective, MTO and ETO are identical (Olhager and Prajogo, 2012). As this thesis focus consists mainly of material flow perspective, using this basis, identifying PPC solutions which mainly focus on material flow should be the same for both MTO and ETO. Rudberg and Wikner (2004) differ the CODP of engineering and production. What is commonly referred to as the ETO industry, where engineering is done after an order is received, is referred to as engineering ETO, production MTO. There is not necessarily distinction between ETO and MTO based on shop configuration either (Gravel and Price, 1988). Therefore, in the context of appropriate PPC solutions, grouping ETO and MTO together is logical.

3.1.1 Criteria for determining applicability of PPC methods

A set of criteria for determining applicability is described by Stevenson et al. (2005).

- Inclusion of the customer Enquiry Stage for delivery date determinations and capacity planning.
- Inclusion of the Job Entry and Job Release stages, focusing on due date adherence.
- Ability to cope with non-repeat production, i.e. highly customized products.
- Ability to provide planning and control when shop floor routings are variable, i.e. general flow shops and job shops.
- Applicability to Small and Medium sized Enterprises

The goal of PPC in MTO manufacturing is reduced lead times, stable lead times, and high due date adherence. Increased demand for customized products lead to increased competition, which increases the importance of reduced lead time. In the MTO industry there are different key characteristics compared to MTS, such as the importance of accurate and competitive delivery date quotations at the customer enquiry level (Stevenson et al., 2005). When assessing the applicability of MRP in the HVLV industry, the size of the company becomes an important factor for success. The need for tailoring of an existing MRP system to fit the HVLV industry increases the already high cost of implementation. This is a barrier to implementation for ERP implementation in SME dominated markets (Stevenson et al., 2005).

Selecting between Kanban, CONWIP, and DBR, Darlington et al. (2015) use Value Stream Mapping (VSM) and existing MRP systems to extract financial and operational data to create tools for capacity planning, WIP monitoring, and system simulation. The basis for their study

is lack of empirical literature within the field of Lean manufacturing investigating shared-resource problems. The VSM is a way of visually represent routing complexity and variability, that is the complex nature of the routing (which route a product takes, and how many processes are included), and variability of routings between products.

3.1.2 Summary of Characteristics in MTO Manufacturing

When the CODP is moved upstream, more processes experience more variety, because the outcome of the processes depend on the specifics within the customer orders placed. This is also an effect of moving from MTO to ETO. Although the variety in shop characteristics might be similar, ETO includes design after an order is accepted. Depending on the degree of engineering done, the processes necessary to complete a product might be unknown at the time of the order acceptance. This will severely affect the ability to estimate cycle times and routings, as well as increasing the risk of rework and reengineering. Various performance indexes experience the variance based on similar characteristics as the “bullwhip effect”: If something is wrong with the planning, its effect will spread throughout the internal value stream, affecting lead times, delivery date adherence, quality, inventory, etc.

3.2 Production Planning and Control

Assessing appropriate PPC methods for MTO production has previously been done (Thürer et al., 2016a, Stevenson et al., 2005). Applicable solutions depend on process time and routing variability (Thürer et al., 2016a) or shop configuration (Stevenson et al., 2005). Shop configuration may be a less specific way of highlighting process time and routing variability, as especially routing variability will decide on the appropriate shop configuration between the four configurations Pure Flow Shop (PFS), General Flow Shop (GFS), General Job Shop (GJS), and Pure Job Shop (PJS). Stevenson et al. (2005) argue that two PPC methods are applicable across all configurations. Enterprise Resource Planning (ERP) systems is the furthest advancement and general description of systems that incorporate Material Requirements Planning (MRP) and Manufacturing Resource Planning (MRP II) logic. ERP systems is the most commonly used system, used by one third of manufacturing companies studied by Sower and Abshire (2003). The second method applicable across all shop configurations are Theory of Constraints (ToC), and its method Drum-Buffer-Rope (DBR). The second most applicable solutions are Workload Control (WLC), applicable in GFS and GJS, and CONWIP, applicable in PFS and GFS. Kanban is suitable for PFS alone, POLCA is suitable for GFS alone. Thürer et al. (2016a) argue that the only solution to be applicable with

routing and processing time variability is COBACABANA, a card-based version of WLC (Land, 2009).

When implementing a pull systems there are some elements that should be addressed first. Reduction of demand amplification, defect rate, disruption through breakdowns and changeover time, as well as stable work through standardization (Bicheno and Holweg, 2016). Different pull methods exist, some developed to overcome the limitations of Kanban, such as Drum-Buffer-Rope (DBR) and Constant Work in Process (CONWIP) (Bicheno and Holweg, 2016). Advantage of Pull versus push production is that pull production controls WIP, push production controls throughput (Spearman et al., 1990). WIP is easier to control than throughput, and gives inherently better control over flow times. Throughput is near constant whenever production is near capacity. WIP however, will continue to grow as long as input is higher than output, even in perfect conditions. Furthermore, accumulation of inventory, the consequence of increased WIP, will have a negative effect on throughput, increasing the flow time, which in turn increase the difference of input versus output given steady demand. What this shows, is that “worst case” of push is worse than “worst case” of pull: a pull system will carry less risk. When demand is higher than capacity, a pull-system handles the excess better than a push system: rather than accumulating WIP, there will be accumulating pre-shop orders. Because these are easier to manipulate, shortage of capacity is easier to manipulate. Throughout the literature review of pull production, the term “card” is often used. It should be noted that it in addition to being a physical card is transferable to represent other types of signals.

ERP is the most commonly used system. It sets up an order in terms of its material (Material Requirement Planning, MRP) and resource requirements (Manufacturing Resource Planning, MRPII). In addition to the push based logic used in MRP as a separate PPC being of interest, the synergy a PPC method has with MRP systems is of interest. However, as discussed by Stevenson et al. (2005), MRP (and beyond, MRPII, ERP) may be difficult for many MTO companies, often due to company size. Because of this, the reviewed PPC methods are Kanban, DBR, CONWIP, POLCA, and WLC.

3.2.1 Kanban

Basic classification of Kanban divides it into two types: production Kanban, and move or withdrawal Kanban (Bicheno and Holweg, 2016), as well as WIP Kanban (Thürer et al., 2016a). With a MTO environment, there will be a high number of different products (or a

large number of variations of the same product). In such a situation, two variations of Kanban has potential fit: **Sequential Operations Kanban** and **Generic Kanban** (Bicheno and Holweg, 2016). **Sequential operations** sets up a number of sub-assemblies in between stations. This allows for quick response to changes, but a set bill of material is needed, which excludes any engineering changes done (ETO). It also accumulates inventory quickly, and is not fit for an environment with a large number of products. **Generic Kanban** authorizes feeding work centers to make a part without specifying which part is to be made. The specification of what part to make is specified in other ways, through a manifest or a broadcast system.

Although Kanban traditionally started out as a tool for pull production in a stable, repetitive environment, efforts towards address its limitations and increase its applicability to other environments have been made. Junior and Godinho Filho (2010) review 32 publications on Kanban where comparison between traditional Kanban and later alterations are made, where differentiation from the traditional system as well as benefits achieved are highlighted. Functionality of Kanban in different environments can be derived from the pinpointed advantages identified by the authors (Junior and Godinho Filho, 2010). 19 different advantages are identified, out of which 8 hold interest for a MTO environment. Out of a total of 32 papers reviewed, 28 fill one or more of the advantages:

- Can be used effectively in environments with
 - Unstable demand
 - Variability in processing times
 - Complex material flow
 - High variety of items/products (mix)
 - Low machine reliability/high probability of breakdowns/halt in production
- Allows better coordination in production systems with assembly operations
- Gives more flexibility for changes in control periods

From Junior and Godinho Filho (2010)

Regenerative Kanban (Seidmann, 1988) is a technique that addresses unstable demand, variable processing times and complex material flow (Junior and Godinho Filho, 2010). These methods could show promise, however both literature availability and the tailored nature of these modifications suggests that, if they do not already start to look like similar, card-based

system, they either need additional tailoring to fit a more general environment, or is incomplete.

Kanban in job-shop have been attempted (Gravel and Price, 1988). Kanban is used to coordinate merging production flows in an environment with variable routing and shared resources. Conducted as a simulation study comparing an as-is system (FIFO assignment with no Kanban) with three various assignment techniques: Operation-weighted critical ratio, Shortest processing time, and Operation-weighted critical path. Assignment is in essence sequencing: which Kanban (production order) to start at what available machine. A Kanban card contains information regarding which operation to perform, which machine to use, what pieces required, the resulting part number as well as direction after completion. Stock accumulation at different machines are periodically revised by a production supervisor, who in turn fill out the decision board used. This means that release of work is periodic and controlled by the various sequencing rules described. Thürer et al. (2016a) argue that the semantics is wrong, and that the system simulated by Gravel and Price (1988) is different from Kanban in that the Kanban's are used to control production flow rather than capping inventory.

Generic Kanban does not feature direct product links (Bicheno and Holweg, 2016). The purpose of a generic Kanban system is to limit the number of jobs in the system, the WIP, in an attempt to modify Kanban to fit dynamic environments (Chang and Yih, 1994). A First-Come-First-Serve (FCFS) sequencing rule is utilized. Generic Kanban suits a dynamic environment where production follows received demand, that is, MTO. However, as the results of Chang and Yih (1994) rely on the simulation model utilized, generic Kanban seems suitable for a very specific type of HVLV environment. Chang and Yih (1994) describes the environment: the mix is low, only two different products, the production line is simple and directed, and the only dynamic parameter is demand fluctuation. The generic Kanban as described by Chang and Yih (1994) could be compared to single line CONWIP. If the number of Kanban's is fixed, the two systems both keep WIP at a constant level. The main difference is that CONWIP has one number to represent WIP throughout the entire production line, whereas generic Kanban sets a number between each process step. While CONWIP is simpler with only one input, generic Kanban is more flexible due to the fact that every Kanban number can be manipulated. This in turn makes the system more robust and can handle higher variability in processing time (Chang and Yih, 1994).

Gravel and Price (1988) show the use of Kanban without a signal from the end process that triggers demand upstream. In its place, a number of Kanban's and a Kanban-lot size is set, and simulated using three different sequencing rules, compared with no workload control and a de facto FIFO system. This is in essence capping work in process by the Kanban lot-size. As the definition of the system as a Kanban system is arguably wrong (Thürer et al., 2016a), it is in fact a CONWIP system that release a set of orders onto the shop floor when the shop floor clears.

3.2.1.1 Inputs

Kanban is a system that controls inventory (Thurer et al., 2016). In general, Kanban works as a two-bin system, using the reorder point (ROP) to derive the formula for number of Kanban cards (Bicheno and Holweg, 2016).

$$ROP = D \times LT \times SS$$

Where D equals demand during lead time (LT), and SS is the safety stock. N is the number of Kanban cards and Q is the container quantity, then the number of Kanban's is calculated as follows:

$$N = ROP/Q$$

Another approach, rather than using a safety stock, is to buffer with lead time, adding a factor of safety lead time (ST), which changes the formula:

$$N = (D \times (LT + ST))/Q$$

When the process have changeover, the calculation changes to include changeover, batch run time, queue time and delivery time in lead time, changing the above formulas. Batch run time, changeover time and queue time is referred to as Every Product Every (EPE), and is the interval in days of cycling a product. When changeover is included in a Kanban, the signal used is often a triangle.

$$LT = EPE + Delivery\ time + safety\ time$$

The problem of traditional Kanban is that, even for generic Kanban, it assumes repetitive production as well as a fairly level schedule. This is indirectly shown in the formulas: ROP is derived from demand, and the fact that lead time is such a deciding factor shows that variation in processing times and rework will make the final numbers regarding N, number of

Kanban's, at a high error margin. Parts included in a traditional Kanban should be use every day, while large and expensive parts should be excluded from the system. Further, the environment in which Kanban fits best is characterized by simple bill of materials (BOM), short lead times and small order quantities (Darlington et al., 2015). Darlington et al. (2015) characterize Kanban as a “stock based method”, which holds up towards the Kanban calculations mentioned. However it should be noted that it is possible in traditional Kanban to link the number of cards towards time rather than inventory.

Another issue traditional Kanban has over other PPC methods in HVLV, is that it is established between two stations, meaning that in order to have Kanban control over the entire system, Kanban needs to be established between each station in a directed way (Thürer et al., 2016a). Although Kanban do allow for some routing variability through the use of *multi-loop Kanban's*, a high mix or any customization needed prove difficult and in effective due to the fact that there has to be established multi-loop Kanban's to all routings that enters the system. *What Kanban can be used for in a HVLV environment, is to act as a replenishment line: Kanban systems naturally provide a mechanism for sharing a resource among different routings* (Thürer et al., 2016a). Further, products might have a change in order characteristics as it matures. To be aware of the state of each product, it is suggested to continuously look at the CODP (Amrani et al., 2010).

Darlington et al. (2015) reject Kanban in a HVLV environment due to the necessary super market creating prohibitive space and cost implications based on material cost, deep BOM (large number of parts), large number of engineering changes, and large variation in size and work content of the product in their case study.

3.2.1.1 Summary

What can be concluded for Kanban is that it cannot be used as a system for PPC in HVLV production. When addressing Production Planning and Control (PPC), the main factors controlling how an order flows through the value stream are: Order acceptance, sequencing, release and shop-floor dispatching (Thürer et al., 2016b). Kanban can be used as part of a bigger PPC system such as WLC, replenishing the main production line with high-running parts (Stevenson et al., 2005). In addition, as a product matures, the CODP can shift, and Kanban might prove effective as the mix lowers due to standardized and more predictable customer demand (Amrani et al., 2010).

What hinders Kanban from being a solution in HVLV environments is the need for product specific information. High mix cause high inventory if product specific Kanban is used. High mix cause an overly complex system if *multi-loop* Kanban is used. If generic Kanban is used, processing time variation cause either starvation at the control point, or accumulation of inventory. In addition, because Kanban keep information locally at the station that it controls, there is nothing balancing the workload elsewhere on the shop floor.

3.2.2 CONWIP

CONWIP is introduced as an alternative to Kanban in achieving pull production, introduced by Spearman et al. (1990). Described as a generalized Kanban (Spearman et al., 1990, Thurer et al., 2016) (not to be confused with generic Kanban systems (Bicheno and Holweg, 2016)), the main differentiation is that the card loop in a CONWIP system includes the entire production line. A card is attached to a container of parts at the beginning of the production line, and detached at the end of the line, allowing it to be attached to a new work order. Each container passing through the line that is included in the CONWIP system has roughly the same amount of “work”; that is, total processing time at the bottleneck. Whenever a new order enters the system, a *system entry time* is noted on the CONWIP card. Within the system, “*First come first served*” (*FCFS*) sequencing is used to ensure steady throughput in the system, with the exception of rework running through the system.

Thurer et al. (2016) describe CONWIP as a solution to *the low variability order control* problem. Because the goal of CONWIP is to ensure steady workload, every product included in the CONWIP system has to have approximately the same workload at the bottleneck (and have the same bottleneck). This is because every product will be represented by a CONWIP card. The number of cards set the level of work due to be released into the system at the same time. If work orders differ in workload, the CONWIP system will have varying levels of work (Stevenson et al., 2005). The purpose of CONWIP is described by Prakash and Chin (2015) as avoiding local and global WIP buildup.

A literature review on modified CONWIP (Prakash and Chin, 2015) show how different modification of CONWIP has received attention. Of 170 papers reviewed, 70 discuss original CONWIP. Notable modifications are (number of papers out of 170 in brackets): Segmented CONWIP (15), Closed-loop MTS CONWIP (10), M-CLOSED CONWIP (10), Dynamic dispatch rule CONWIP (20), Hybrid Kanban-CONWIP (15), CONLOAD (10).

Demand triggers the release of new orders, which in return pushes the order through the production system (Gastermann and Stopper, 2012, Gastermann et al., 2011). Gastermann and Stopper (2012) states that a CONWIP system requires certain inputs: Anticipation Horizon, the amount of work scheduled and released based on due dates; Capacity Trigger, the amount of work the production line can handle; Work-In-Process Cap, maximum amount of work the system is allowed to concurrently work on; and Dispatching and Processing Rule, the sequence in which work is conducted. Riezebos et al. (2009) argue that the main input for CONWIP is to set parameters per routing.

Matt et al. (2014) use CONWIP in an ETO construction environment in order to reduce overproduction which leads to increased time searching for components on-site, as well as increasing on-time delivery to site, consequently reducing downtime on-site and budget overruns. The CONWIP system adapted is a “first in system first served”, with the exception of rework which has highest priority. Similarly, Kjersem et al. (2015) investigate lean in ETO, where CONWIP is used in the production of the hull in the ship building process. In both of these cases, the CONWIP is used as a tool to release work through the entire process line used. Slomp et al. (2009) use CONWIP in an MTO environment in order to reduce WIP in what is described as a CONWIP/FIFO/takt time control system. Unlike Matt et al. (2014) and Kjersem et al. (2015), Stump and Badurdeen (2012) use many CONWIP lines, divided into sub-cells (fabrication, assembly and inspection CONWIP). CONWIP is used for both a workload limitation, as well as a tool to create steady flow of production.

A comparative simulation between CONWIP and POLCA is conducted by Frazee and Standridge (2016) with numbers based of a high-mix/low-volume environment. Two simulations is done: one with constant values regarding order arrival (demand), order size, and operation times, and one with random numbers for the values mentioned. The simulation include two products, each with four variations that go through a flow shop with twelve distinct processes. In both scenarios, the CONWIP solution comes out on top in terms of

Stump and Badurdeen (2012) describe CONWIP used in a MTO scenario building high-end boats, where the case company is described as an assembler. Currently, no PPC system is in place. The final product is based on a standard model that is customized to customer specifications out of an on-catalogue range of variations. The company does produce to stock, but 70 % is MTO. Manufacturing cells is divided into three main cells controlled by CONWIP (Fabrication, assembly, and inspection), in addition, a components cell supplies the main cells. Two main routings (Jumbo and Mini) is described. The main need before

implementation is to establish number of CONWIP cards, and sets the WIP cap. This is done using simulation, starting with a high number of cards. This is done because too few CONWIP cards will starve the production cells, causing loss of efficiency and increased lead time. Both WIP and lead times is reduced.

3.2.2.1 Inputs

CONWIP assign work to a card by accessing a *backlog*, or pre-shop queue. The first order on the list with raw materials and/or parts/sub-assemblies ready is released onto the shop floor (Spearman et al., 1990). The backlog is managed by production and inventory control staff: It can be generated from a master production schedule, as well as managed and rearranged. That is; there is no specific pre-shop release rules inherently present in CONWIP. The only pre-shop consideration enforced in a CONWIP system as described by Spearman et al. (1990) is that, under no circumstance should work be forced onto the shop floor without a card present. Spearman et al. (1990) establish four initial parameters necessary for a CONWIP system:

1. Card count, m . Determine maximum WIP in the system.
2. Production quota, q . Target production quantity for a period.
3. Maximum work ahead amount, n . That is, $q+n=\text{maximum production quantity}$
4. Capacity shortage trigger, r . A function that triggers capacity increase, such as overtime at the end of a production period t .

It is advised to set the parameters higher than necessary at first, and take a cautious approach, similarly to how Kanban implementation is advised. The operation of a CONWIP system should be regulated by its bottleneck resource. Comparing CONWIP to Kanban and push-systems, Spearman et al. (1990) measure WIP, finished goods inventory (FGI) and fraction of jobs that are late (service level). Prakash and Chin (2015) differentiate and categorize CONWIP solutions characteristics relating to: order release mechanism, aggregation of workload measure, workload control mechanism, workload control method, determination of card count, and dispatch rule.

In terms of routing characteristics, the inherent loop of a CONWIP system has three restrictions that may hinder applicability: 1. Work have to enter and exit the shop floor at determined points in order to control the system with a CONWIP loop; 2. Orders must visit the same stations, that is, the flow of the orders must not split; 3. The number of stations covered by a CONWIP loop should not be too long (Thürer et al., 2016a). The first two rules ensure that the load on each station on the shop floor have as stable workload as possible. If

the orders split, some machines may experience overloading, while others experience starvation. The way that this claim is made by Thurer et al. (2016), the first rule is redundant, as a rule that says all orders must visit the same stations in the same sequence implies that entry and exit point have to be the same for every order. The final point is stated because number of jobs and number of stations is in relation with each other. To avoid starvation, each station should at any time have at least one job. A system with 5 stations should use at least 5 cards, a system with 100 stations should use 100 jobs (Thurer et al., 2016). While this logic on paper seems on par, a closer look might tell a different story: if a system of 5 stations have 5 cards, each station would only have at least one job if the processing time of each station is the same. If station 1-4 have a processing time of 1 unit, and station 5 have a processing time of 10 units, the only station with full utilization will be the station with 5 units processing time. In addition, there will be propagation of inventory in front of the bottleneck. Logically, in said scenario, a released job will reach station 5 in 4 units of time, given that the lead time of station 1-4 is 4 units of time. This means that when station 5 finish a job, there has to be a new job readily available either from station 4, or as inventory. If two cards are used, there will be inventory of 1 job in front of station 5 idle in 6 units of time, which is better than inventory from 4 jobs idle for an average of 28 units of time.

The requirements presented by Thurer et al. (2016) is contradicted by Prakash and Chin (2015) who claim that CONWIP shows success in the semi-conductor industry job-shop MTO environment due to the systems high product-routing flexibility. The rules presented by Thurer et al. (2016) is also contradicted by Stevenson et al. (2005) (oddly enough, Stevenson have contributed to both sources). Stevenson et al. (2005) present CONWIP as a suitable solution in a general flow shop, where as Thurer et al. (2016), both implied by the rules presented and as a direct claim, argue that CONWIPs suitability is restricted to a pure flow shop only. In their own words: *CONWIP ... can only be really be applied in a pure flow shop (i.e. where all orders visits all stations in the same sequence).* (Thurer et al., 2016,p. 113)

An advantage of CONWIP is that, given that bottleneck processing time is fairly equal, the WIP is fairly constant (underlying the name CONWIP). This means that flow times are stable (Spearman et al., 1990). This has two benefits: in relation to important KPIs for HVLV production (Stevenson et al., 2005), it allows for predictable delivery date determination, given that the CONWIP is used for an entire system. If CONWIP is used for fragmented parts of the entire production line, it allows for stable and predictable feeds to an assembly.

In a CONWIP system, WIP naturally accumulates in front of the bottleneck (Darlington et al., 2015). This makes the system more robust to variation, and an attractive solution for a production system with maintenance problems, a shifting product mix, or any other issue that would cause the bottleneck location to change.

3.2.2.2 Summary

CONWIP is a PPC method that ensures a constant, limited level of work in process. This is done by finding the capacity of the bottleneck(s) in the production, and identifying the amount of work necessary at the bottleneck for each work order. Work orders with roughly equal work amount necessary at either the bottleneck or the entire system are grouped together in a production loop, using cards to release a new job whenever a job is finished.

3.2.3 Drum-Buffer-Rope

Drum-Buffer-Rope (DBR) is a bottleneck-oriented concept developed from Optimized Production Technology (OPT). DBR is the same as Theory of Constraints (TOC). DBR schedules production processes to run in accordance with the pace of the bottleneck(s) in the production system as a whole. The basis of DBR is the concept of throughput accounting (Goldratt and Cox, 2016, Wahlers and Cox III, 1994). Throughput accounting is based upon three measures: throughput, inventory, and operational expense. As its name suggests, DBR consists of three core mechanics: the “Drum”, the “Buffer” and the “Rope”. The Drum controls the pace of the system, and needs to be correctly identified as the systems bottleneck in order to maximize throughput. As DBR is a time-based PPC systems, the Drum is identified through time-based capacity planning, identifying the resources that are most utilized and where the most WIP is located. The Buffer is defined by ensuring that the Drum will have as high utilization as possible. This means that process times, breakdowns, and downtime have to be taken into consideration. The Buffer size is defined by work load measured in time (hoursworth, daysworth, etc.). The Rope as a consequence of the Drum definition includes all resources located upstream of the Drum. Its function is to signal the release of work when new work is started at the Drum, providing space in the Buffer inventory.

Gupta and Boyd (2011) suggest three levels of guidance for managing processes. The first level suggests that TOC will only have a significant impact on the system as a whole if it is focused on the constraint in the system. The second level suggests that, based on plant type (V-A-T classification), common issues can be predicted and attributed to either inventory or

capacity misallocation. The final level is the DBR system itself. This is based on Goldratt and Cox (2016), who suggest that TOC contains “five focusing steps”: Identify the constraint, exploit the constraints existing capacity, subordinate the rest of the system to the constraint before acquiring additional capacity, and return to the first step if the constraint is broken.

One study implementing DBR is a comprehensive case study over 24 months that target a multinational enterprise with multiple plants (Darlington et al., 2015). First, the most suitable plant for improvement is identified analyzing the financials of all plants and comparing where a WIP reduction would have the greatest impact. Value Stream Mapping (Rother and Shook, 2003) is used to map the system as a whole. Contrary to the VSM previously developed, which states that a single product or product family should be mapped. In this case, however, the team found it necessary to map all products running through the shop floor, as there was an abundance of shared resources. Mapping all operational steps present on the shop-floor, the consequent material and information flows of each individual product is mapped, separated by color-coded lines. Rather than the conventional VSM method, this highlights the shared resources, and gives any practitioners a visual representation of candidate processes to be set as the “drum” of a DBR system. In addition, this allows for routing variability to be visibly represented, aiding in the initial choice of appropriate PPC method. Darlington et al. (2015) develops a time-based capacity planning tool within an existing MRP system, producing a utilization analysis of all resources within the analyzed system, where demand (in hours) was compared to available capacity (in hours). The highest utilized resources were made Drum-candidates. To determine which of the candidates that was the actual bottleneck, a separate WIP monitoring tool was developed, which helped exclude some Drum-candidates. Finally, coefficient of variation (VAR) was measured for the lead time characteristics of the system as a whole. The final decision on Drum definition was made with the rule *when bottleneck feeds bottleneck it is best to deal with the first one first*. The identified and defined Drum was located early in the production process, and as a consequence, the Rope becomes short, and the system as a whole covers little of the shop-floor as a whole. The Buffer size is identified by simulating the defined system with the identified variations, such as lead time variations, downtime, and breakdown records. Contrary to common DBR definition, Darlington et al. (2015) have to implement a separate transportation buffer that buffered downstream process variation.

In a pure-job shop, it is argued that DBR is better suited than CONWIP (Spearman et al., 1990). Where CONWIP control WIP, DBR control release rates. DBR allows greater

variation of lead time at the bottleneck, as the drum releases work based on processed work at the bottleneck rather than releasing jobs altogether. It also allows for greater routing variability, given that the bottleneck remains the same. Darlington et al. (2015) argue that DBR is a functional strategy in a general job shop, and compares and shows superiority over the CONWIP system. This coincides with Stevenson et al. (2005), who also argue that DBR (or TOC) is superior to CONWIP in a general job-shop, but argues WLC to be the better option all together. Darlington et al. (2015) choose DBR over CONWIP in their case study due to three factors: complexity of the routings, amount of work content, and lead time variation.

DBR comes with some issues, however. Its “buffer” is inventory stored in front of the “drum” to ensure that the bottleneck always is working. This means jobs have to be released in such a way that it fills the buffer whenever new work enters the bottleneck workstation. To ensure this, the “rope”, the sequences of work needed for a job before the bottleneck work station have to be fairly predictable. In addition, for the shop-floor to have only one buffer, work time per workstation both prior and after the bottleneck must be scheduled in such a way that work does not accumulate as waiting inventory.

3.2.3.1 Inputs

DBR schedules based on the pace of the bottleneck. Because only one source of workload is necessary to calculate in order to buffer the system, the issue faced in HVLV production regarding uncertainty is decreased due to less sources of error being present in the system design (Stevenson et al., 2005). The main issue faced regarding DBR is that there is an assumption that the bottleneck(s) are fixed. In terms of shop configuration, this can be related to flow shops rather than job shops. HVLV production with a general flow shop configuration is more likely to benefit from DBR than general or pure job shops.

An argument against DBR is that it does not include the customer enquiry stage and the job release stage (Stevenson et al., 2005). The Rope does tell the system *when* to release, but not *what*, meaning that a release rule have to be specified in addition to the DBR system.

3.2.3.2 Summary

Similarly to CONWIP, DBR seems most fitting for a general flow shop, rather than a general job shop. In addition, there seems to be a fair bit of overlapping between multiple solutions. The problems that occur with either CONWIP or DBR when the systems are too complex

(routing variability, process time variability, bottleneck variability) seems to facilitate “fixes” that in practice turn the CONWIP or DBR system into some sort of WLC.

3.2.4 Paired-Cell Overlapping Loops of Cards with Authorization

Paired-Cell Overlapping Loops of Cards with Authorization (POLCA), is a hybrid push-pull system designed for HVLV in order to cope with complexity and dynamism in order to reduce throughput times (Shi-chao et al., 2012, Krishnamurthy and Suri, 2009) POLCA is as its name suggests a card-based system, in which a card signals downstream capacity availability.

Prerequisites for implementing POLCA is to have a High-Level Material Requirements Planning (HL/MRP) system or any system that can obtain rough-cut capacity and lead time estimates, as well as cellular manufacturing (Krishnamurthy and Suri, 2009). The selection of job is done by reviewing authorized jobs and using High-Level Materials Planning Requirements (HL/MRP). A HL/MRP system does not work on an operations level, rather, it helps planning flow between cells. That is, material movement between cells is controlled by the POLCA system, the mix is controlled by the HL/MRP. Furthermore, because the POLCA system controls flow between cells, material flow within the cells themselves need different control strategies, and in some instances, Kanban can be viable (Krishnamurthy and Suri, 2009, Stump and Badurdeen, 2012). The main problem of planning cellular manufacturing is synchronization of processes between the cells (Riezebos, 2010).

POLCA is a system which use MRP planning in its functionality. The MRP system calculates an earliest release date, that is, the earliest expected date in which the order is confirmed and materials is available (Thürer et al., 2014b). This is the same as a standard MRP system. The difference is that the POLCA system puts a restriction on the routing of the order. Whereas the MRP system push the order, that is, it is released at the moment of the earliest release date, a POLCA system release the order only after there is available capacity within the routing of the order (Thurer et al., 2016). The loops of a POLCA system works as follows (simple example): Given three stations, A, B, C, a pair of stations represents pairs that are included in a products routing. For a product with routing A-C-B, there are two cards: AC and CB. When earliest release date is realized at station A, product will begin, given that there is an AC card available. This card is assigned to the product until it has finished production at both station A and station C. Production at station C will not start until a card CB is available. Production at station B will start at the earliest release date without a new card because it is the end of the products routing. Table 3 shows an example of a POLCA release list. Note that

this is the release list for cell A. Because there is only one required POLCA, this is the first cell in the various routings of the order numbers in the table.

Table 3: POLCA release list (Riezebos, 2010)

Release list cell A (released on 26-3-2008 (dd-m-yyyy))							
Order number	Material check	Earliest start date	Route in cell A (proc. time)			Next cell	Required Polca's
			1	2	3		
CL002	not ok	21-3-2008	M1 (10)	M5 (20)	M8 (10)	B	1 A/B
CL003	ok	24-3-2008	M2 (30)	M1 (40)	M3 (50)	B	1 A/B
CK001	ok	25-3-2008	M3 (10)	M1 (60)		C	1 A/C
ST253	ok	26-3-2008	M1 (110)	M2 (140)		B	1 A/B
CK005	not ok	26-3-2008	M4 (20)	M1 (40)	M5 (70)	C	1 A/C
CL015	ok	26-3-2008	M2 (40)	M3 (60)	M5 (60)	B	1 A/B
HY563	ok	28-3-2008	M3 (100)	M8 (120)	M2 (10)	B	1 A/B
ST237	ok	29-3-2008	M1 (180)	M2 (120)	M5 (120)	B	1 A/B

POLCA implementation is a four stage implementation process (Krishnamurthy and Suri, 2009): Pre-POLCA assessment; design of the POLCA system; launch of the POLCA implementation; post-implementation evaluation. Assessment is to check if the cells have capacity to be able to cope with the required throughput. Design requires routing identification, release authorizations, workload values, documentation of the POLCA procedure, and computing the number of POLCA cards in each loop. At the launch of POLCA, training and education, as well as frequent reviews is needed. After implementation, key metrics needs to be tracked, and qualitative benefits measured.

Riezebos et al. (2009) argue that the main input for POLCA is to set parameters per combination of two cells. Later, Riezebos (2010) argue that POLCA use three categories of tools and methods to facilitate the improvement of throughput time control. These categories are routing, release and facilities: Routing, rout specific cards attached to jobs; Release, list jobs by earliest starting times; Facilities, enable operating the system in various circumstances. Where Kanban is described as product specific, CONWIP is described as product anonymous. Riezebos (2010) state that POLCA should be described as route-specific. Case studies show promise of POLCA implementation when flow is directed (Krishnamurthy and Suri, 2009) for both pure flow shop (PFS) (products all go through the same routing) and general flow shop (GFS) (products go through a variation of routings, but flow is unidirectional). POLCA is inefficient when routing variability is low (Stump and Badurdeen, 2012). Managing a POLCA-system is more difficult than managing CONWIP, and the authors chose in its place to implement CONWIP-loops rather than POLCA.

3.2.4.1 Inputs

The POLCA system has three core elements: order confirmation and material availability from the MRP system, and availability of capacity. Earliest release date (MRP) have to be correct in order to prevent starvation or work on the wrong orders. This occurs if the date is determined early or late respectively (Thurer et al., 2016). The capacity is determined available from two variables: routing (card pair) and card count (Riezebos, 2010). Card count is based on lead time from both stations, throughput rate in the specific card pair, waiting time per card pair, and travel time both to and from the receiving station. Krishnamurthy and Suri (2009) compute the number of cards per pair of cells by the formula:

$$N_{i/j} = [LT_i + LT_j] \times NUM_{i/j}/D$$

LT is the lead time of each cell. NUM is the number of jobs that go from cell i to cell j . D is the duration of the planning period. All durations measured in days. Riezebos (2010) build upon this formula to include waiting times both in front of the first cell in the routing, but also the next step, as the hold-up in front of each cell contributes to delay of production. The need for average lead times is also mentioned, but is rhetorically implied by Krishnamurthy and Suri (2009). Further, Riezebos (2010) include the travel time between cells. Lastly, it is mentioned that waiting and travelling time can be excluded and exchanged with an estimated safety allowance, giving the formula:

$$N_{i/j} = [LT_i + LT_j] \times (1 + \alpha) \times \lambda$$

Where α represents waiting and travel time divided by lead time, λ represents throughput rate of jobs (Riezebos, 2010), same as $NUM_{i/j}/D$ (Krishnamurthy and Suri, 2009).

Routing is order specific, but in order to have an estimate of the card count, the routing have to be set before release of the order onto the shop-floor.

3.2.4.2 Summary

POLCA seemingly fits very well with HVLV manufacturing. The only restriction is set on routing, with release being based on MRP logic paired with available capacity.

POLCA does have two main drawbacks: it is dependent on a MRP system, which have shown itself difficult to use effectively in many MTO companies, especially SMEs (Stevenson et al., 2005). In addition, due to its paired-cell nature, the routings have to be unidirectional.

3.2.5 Workload Control

Workload control is a method of controlling total WIP throughout a production cell through order release methods, as well as flow of work between stages of the production cell through shop floor dispatching rules (Mortágua et al., 2014). This is done by holding back accepted jobs, allowing for transparent shop floor situation that allows for faster feedback regarding job status (Oosterman et al., 2000). Three levels of control are considered: order entry level, the acceptance of a job; order release level, the release of a job; priority dispatching, the selection of a job in front of a work station at the shop floor (Marangoni et al., 2013). Lead time in WLC is depicted in Figure 1. Stevenson (2006) includes a third level of control, namely the customer enquiry stage used in the LUMS/LUMS COR order release system. There are four terms used to address each operation step which the WLC considers: station, stage, work center and operation. These will throughout the discussion of WLC be used interchangeably. Stated by Thürer et al. (2012), Workload Control (WLC) protects throughput from variance, the key to achieving lean. WLC is argued by Mortágua et al. (2014) to be the most promising production planning and control (PPC) method in HVLV environments, especially for small-medium sized enterprises (SME). Bertolini et al. (2015) agree with this statement, claiming that WLC is among the pull systems showing the most promise in MTO-HVLV due to its handling both shifting bottlenecks as well as routing with very high variability. WLC share the characteristic that work is not released until WIP is below a set threshold, however, contrary to CONWIP, WLC have a fluid value for work content; the contribution a job has to the overall workload is total throughput time divided by routing position, that is, at which production stage the product currently possess. Simply, this means that WLC controls workload level, while CONWIP controls number of jobs currently active in the system. Because of this, there is no need to convert a production order into workload content (Riezebos, 2010). WLC inspired card-based PPC is suggested by Thürer et al. (2015) in the form of “Control of Balance by Card Based Navigation” (COBACABANA).

The concept of WLC with periodic release is well described by Oosterman et al. (2000). The claim is that the goal of WLC is to control direct load at a low and stable level. That is, to control the workload at each stage of production. Workload calculations is described, through three equations, summarizing three concepts that measures direct load, aggregated load and shop load as the deciding factor for further release of jobs onto the shop floor. The argument is that measuring direct load is working in a pure job shop, because the work will reach the measured stage shortly after release. However, if the flow is directed, there will be a time lag,

because the job needs to pass through stages upstream before reaching the stage where direct load is measured. Aggregated load measures the direct load queueing in front of a station, and in addition considers the load contribution of all work released that currently sits upstream. Shop load measures aggregated load, and adds the workload still to be finished downstream. The calculations is based on a method called load conversion from the theory of Wiendahl (1995), chapter 6.

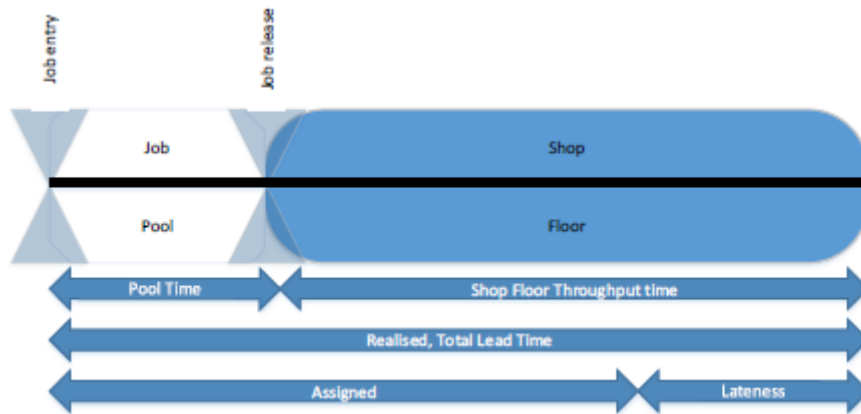


Figure 1: Lead Time WLC (Oosterman et al., 2000)

Thürer et al. (2012) investigates release methods for WLC, in which CONWIP is included as a release method. Comparing continuous release methods and combined continuous and periodic release methods, it is concluded that continuous release methods outperform all forms of pure periodic release methods. This adds to the research by Oosterman et al. (2000), as it investigates continuous release, hybrid (continuous and periodic) and pure periodic release. Pure CONWIP as a release method is outperformed by other methods. The main focus of Thürer et al. (2012) is to investigate how the release method “Lancaster University Management School Corrected Order Release” (LUMS COR), a hybrid periodic/continuous release method compare to other release methods. LUMS COR periodically releases work onto the shop floor from the pre-shop pool by comparing the workload a job contributes to the work centers in its routing against the workload limits set (Thürer et al., 2014). The periodic release method is aggregated load released up to the upper bound workload limit, the continuous methods are “Superfluous load avoidance release” (SLAR) and “Work center workload trigger planned release date” (WCPRD). SLAR releases job(s) if any work center is starving, that is, no urgent jobs queueing in front of the work center. WCPRD release a job if the direct load in front of any work center falls below lower bound selected threshold. The conclusion is that WLC will improve performance, even if dispatching such as Planned

Operation Starting Time (PST) is in place. To further enhance WLC performance, Thürer et al. (2014) investigates how Customer Enquiry Management (CEM) impacts the effects of WLC. CEM is used to set more precise due dates, which increases the precision of pool selection rules and shop-floor dispatching rules that include due-dates in their job selection, such as PRT, pre-shop pool selection (Mortágua et al., 2014) or PST, shop-floor dispatching (Thürer et al., 2012).

Mortágua et al. (2014) simulate how different release methods function in flow shops with unidirectional flow. The simulated approaches are workload bounding and workload balancing. Workload bounding releases work from the pre-shop pool if production levels are within an upper limit in all production stages. Workload balancing releases work from the pre-shop pool if the release of work improves the balance of work through the production stages, and allows for levels above selected threshold at up to 50% of the production stages. Rules are set for selecting jobs from the pre-shop pool, Mortágua et al. (2014) tested three: earliest planned release time, shortest processing time, largest total work content. In the scenario simulated by Mortágua et al. (2014) represent a hypothetical six-stage MTO production system in two shop configurations, pure flow shop (PFS), and general flow shop (GFS). The difference between PFS and GFS is well described by Oosterman et al. (2000): The routing in PFS includes all stages, the routing in GFS includes a variable number of stages, but always in an ascending order, meaning that the flow is unidirectional. Job sequencing at each machine is done by giving priority to the job that is most urgent to avoid variance in lateness across jobs. This is done by focusing on throughput time, due date, and remaining routing for the product. The results show that pool selection rule impact the performance of WLC in the GFS has little impact, and that workload balancing shows slightly better results on both tardy jobs and standard deviation of lateness on jobs when throughput time is short. In a PFS, pool selection rule shows much higher impact on results, especially if the release method is workload bounding. As a result, workload balancing should be chosen as release method in a PFS, as it is more robust.

An effort combining WLC with MRP was done in the case study by Melchert et al. (2006). 2600 stock keeping units (SKU) are available on catalogue. Each with different product and packaging characteristics (routing). Workers and machines are specialized, and lack cross training, and the current PPC is an MRP software. The stated goal of improvement is to reduce lead time and operational cost. The complexity of the routings was addressed by cross training workers and replacing specialized machines with more flexible, simplified work

stations. Set-up procedures of was implemented to reduce set-up time. Two modules was implemented: a worker allocation module taking advantage of a cross-trained workforce, reallocating workers to work centers scheduled for work, and highlighting capacity and efficiency rates. A customer enquiry model was implemented next, allowing delivery date estimation based on expected pool delay and queuing time at the shop floor represented by the assigned time line in Figure 2.

3.2.5.1 Inputs

Workload Control is the most input-heavy of all the PPC's discussed in this thesis. In order to be utilized to its fullest potential, WLC needs workload to be measured continuously, or at least periodically (based on release method) in order to control the workload on the shop floor. Oosterman et al. (2000) show three configurations (Figure 2).

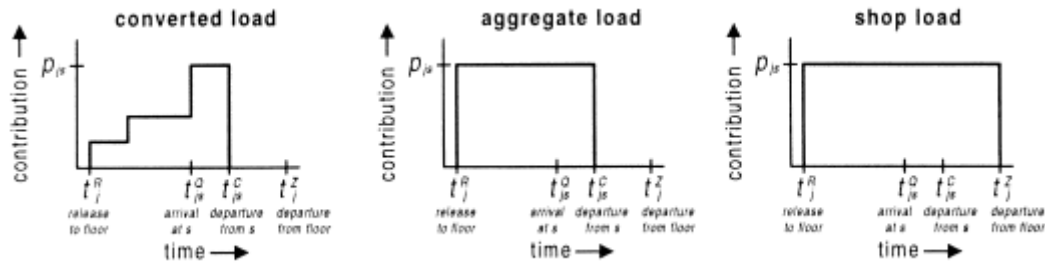


Figure 2: Workload contribution of job j across time

The workload measured is then used to release work onto the shop-floor based on release rule. A number of release rules are described by Thürer et al. (2012b). As can be seen in table 4, other than with CONWIP, a workload norm is set, and work (jobs) are released onto the shop floor when the rule dictates release.

Table 4: Release and dispatching rules (Thürer et al., 2012b)

Abbreviation	Full name of rule	Classification	Brief description
ConWIP	Constant Work-In-Process	Continuous order release method	Releases jobs if the number of jobs on the shop floor falls below a predetermined level (lower bound)
LUMS COR	Lancaster University Management School (LUMS) Corrected Order Release	Periodic and continuous order release method	Combines periodic with continuous release. Jobs are pulled onto the shop floor in between periodic reviews if a work center is starving
Periodic	Corrected aggregate load	Periodic order release method	Releases jobs periodically up to the workload norm (upper bound)
SLAR	Superfluous load avoidance release	Continuous order release method	Releases jobs if a work center is starving or there are no urgent jobs queuing in front of a work center
WCPRD	Work center workload trigger planned release date	Continuous order release method	Releases jobs if the direct load of any work center falls below a predetermined level (lower bound)
FCFS	First-come-first-served	Shop floor dispatching rule	The job which arrived at the work center first is chosen from the queue
PST	Planned operation start time	Shop floor dispatching rule	The job with the earliest planned start time at a particular work center is chosen from the queue

In addition to having more complex release rules than all other PPC's, the detailed overview required for (some of) the rules allow for more sophisticated shop-floor sequencing rules (Thürer et al., 2016b).

The Workload Control Paradox is a problem faced when comparing empirical and theoretical information surrounding WLC. Simulations of WLC practices often suggest very small benefits regarding lead time, or even an increase in lead time. Empirical reports, however, report on reduction in throughput times at 40-50% (Stevenson et al., 2005).

3.2.5.2 Summary

Workload Control is the solution most often cited as optimal for the job shop, being purposefully designed with that environment in mind. Depending on how it is defined, WLC requires more information than other PPC's, however, as can be seen in table 4, CONWIP remains a valid release rule within the WLC terminology.

3.3 Sequencing rules

Sequencing rules are rules that determine the order of any queue in front of any machine/processing step. Any system with a mix of more than one that share resources run the risk of multiple different work orders queuing in front of any step. Even orders containing the production of the same product could have different characteristics in terms of due date. As with release rules, the most common sequencing rule is FIFO.

Sequencing rules are primarily used in CONWIP and WLC, where the possibilities of routings include lack of direction as well as variation. Gravel and Price (1988) apply sequencing rules in a Kanban system, but as Thürer et al. (2016a) argues, it does not carry the characteristics of a Kanban system. The system applied by Gravel and Price (1988) is arguably a CONWIP system with periodic release.

List of sequencing rules:

Thürer et al. (2016b)

- First-Come-First-Serve (FCFS)
- Earliest Due Date (EDD)
- Planned Release Date (PRD)
- Shortest Processing Time (SPT)
- Lowest Workload Imbalance (LWIB)

- Capacity Slack Corrected (CSCOR)
- Modified Capacity Slack (MCS)

Gravel and Price (1988)

- The Operation-Weighted Critical Ratio
- The Shortest Processing Time
- The Operation-Weighted Critical Path

Gravel and Price (1988) use Kanban as a means to coordinate assembly operations. A simulation of the three rules show that all rules outperform FIFO as a shop-floor sequencing rule. In the case study however, The Shortest Processing Time was chosen due to a lack of computing power available at the case company. The way the Kanban works in this scenario, is that a centralized board displays all available Kanban's. On this board, all waiting operations are displayed sequentially from top to bottom, sorted by their processing time. An operator selects the first operation where both a Kanban as well as all stock tickets are available.

The Operation-Weighted Critical Ratio

Operations present on the routings that are most likely to be congested is assigned to free a free machine (processing step). The order assigned to any free process step is the order on a path i whose next task is on the free processing step, and with the maximum value $Max_{(m,i)}$ for the following ratio:

$$K_{(m,i)}t_{(m,i)} / T_m$$

$K_{(m,i)}$ is the numbers of operations remaining that use machine m on path (or routing) i .

$t_{(m,i)}$ is the remaining machine time required of machines of type m on path i

T_m is the total time required of all paths of machine(s) type m .

This ratio is calculated for all machines (or processing steps), not only those who are free. This is to get an overview of future congestions (queuing).

The Shortest Processing Time

Any free machine (processing step) is assigned the waiting operation with the shortest processing time.

The Operation-Weighted Critical Path

The operation on the routing with the highest operation-weighted factor are assigned to a free machine (processing step). The values of $K_{(m,i)}$ and $t_{(m,i)}$ are the same as for the critical ratio.

$$K_{(m,i)}t_{(m,i)}$$

3.4 Summary

The general idea regardless of what pull system to implement is to start “loose” and tighten up production as experience grows. That is, to start with a generous amount of allowed work, and reduce over time (Bicheno and Holweg, 2016).

There exists a number of attempts to adapt the various PPC systems to fit with more complex environments, and moving the CODP upstream, in essence moving the PPC methods from MTS to MTO (or from low variety/high volume to HVLV). The description and functionality of many of these solutions blur the difference between different methods, such as Kanban and CONWIP (Chang and Yih, 1994), CONWIP and DBR (Spearman et al., 1990), and CONWIP/DBR and WLC. This highlights the need to answer the main questions of this thesis: how is work released, how is the parameters upon which release is instigated measured, and how does work flow on the shop floor. The functionality of one method might have favorable characteristics in a certain manufacturing environment, but lack support for load balancing (Thürer et al., 2016a), fail to address all PPC levels, and be too costly (Stevenson et al., 2005).

CONWIP is a PPC that can handle routing variability, but there has to be lead time stability. DBR can handle lead time instability, but is more restricted by routing, as all routings need to share bottleneck in order for it to be effective. If either one characteristic exists without the other, an approach of CONWIP or DBR is in order. If, in varying degree, both lead time and routings have significant variability, an adaptation of either CONWIP or DBR the push the system towards WLC must be made, or a WLC system should be the first to be implemented. Both in pre-shop and shop-floor sequencing, FCFS sequencing is often used. It should be noted that this sequencing rule is the same as “first-in-first-out” (FIFO).

There is similarities between CONWIP, DBR, and WLC. Both CONWIP and DBR turn into WLC systems if there are “fixes” needed in complex systems. If the systems show stability over time, any WLC system implemented very well turn into a CONWIP or DBR system in practice. Because CONWIP and DBR are simpler than WLC, these systems are easier to

implement first if there are many unknowns in the system with the idea of “it is better to do something than nothing”. When investigating any issues that occur, the PPC system could be made more complex, turning a CONWIP or DBR system into a WLC system with greater care for workload in each process step, with focus on pre-shop release.

4. Discussion

As can be seen in chapter 3, there are similarities between the different PPC methods. This chapter will discuss how the PPCs can be differentiated, and which ones show the highest level of applicability given the environment in which it is implemented. Similarly to VSM implementation in HVLV described by Strøm (2017), PPC implementation in MTO show many of the same characteristics. A literature review that suggest a method will, regardless of the seemingly perfect fit between PPC method and environmental factors, need specific tailoring to that specific environment (Brink and Ballard, 2005).

4.1 Defining the MTO industry

As seen in the literature, the definitions of MTO and ETO, and HVLV differ somewhat. As the environments from various case studies show, even those companies that semantically can be categorized under the same definitions, have vastly different production characteristics (Stevenson et al., 2005, Rudberg and Wikner, 2004, Olhager and Prajogo, 2012, Krishnamurthy and Suri, 2009). Defining both ETO and HVLV under the category MTO helps to create a clear picture, because, rather than defining manufacturing characteristics based on CODP or the rather vague terms “Variety” and “Volume”, this forces the review of PPC methods to be characterized by the actual manufacturing performance and system characteristic in which they are to be applied, rather than to have its applicability generically be claimed to fit their CODP or Variety/Volume category. With the general term for MTO in mind, the characteristics should then be defined. By identifying the critical contingency factors for system selection, each system can be defined to a specific MTO scenario which show more accuracy rather than generic MTO/ETO/HVLV terminology.

With this definition of the MTO industry, the characteristics described in table 2 become important in order to highlight the contingency factors for choosing a PPC system. The CODP, or the “Variety” or “Volume” from the HVLV terminology becomes less relevant. What is important is that the differentiation between manufacturing environments can be attributed to specific factors.

4.1.1 Criteria for determining applicability

As we saw in chapter 3 regarding applicability, Stevenson et al. (2005) proposed five criteria for applicability. Again, some of these criteria are rather vague. Ability to cope with customized products, and ability to cope with variable routings is applicable for most of the suggested solutions, only excluding Kanban as an overall shop-floor control method. When

paired with the environmental factors proposed by Birkie and Trucco (2016a), as well as the other identified factors (Strøm, 2017), further determination for applicability can be identified.

For PPC in the MTO industry, and as a basis for PPC selection, the primary goals of a PPC as described by Stevenson et al. (2005) is used. Reduced WIP, minimized throughput times and lead times, and improve delivery date adherence are all goals the PPCs discussed in this paper aim to achieve. The problem with previous definitions however, is that terms like “variability” and “variety” (HVLV) can describe both complexity and dynamism as described by (Birkie and Trucco, 2016a). Variability in routings for instance might indicate a high product mix, where the routings are known and fixed, but a high number of products run to the same system. In this case, if the terminology described by Birkie and Trucco (2016a) is used, the complexity is high. If there on the other hand is a higher degree of innovation and/or reengineering, then the routings, which would still be described as variable, become unknown. This would be described by Birkie and Trucco (2016a) as dynamism. This distinction have a direct correlation to the choice of PPC system.

4.2 Production Planning and Control

Pure pull production in MTO manufacturing is impossible. Of course, the nature of MTO ensures that any order on the shop floor is produced by the pull principle, however, on the shop floor there has to be push functions. Full utilization of any station relies on flexibility, due to the fact that the products going through the shop floor have different characteristics. If stations are rigid, there will never be full utilization on any machine other than the bottleneck (if there is one). In this case, the goal is not to strive for full utilization, rather, it is important to ensure that the bottleneck(s) never starve.

4.2.1 Kanban

Kanban as the main manufacturing system in MTO is not applicable (Stevenson et al., 2005), as control of individual products in the manufacturing system leads to accumulation of inventory and WIP. This is caused by two factors: routing variation and processing time variability which represents both internal complexity and dynamism. Thürer et al. (2016a) explain that the common way to overcome routing variability problems relating to Kanban is to increase the number of Kanban cards. This is disproved as a solution (Thürer et al., 2015) because the cause of an increase in Kanban’s working is that it creates an inventory buffer.

However, in conjunction with a higher level PPC, Kanban control can be used on the shop floor. This is further described by Amrani et al. (2010) who explains that over time, the environmental factors change for different parts, allowing for more predictable demand of parts that are used in a variety of end products. This in turn allows for a MTS of the parts which in turn is an ATO strategy for said parts. In essence, this is what VSMM is, where Bertolini and Romagnoli (2013) and Kjersem et al. (2015) apply techniques to build a lean value stream in an ETO environment, where Bertolini and Romagnoli (2013) find reoccurring product families, and Kjersem et al. (2015) finds part of the value stream where the product differences are small enough to allow for a simple, unidirectional flow. These are examples where Kanban could find use.

4.2.2 CONWIP

In many ways, CONWIP can be seen as a simplified Kanban, and given the size of the production cell in which it is applied, can function in all scenarios. CONWIP is in essence as complex as the inputs it is given, with the overlying aspect of setting a cap on WIP.

CONWIP used in ETO is present in the cases of Kjersem et al. (2015) and Matt et al. (2014). Where Kjersem et al. (2015) describe the hull production in ship manufacturer, Matt et al. (2014) use CONWIP in the production of aluminum frames used in construction. In both cases, the complexity of the products is low. The fact that each project requires one-of-a-kind engineered parts causes the product diversity to be high, however the mix share a similar routing, going through the same processing steps (CI 1-3, table 2). Slomp et al. (2009) do the same thing when implementing CONWIP in MTO, namely applying a lean production system on a part of production (here the production of conductors for low- medium switchgear systems) that has the same processing steps, the production of copper strips. Again, each order will have customized specifications, but the routing is the same. This allows for CONWIP to be applicable.

In the case of Stump and Badurdeen (2012), CONWIP is used in a MTO scenario of low internal complexity and dynamism. CONWIP loops is used on three manufacturing cells (fabrication, assembly and inspection), where each cell internally utilizes push production.

The conceptual description of a CONWIP planning software as suggested by Gastermann and Stopper (2012) seemingly contradicts the conclusion of Stevenson et al. (2005) in that CONWIP is used in an environment where the environmental factors of Birkie and Trucco (2016a) (table 2) suggest WLC as the optimal solution, backed up by (Mortágua et al., 2014)

and Thürier et al. (2012a). Gastermann and Stopper (2012) argue that for a small and medium sized enterprise (SME) with limited capacity and capabilities, implementation of more sophisticated software might create more problems than they solve. However, Gastermann and Stopper (2012) concludes that CONWIP for SME is a very suitable solution given *low WIP*. The latter statement questions whether it is the CONWIP solution they apply that contribute to the improvements, or that the shop-floor dispatching rules is the actual beneficial implementation.

CONWIP remains a simple solution, but it is very sensitive to routing variability in the form of complexity in the manufacturing environment. In order to prevent both starvation at resources as well as congestion at bottlenecks, a CONWIP would need a pre-shop sequencing tool, and a shop-floor sequencing tool to make sure that orders with conflicted routings do not arrive at the same resources at the same time. This would push the CONWIP towards a WLC system, which is what the literature review of Prakash and Chin (2015), although not stated, hints at. The modifications directly address workload measure, workload control method, and dispatch rule, all characteristics that are missing from the original CONWIP, but exists in WLC. This again shows how, when modifying CONWIP, it will in essence resemble a WLC system. Zäpfel and Missbauer (1993) argue that Kanban and CONWIP fits in simpler environments, whereas WLC fit in more complex environments, which is backed up by the previous examples.

4.2.3 DBR

DBR and CONWIP share many characteristics, especially if the CONWIP is designed with workload balancing at the bottleneck. As a general solution, DBR requires more specifics compared to CONWIP, because it is in essence bottleneck control. This means that there have to be a bottleneck, it needs to be stationary, and the more possible bottlenecks, the increasingly difficult will the DBR system be to implement. What DBR lose by requiring specifics, it gains by requiring less information. As the inclusion of the customer enquiry stage is important in due date setting and adherence, the specific requirement of the DBR have less margin of error compared to other methods. As dynamism increase, given that the bottleneck remains the same, the need to “guesstimate” lead times are restricted to one resource. Less variables means less variation, and the predictions possible to make regarding due dates will increase in precision. Spearman et al. (1990) show that DBR is sensitive to errors due to requiring decisions continuously being made about release times, whereas CONWIP release automatically when an order is finished.

4.2.4 POLCA

Even though CONWIP can be applied in all situations, a more specific approach is POLCA. This is when the routing variation is high (Stump and Badurdeen, 2012). Specified in context of Birkie and Trucco (2016b), that is when CI 2-3 *Table 2* is high. Even though the implementation of CONWIP loops as suggested by Stump and Badurdeen (2012) share similarities with a POLCA system, it is argued that due to the low routing variation, a POLCA system will not outperform a CONWIP system, and that the high dynamism levels cause the POLCA system to need to many inputs, further increasing implementation difficulties.

Workload and routing predictability is a necessity for POLCA to be viable. POLCA loops in between stations determine availability and release of work. Because of the design of the POLCA system, orders released is based on their routing and workload. Number of POLCA cards is determined by lead time and number of jobs that go between cells. This is determined over the set planning period. Unaccounted practical issues means that the standard POLCA practice is to round the calculated number of POLCA cards up. In addition, “safety cards” is introduced to allow for cells to be used if a certain product is halted in a cell due to lack of supply, quality defects or due date postponement. This shows that POLCA is very responsive to change, and is based on knowledge influenced by dynamism. As such, and much like Stump and Badurdeen (2012) claims, variability or instability (dynamism, *Table 2*) cause a POLCA system to be very complex and lack precision. Similarly, in a comparison between CONWIP and POLCA, Frazee and Standridge (2016) show that both in a high complexity/low dynamism and a high complexity/high dynamism environment, CONWIP outperforms POLCA in that WIP is held at a lower level given the same lead time and consequently throughput. Questions raised regarding the CONWIP and POLCA design in the simulation arise, but even if there could be false results due to faulty calculations, it shows that CONWIP is easier to use. The results in the low-dynamism environment show CONWIP and POLCA performing similarly, whereas in the high-dynamism environment, CONWIP outperforms POLCA with good margin. Still, there could be accumulation of WIP in a production cell in a CONWIP system that a POLCA system could handle due to the hold-up of work if a cell is at a halt (Frazee and Standridge, 2016).

Krishnamurthy and Suri (2009) shows three implementations of POLCA through case studies. Two of which is characterized as expected with high levels of complexity, and low levels of dynamism. The latter, an ETO company shows higher level of dynamism. The solution in POLCA implementation was however to implement the POLCA techniques to routings of

products with less dynamism. As with all MTO production described in this thesis, demand is unpredictable and unstable, but no other dynamism values are identified. As such, all three cases described by Krishnamurthy and Suri (2009) validate the statement of POLCA being a fit when the environmental factors show high complexity and low dynamism.

Furthermore, because of POLCA and the amount of inputs needed, rate of innovation, (DI 2, *Table 2*) depending of the level of change each innovation would add to the shop-floor, would introduce new product routings, further proving POLCAs lack of handling environmental factors that increase dynamism.

4.2.5 WLC

As the WLC scenario described by Mortágua et al. (2014) is a simulated, hypothetical MTO environments, the environmental factors present are hypothetical. External complexity is assumed low (CE 1, 3 table 2). In addition, all internal dynamism is assumed low based on the fact that throughput time and work content is known (DI 1-3, table 2). The authors targets with their WLC approach, is internal complexity; even though flow is unidirectional, all stages of the process is assumed equally likely to be next process stage. As such, both process interdependencies and variety of interactions is high (CI 2-3 table 2). In addition, external dynamism is assumed low, as supply of material is not considered (DE 4, table 2). The conclusion made by the authors also seems to confirm the “workload control paradox”, where a simulation show sub-par results (Stevenson et al., 2005). Mortágua et al. (2014) get simulation results that suggest that WLC will reduce WIP, but at the same time reduce due date adherence. Seeing as due date adherence is a key performance index for MTO companies (Stevenson et al., 2005), this suggests that WLC is a sub-par solution. This result seems to have to be accompanied by case-studies, as empirical reports show that WLC perform well.

Simulation is a reoccurring theme for WLC. Thürer et al. (2012a) simulate WLC with 6 different release methods (Immediate release, periodic release, LUMS COR, WCPRD, SLAR, CONWIP) and 2 dispatching rules (FCFS and PST). Again, dynamism and external complexity is hypothetical, and set low, but contrary to Mortágua et al. (2014), Thürer et al. (2012a) include fluctuating production flow, simulating directed flow in increments of 25%, meaning that a pure job shop environment and a general flow shop environment is simulated. By doing so, external complexity (CE 1, table 2) is simulated and optimal release method is found given any scenario.

In WLC, the mathematical solutions for both release methods and sequencing within the work cell require known data. Thürer et al. (2012a) use two dispatching (sequencing) rules: FCFS and PST. First-Come-First-Served (FCFS), require no data other than arrival time, but Planned operation Start Time (PST) require specific information regarding the product, namely processing time, due date and remaining operations. As such, for WLC to be an effective solution, internal dynamism has to be low. FCFS is one of the most common dispatching rules, and is applied to a CONWIP system by Matt et al. (2014), with the exception of giving rework highest priority. As this is a dispatching rule, there could be questions regarding WLC in the context of ETO, or any high-dynamism environments. Although not related to a release method, Girod et al. (2014) propose a hybrid-dynamic manufacturing system that in many ways function as a shop-floor dispatching rule. If high runners can be identified, these have first priority both on order release and in the queue in front of work centers. As Girod et al. (2014) is designed to have flexibility regarding PPC system.

As both complexity and dynamism influences WLC, Oosterman et al. (2000) show why flow direction impacts the optimal solution regarding release methods. As the simulation performed is a simulation of four extreme variants of job shops (pure job shop, no directed flow; general flow shop, fully directed flow; varied routing length, varied number of stages visited; constant routing length, all stages visited), the result show the need for the variables to be known in order to select the optimal solution. The solution optimal in a pure job shop, is the worst solution for a general flow shop, which is a statement that gives validity to the adaptation of Bohn (1998) in Sousa and Voss (2001) that state the need to know which process variables are affected or affect implemented practices in order to achieve “best in class practice”.

Marangoni et al. (2013) argue that it is possible to implement WLC in an ETO environment with high degree of dynamism in the form of variability in routing variability and lead time variability (internal dynamism). This is done by basing families on routing, and mapping each family for improvements towards the lean value stream (Rother and Shook, 2003). However, the solutions found suggest that the dynamic impacts of both internal and external factors hinder WLC as suggested by Marangoni et al. (2013) to be utopic visions of the future state. The suggestions are either to make dedicated routing, which is a waste generating solution rather than a waste eliminating implementation; because the processing stages on the shop floor is shared by the majority of the products, dedicating machines to the routing of a single

family is dependent on volume of selected family to give a throughput high enough to avoid the stages to go idle. In addition, by removing stages from other routings, bottlenecks could be created. As such, WLC cannot be applied to an ETO environment (or a highly dynamic MTO environment).

To summarize, WLC is an excellent solution, given predictable production and supply. As such, the technique proves a good solution to handle demand unpredictability and instability. In addition it is a way to achieve levelling in MTO (Thürer et al., 2012a). This means that it is suitable for a MTO environment, but needs further investigation for the ETO environment.

One application that suggests that WLC the best PPC solution for MTO, is its ability to handle routing variability. Because entire routings are analyzed, work is released in such a way that the current pool of orders continue to be released onto the shop floor with as high utilization as possible while preventing accumulation of work at the bottleneck(s), or even prevent bottlenecks from occurring.

The main argument against both CONWIP and DBR is that they do not include the customer inquiry and job release stage (Stevenson et al., 2005). These difficulties can be overcome by including pre-shop sequencing, pushing the simpler CONWIP and/or DBR systems towards a WLC system. This will increase the complexity of the capacity planning, but in return, the ability to increase due date setting at the customer enquiry stage increase.

4.2.6 PPC Summary

Table 5 shows the literature on PPC in MTO and how it relates to the environmental characteristics present in the relative literature.

Table 5: Literature of PPC methods in the MTO industry

Author(s)	PPC	Application Area	Key Findings	Environmental Characteristics
Bicheno and Holweg (2016)	Kanban	High mix	Kanban variations do allow for product anonymous Kanban's	Large number of variations on the same product. High complexity, low dynamism
Bertolini et al. (2015)	WLC	MTO	Simulate WLC in a HVLV environment. Conclude that there is a tradeoff between WLC and push production (MRP), where WLC reduce WIP, MRP increase due date adherence	Job-shop configuration. SME
Chang and Yih (1994), Chang and Yih (1998)	Kanban	MTO	Generic Kanban, anonymous Kanban's, suits an environment where the mix is low.	Low complexity, low dynamism. High demand fluctuation
Darlington et al. (2015)	Kanban	MTO, ETO	Kanban not feasible due to being stock based	Simple routings, engineering per order. High dynamism, low complexity.
	CONWIP		CONWIP rejected due to complex routings and lead time variation.	
	DBR		Identifying the bottleneck require external tools in complex systems.	
Gastermann et al. (2011), Gastermann and Stopper (2012)	CONWIP	ATO	CONWIP works very well when WIP is low.	Medium sized company.
Gravel and Price (1988)	Kanban	MTS	Three sequencing techniques to choose what order to start next, given available capacity. Release is in	Seasonal production. High complexity, low dynamism. Job-shop.

essence CONWIP,
rather than Kanban

Junior and Godinho Filho (2010)	Kanban	N/A	Kanban modifications reviewed. Appropriate Kanban modifications are tailored to specific environments, and more often than not resemble other card-based solutions such as CONWIP or POLCA.	N/A
Kjersem et al. (2015)	Kanban	MTO	Kanban as a supply line to hull production in shipbuilding.	Simple routings, repetitive work tasks.
	CONWIP		Assembly platforms restrict WIP to accumulate, with dedicated workforce working on jobs in a FIFO sequence.	
Krishnamurthy and Suri (2009)	POLCA		Step-by-step implementation method for POLCA. Shows great promise in complex systems that need to be responsive. More dynamic systems could only implement POLCA on certain products.	High complexity
Marangoni et al. (2013)	WLC	MTO	Use VSM to determine WLC configuration in a HVLV environment.	Complex Bill of Material, convergent workflows.
Matt et al. (2014)	CONWIP	MTO	CONWIP is a good system for off-site production within the construction industry. The backlog nature of CONWIP allows for control over what to produce next.	Simple routings for aluminum frames with one-of-a-kind specifications.
Melchert et al. (2006)	WLC	MTO	Simulations show that WLC is a system that handles the	High variety of products, low

			characteristics of MTO well.	quantity, variable demand.
Mortágua et al. (2014)	WLC	MTO	Investigate how WLC with different workload strategies and pre-shop sequencing rules. Show that WLC can work in a pure flow-shop.	Small to medium size. Unidirectional workflow, simple routings.
Oosterman et al. (2000)	WLC	N/A	When designing the WLC system, the key characteristic to performance is direction of workflow.	Job shop.
Riezebos et al. (2009)	Kanban	MTS/MTO/ETO	Kanban is restricted to MTS due to being product specific.	N/A
	CONWIP		CONWIP is mainly a MTS strategy, but can function in MTO given simple routings.	
	POLCA		Works for all CODP configurations. POLCA is a flexible system because the only input needed is the succeeding cell at any point in manufacturing.	
Riezebos (2010)	POLCA	MTO	Need for more case-studies on POLCA. Increase focus on cell-specific release. POLCA in case-study reduce lead time and increase due date adherence.	High complexity, complex product flows (routing). Small to medium size.
Shi-chao et al. (2012)	POLCA	MTO	The POLCA card count can be optimized through programming.	Unidirectional workflow.
Spearman et al. (1990)	CONWIP	N/A	CONWIP function with lower WIP than Kanban. It also functions in jobs with short production runs.	High demand fluctuations.
	DBR		DBR works in a pure job-shop. DBR and CONWIP end up as similar systems when applied to flow-systems	

Stevenson et al. (2005)	Kanban	MTO	Kanban cannot cater with high routing variability and lack of repetition	High complexity and dynamism, VMC and RBC classified with "Variety" and "Volume". Introduce company size as a notable selection criteria
	CONWIP		Families of common routings necessary. Would need batching in MTO, which leads to pore due date adherence.	
	DBR		Stationary bottleneck a prerequisite for DBR to work. Can work in the correct circumstance.	
	POLCA		Accommodates high routing variability, but struggles with variations in routing directions.	
	WLC		Designed for MTO. Accommodates non-repeat production and variable routings. Increases the PPC choice problem due to having many different configuration options.	
Stump and Badurdeen (2012)	CONWIP	MTO	Two customized products and three production cells controlled by CONWIP.	Simple routings, low dynamism.
Thurer et al. (2016), Thüerer et al. (2016a), Thüerer et al. (2014a), Thüerer et al. (2016b), Thüerer et al. (2014b)	Kanban	N/A	Kanban works as a replenishment line in complex and dynamic systems.	Pure Flow Shop, low complexity and dynamism
	CONWIP		Argue that pure CONWIP only work in pure flow shops due to all jobs having to enter and leave at the same point, and that flow of work should not be split	Pure Flow Shop, low complexity and dynamism. Allows for a greater mix of products compared to Kanban
	POLCA		POLCA is an effective tool for MTO companies, but is very sensitive to its inputs	High complexity, low dynamism. Allows for high routing variety, but

		needs directed flow. Accurate information required when scheduling.
WLC	WLC effective in job shop. Card-based WLC termed COBACABANA.	No restriction to routing, and allows for high processing time variability.

4.3 Pre-Shop Sequencing

Of the four appropriate PPC methods discussed (CONWIP, DBR, POLCA, WLC), only WLC inherently include a pre-shop sequencing. CONWIP and DBR tell when a new order should be released, but not which one. Because any release rule is better than to have no release rule, any PPC method that gets implemented should use FIFO sequencing in the pre-shop release pool. This will stabilize production, reducing throughput variation and tardiness. In addition, by using a sequencing rule, the rule can be revised based on performance, and improvements can be made. It may also provide improved performance measurements throughout the shop-floor, allowing for a more accurate review of the PPC method chosen as a whole, in addition to the obvious review of the release rule itself.

POLCA release a product with an open routing, and release based on earliest release date set by the MRP system. This does function as FIFO, both pre-shop and on the shop floor.

However, with more complex systems (Birkie and Trucco, 2016a) (table 2), the sequencing required to meet the goals of a PPC method in MTO (Stevenson et al., 2005) will become to general and sub-optimized. The common POLCA system (Krishnamurthy and Suri, 2009, Riezebos, 2010) will therefore benefit from increasing the sophistication of the pre-shop sequencing when system complexity increases.

The main argument against both CONWIP and DBR is that they do not include the customer inquiry and job release stage (Stevenson et al., 2005). These difficulties can be overcome by including pre-shop sequencing

4.4 Summary

Discussing PPC methods in MTO, there is a number of variables that affect the outcome of any chosen solution. This in turn cause a specific scenario to have potentially more than one suitable solution. Due to this, there is an incentive to create a single solution that inherently is

simple, but can be made more complex as the system matures and operators gain an understanding of the actual impactful variables, as well as what KPIs that needs to be improved/focused on. This in turn suggests that WLC can be divided into simpler solutions that inherently will use CONWIP or DBR logic, with a guided system towards maturity and increased complexity, increasing the actual control of the system, setting stricter and more rules using an increased amount of measurements to increase the precision of the system in terms of setting and adhering to the due date while at the same time striving for reduced total lead times (or similar improved production performance).

POLCA on the other hand uses slightly different logic, and have a stricter inherent ruleset. It relies on known routings as well as accurate estimation of the earliest release date, as planned with an MRP system. The system handle environmental factors relating to complexity better than the system that can be defined as WLC (CONWIP, DBR, WLC), but handle dynamism poorly. This means that POLCA is very suitable for MTO companies with high mix that do little to no reengineering, and have a stable and directed workflow on the shop-floor.

The biggest problem occurring when reviewing literature on PPC, is that there are rarely any sources citing failed applications. Perhaps due to the nature of the methods, especially CONWIP and DBR, to start loose and tighten with experience, if a case-study is conducted and these methods are implemented, production will see improved performance due to the share nature of highlighting and attempting to fix the already identified performance goals. There are however many examples of attempted implementations that needed adaptations in order for the methods to fit the environment. This again hints that a simpler method like CONWIP and DBR should be attempted, with the goal of increasing the sophistication of the system with experience, and using the WLC theory as a baseline for what is to be achieved once the system has matured. As discussed, POLCA is more rigid and “set in its way”.

5. Conclusion

Defining MTO has in this paper included ETO as well as HVLV where it is appropriate. Using this definition, a number of critical selection criteria have been identified, relating to environmental factors that can be categorized as complexity or dynamism.

Assessing PPC in a MTO manufacturing environment, it is clear that the choice of PPC systems in reality is a choice between three (four) solutions: Push production (MRP), push/pull hybrid (POLCA), or WLC (The fourth solution is to have no PPC system at all). Out of the three systems, WLC allows for the simplest implementation, because CONWIP or DBR (or both) can be implemented first as incomplete systems. As experience is built, the system can be tailored, and if need be, developed into a WLC. POLCA is a more complex system, as well as a more strict system. It does allow for a hybrid push/pull that works very well within the MTO sector, but is very prone to errors both in routing predictions and earliest release date. It also relies on a MRP system, which has proven to have difficulties in the MTO sector, especially for companies considered small or medium sized.

Relating back to environmental factors, it is clear that, while CONWIP and DBR most certainly fit in some MTO environments, complexity is the biggest differentiator between these methods and POLCA and WLC. BOM size and complexity and product mix both increase the number of possible routings, and CONWIP will quickly underperform in a complex environment. DBR performs better, but is reliant on a stable, identifiable bottleneck. Both CONWIP and DBR handle dynamism poorly. POLCA is tailored to a complex environment. The nature of only releasing work to the next work station when there is available capacity downstream carries some of the functionality of WLC with it, without needing the same quantity of information, making it perfect for a complex environment. However, it is vulnerable to dynamism, and requires unidirectional workflow.

WLC can be concluded as being the most suitable PPC for MTO. It performs well in all defined MTO environments, with its only drawback being that it is an information dependent system that is prone to becoming complex (the system, not complex as in the environmental factor). In addition, both DBR and even more suitable, CONWIP can be implemented with loose restrictions in order to get to know the system, and building the system towards a WLC as it matures and its operators gain experience. In addition to the operators gaining experience along the way of implementation, the precision of information can continuously be

monitored and checked, allowing for both more information being available regarding both routings and workload.

5.1 Further Work

Scheduling and sequencing jobs are processes apply to more than the manufacturing industry. Both service jobs, as well as administrative tasks will usually have a flow of work. Expanding this way of thinking to both administrative jobs as well as engineering jobs can further increase effectiveness: doing the right things at the right time.

When assessing PPC methods in MTO, there are a number of options in terms of validating finds. Single case-study can be used to describe various functions of a selected method, such as PPC configurations, implementation, or continuous improvement. Multiple case-study can be used to compare different methods in similar manufacturing environments. Simulation(s) can be used to identify most applicable theoretical solutions, as well as how the different PPC methods change depending on the input of variables within the systems simulated.

Reference list

- ADRODEGARI, F., BACCHETTI, A., PINTO, R., PIROLA, F. & ZANARDINI, M. 2015. Engineer-to-order (ETO) production planning and control: an empirical framework for machinery-building companies. *Production Planning & Control*, 26, 910-932.
- AMARO, G., HENDRY, L. & KINGSMAN, B. 1999. Competitive advantage, customisation and a new taxonomy for non make-to-stock companies. *International Journal of Operations & Production Management*, 19, 349-371.
- AMRANI, A., ZOUGGAR, S., ZOLGHADRI, M. & GIRARD, P. Supporting framework to improve Engineer-to-Order product lead-times. 5th IFAC Conference on Management and Control of Production and Logistics, MCLP'10, 2010 Coimbra. 102-107.
- BERTOLINI, M. & ROMAGNOLI, G. Lean manufacturing in the valve pre-assembly area of a bottling lines production plant: An Italian case study. 2013 International Conference on Industrial Engineering and Systems Management, IEEE - IESM 2013, 2013 Rabat. IEEE Computer Society.
- BERTOLINI, M., ROMAGNOLI, G. & ZAMMORI, F. Assessing performance of Workload Control in High Variety Low Volumes MTO job shops: A simulative analysis. Industrial Engineering and Systems Management (IESM), 2015 International Conference on, 2015. IEEE, 362-370.
- BICHENO, J. & HOLWEG, M. 2016. *The Lean toolbox: The essential guide to Lean transformation*, Picsie Books.
- BIRKIE, S. E. & TRUCCO, P. 2016a. Understanding dynamism and complexity factors in engineer-to-order and their influence on lean implementation strategy. *Production Planning & Control*, 27, 345-359.
- BIRKIE, S. E. & TRUCCO, P. 2016b. Understanding dynamism and complexity factors in engineer-to-order and their influence on lean implementation strategy. *Production Planning and Control*, 27, 345-359.

- BOHN, R. E. 1998. Measuring and managing technological knowledge. *The Economic Impact of knowledge, Butterworth-Heinemann, Boston*, 295-314.
- BOHNEN, F., BUHL, M. & DEUSE, J. 2013. Systematic procedure for leveling of low volume and high mix production. *CIRP Journal of Manufacturing Science and Technology*, 6, 53-58.
- BRINK, T. & BALLARD, G. SLAM—A case study in applying lean to job shops. Construction Research Congress 2005: Broadening Perspectives, 2005. 1-12.
- CHANG, T.-M. & YIH, Y. 1994. Generic kanban systems for dynamic environments. *The International Journal of Production Research*, 32, 889-902.
- CHANG, T.-M. & YIH, Y. 1998. A fuzzy rule-based approach for dynamic control of kanbans in a generic kanban system. *International Journal of Production Research*, 36, 2247-2257.
- DARLINGTON, J., FRANCIS, M., FOUND, P. & THOMAS, A. 2015. Design and implementation of a Drum-Buffer-Rope pull-system. *Production Planning & Control*, 26, 489-504.
- FRAZEE, T. & STANDRIDGE, C. 2016. CONWIP versus POLCA: A comparative analysis in a high-mix, low-volume (HMLV) manufacturing environment. *Journal of Industrial Engineering and Management*, 9, 432.
- GASTERMANN, B. & STOPPER, M. Conceptual prototype of a planning software for the CONWIP production control system. 2012 International MultiConference of Engineers and Computer Scientists, IMECS 2012, 2012 Kowloon. Newswood Limited, 1334-1339.
- GASTERMANN, B. C., STOPPER, M. & KATALINIC, B. Aspects of the production control system CONWIP in small and medium sized industrial companies. Annals of DAAAM for 2011 and 22nd International DAAAM Symposium "Intelligent Manufacturing and Automation: Power of Knowledge and Creativity", 2011 Vienna. Danube Adria Association for Automation and Manufacturing, DAAAM, 621-622.

- GIROD, O., ZHANG, H., CALVO-AMODIO, J., HAAPALA, K. R. & MASON, J. B. A proposed hybrid-dynamic transition phase for high mix low volume manufacturers. IIE Annual Conference and Expo 2014, 2014. Institute of Industrial Engineers, 2583-2592.
- GOLDRATT, E. M. & COX, J. 2016. *The goal: a process of ongoing improvement*, Routledge.
- GOSLING, J. & NAIM, M. M. 2009. Engineer-to-order supply chain management: A literature review and research agenda. *International Journal of Production Economics*, 122, 741-754.
- GRAVEL, M. & PRICE, W. L. 1988. Using the Kanban in a job shop environment. *The International Journal of Production Research*, 26, 1105-1118.
- GUPTA, M. & BOYD, L. 2011. An Excel-based dice game: an integrative learning activity in operations management. *International Journal of Operations & Production Management*, 31, 608-630.
- IRANI, S. 2011. Choosing what works. *Industrial Engineer*, 43, 42-47.
- JUNIOR, M. L. & GODINHO FILHO, M. 2010. Variations of the kanban system: Literature review and classification. *International Journal of Production Economics*, 125, 13-21.
- KJERSEM, K., HALSE, L. L., KIEKEBOS, P. & EMBLEMSVÅG, J. 2015. Implementing lean in engineer-to-order industry: A case study. In: MIZUYAMA, H., HIBINO, H., KIRITSIS, D., UMEDA, S., VON CIEMINSKI, G. & NAKANO, M. (eds.) *IFIP WG 5.7 International Conference on Advances in Production Management Systems, APMS 2015*. Springer New York LLC.
- KRISHNAMURTHY, A. & SURI, R. 2009. Planning and implementing POLCA: a card-based control system for high variety or custom engineered products. *Production planning and control*, 20, 596-610.
- LAND, M. J. 2009. Cobacabana (control of balance by card-based navigation): A card-based system for job shop control. *International Journal of Production Economics*, 117, 97-103.
- LANE, G. 2007. *Made-to-order Lean: Excelling in a High-mix, Low-volume Environment*, Productivity Press.

- MARANGONI, G., ROMAGNOLI, G. & ZAMMORI, F. Multiple value stream mapping: How to implement work load control in complex systems. *Industrial Engineering and Systems Management (IESM)*, Proceedings of 2013 International Conference on, 2013. IEEE, 1-9.
- MATT, D. T., DALLASEGA, P. & RAUCH, E. Synchronization of the manufacturing process and on-site installation in ETO companies. 47th CIRP Conference on Manufacturing Systems, CMS 2014, 2014 Windsor, ON. Elsevier, 457-462.
- MELCHERT, E. R., DE MESQUITA, M. A. & FRANCISCHINI, P. G. Lean manufacturing on make-to-order suppliers: a case study. 12th International Conference on Industrial Engineering and Operations Management (ICIEOM), 2006. 1-9.
- MORTÁGUA, J., FERNANDES, N. O. & CARMO-SILVA, S. 2014. How Continuous Workload-Based Release Methods Perform in Flow-Shops? *Romanian Review Precision Mechanics, Optics & Mechatronics*, 7.
- OLHAGER, J. & PRAJOGO, D. I. 2012. The impact of manufacturing and supply chain improvement initiatives: A survey comparing make-to-order and make-to-stock firms. *Omega*, 40, 159-165.
- OOSTERMAN, B., LAND, M. & GAALMAN, G. 2000. The influence of shop characteristics on workload control. *International journal of production economics*, 68, 107-119.
- POWELL, D., STRANDHAGEN, J. O., TOMMELEIN, I., BALLARD, G. & ROSSI, M. 2014. A new set of principles for pursuing the lean ideal in engineer-to-order manufacturers. *Procedia CIRP*, 17, 571-576.
- PRAKASH, J. & CHIN, J. F. 2015. Modified CONWIP systems: a review and classification. *Production Planning & Control*, 26, 296-307.
- RAJASEKAR, S., PHILOMINATHAN, P. & CHINNATHAMBI, V. 2006. Research methodology. *arXiv preprint physics/0601009*.
- RIEZEBOS, J. 2010. Design of POLCA material control systems. *International Journal of Production Research*, 48, 1455-1477.

- RIEZEBOS, J., KLINGENBERG, W. & HICKS, C. 2009. Lean Production and information technology: Connection or contradiction? *Computers in Industry*, 60, 237-247.
- ROTHER, M. & SHOOK, J. 2003. *Learning to see: value stream mapping to add value and eliminate muda*, Lean Enterprise Institute.
- RUDBERG, M. & WIKNER, J. 2004. Mass customization in terms of the customer order decoupling point. *Production planning & control*, 15, 445-458.
- SEIDMANN, A. 1988. Regenerative pull (Kanban) production control policies. *European Journal of Operational Research*, 35, 401-413.
- SHI-CHAO, L., GUO-ZHU, J. & WEN, K. A mathematical programming formulation approach to the number of POLCA cards. *Management Science and Engineering (ICMSE)*, 2012 International Conference on, 2012. IEEE, 218-222.
- SLOMP, J., BOKHORST, J. A. C. & GERMS, R. 2009. A lean production control system for high-variety/low-volume environments: A case study implementation. *Production Planning and Control*, 20, 586-595.
- SOMAN, C. A., VAN DONK, D. P. & GAALMAN, G. 2004. Combined make-to-order and make-to-stock in a food production system. *International Journal of Production Economics*, 90, 223-235.
- SOUSA, R. & VOSS, C. A. 2001. Quality management: universal or context dependent? *Production and Operations Management*, 10, 383-404.
- SOWER, V. E. & ABSHIRE, R. D. 2003. Successful implementation of advanced manufacturing technology: a cross sectional survey. *International journal of computer applications in technology*, 16, 12-20.
- SPEARMAN, M. L., WOODRUFF, D. L. & HOPP, W. J. 1990. CONWIP: a pull alternative to kanban. *The International Journal of Production Research*, 28, 879-894.
- STEVENSON, M. 2006. Refining a workload control (WLC) concept: a case study. *International Journal of Production Research*, 44, 767-790.

- STEVENSON, M., HENDRY, L. C. & KINGSMAN, B. 2005. A review of production planning and control: the applicability of key concepts to the make-to-order industry. *International journal of production research*, 43, 869-898.
- STRØM, R. E. 2017. Application of Value Stream Mapping in High-Variety/Low-Volume Manufacturing in Context of Environmental Factors.
- STUMP, B. & BADURDEEN, F. 2012. Integrating lean and other strategies for mass customization manufacturing: a case study. *Journal of Intelligent manufacturing*, 23, 109-124.
- THÜRER, M., STEVENSON, M. & PROTZMAN, C. 2016. *Card-Based Control Systems for a Lean Work Design: The Fundamentals of Kanban, ConWIP, POLCA, and COBACABANA*, CRC Press.
- THÜRER, M., LAND, M. J. & STEVENSON, M. 2014a. Card-based workload control for job shops: Improving COBACABANA. *International Journal of Production Economics*, 147, 180-188.
- THÜRER, M., STEVENSON, M. & PROTZMAN, C. W. 2015. COBACABANA (Control of Balance by Card Based Navigation): An alternative to kanban in the pure flow shop? *International Journal of Production Economics*, 166, 143-151.
- THÜRER, M., STEVENSON, M. & PROTZMAN, C. W. 2016a. Card-based production control: a review of the control mechanisms underpinning Kanban, ConWIP, POLCA and COBACABANA systems. *Production Planning & Control*, 27, 1143-1157.
- THÜRER, M., STEVENSON, M. & QU, T. 2016b. Job sequencing and selection within workload control order release: an assessment by simulation. *International Journal of Production Research*, 54, 1061-1075.
- THÜRER, M., STEVENSON, M., SILVA, C., LAND, M. J. & FREDENDALL, L. D. 2012a. Workload control and order release: A lean solution for make-to-order companies. *Production and Operations Management*, 21, 939-953.
- THÜRER, M., STEVENSON, M., SILVA, C., LAND, M. J. & FREDENDALL, L. D. 2012b. Workload Control and Order Release: A Lean Solution

- for Make-to-Order Companies. *Production and Operations Management*, 21, 939-953.
- THÜRER, M., STEVENSON, M., SILVA, C., LAND, M. J., FREDENDALL, L. D. & MELNYK, S. A. 2014b. Lean control for make-to-order companies: Integrating customer enquiry management and order release. *Production and Operations Management*, 23, 463-476.
- WAHLERS, J. L. & COX III, J. F. 1994. Competitive factors and performance measurement: applying the theory of constraints to meet customer needs. *International Journal of Production Economics*, 37, 229-240.
- WIENDAHL, H.-P. 1995. *Load-oriented manufacturing control*, Springer.
- ZÄPFEL, G. & MISSBAUER, H. 1993. New concepts for production planning and control. *European Journal of Operational Research*, 67, 297-320.