Paper:
Wikner, J., Yang, B., Yang, Y. & Williams, S. Decoupling Thinking in Service Operations: A Case in Healthcare Delivery System Design. *Production Planning and Control*

Decoupling thinking in service operations: A case in healthcare delivery system design

Abstract: The notion of decoupling thinking has been well established in the manufacturing operations and supply chain management literature. This paper explores how this decoupling thinking can be applied in service operations and in particular in healthcare. It first reviews the relevant literature on decoupling fundamentals, the front- and back-office distinction, and new emerging decoupling thinking in service operations. Subsequently, a flow based framework including content and process is developed for decoupling thinking in service operations. The framework provides an integrated perspective on customer contact, flow driver, and flow differentiation (level of customisation). The framework hence, through flow differentiation, introduces the concept of standardization versus customization in a service context. This is followed by a healthcare case example to illustrate how the framework can be applied. The managerial implications are primarily in terms of a modularized approach to system design and management. The framework offers potential for benchmarking with other service systems as well as with manufacturing systems based on the shared foundation in decoupling thinking. Finally, suggestions are provided for further research opportunities derived from this research.

Keywords: Decoupling, service operations, healthcare, flow thinking, customisation

1 Introduction

Decoupling thinking has a long historical background in business operations. To get the most out of personal limited resources (labour), Adam Smith (1776) introduced the concept of the division of labour. Using the famous example of pins, he referred to it as the practice of decoupling the (pin-making) process into different steps and assigning each step to a specific worker, thus greatly increasing the overall productivity of the factory. This approach to increasing productivity has been further developed into a foundation for mass production in scientific management (Taylor, 1911) and Skinner's (1974) notion of plant-within-a-plant (PWP). Going beyond individual resources, PWP advocates segmentation of a manufacturing facility both organisationally and physically into homogeneous units. Each PWP concentrates on particular manufacturing tasks with, for example, its own objectives, operating procedures, human management approach, and organisation structure.

Drawing on the PWP concept and including the role of customer contact in organisation design, Chase and Tansik (1983) define decoupling as separating activities of a service organisation, physically or organisationally, and placing them under separate supervision. Accompanying this, a traditional way of defining decoupling in service operations takes into consideration front-office and back-office differentiation. Decoupling here is referred to as "breaking a process into its component back- and front-office activities, segregating those activities into distinct back- and front-office jobs, and, usually, geographically separating the back- and front-offices" (Metters and Vargas, 2000, p. 664). Referring back to the division of labour, scientific management, and PWP, this approach to decoupling, including front-office and back-office activities, does not only involve the resource perspective of Smith, Taylor, and Skinner but also adds the perspective of the customer and how customer value is created in the processes. Such decoupling of processes and activities has long been of interest to the service operations literature (Chase, 1978; Shostack, 1984). It has been empirically examined in different contexts such as banking and financial services (e.g. Metters and Vargas, 2000; Zomerdijk and de Vries, 2007), healthcare (Broekhuis *et al.*, 2009), energy (electricity) supply (Ponsignon *et al.*, 2011), and servitised manufacturing (Pawar *et al*., 2009).

The contact point with the customer is a key issue for decoupling in manufacturing operations particularly in relation to process adaptation (i.e. customisation). The interest in processes for customisation dates back to at least the quality management movement (e.g. Deming, 1982). This is when the actual transformation process was explicitly emphasised and consequently the resources mainly played the role as executors of the processes. In this context the transformation process relates the resources to the needs of the customers, which is in line with the foundations of approaches such as lean thinking (Womack and Jones, 1996). From a process perspective the driver that triggers the execution of a process is a key attribute. The process based approach to early decoupling thinking emphasised strategic placing of inventory at key positions in order to decouple the flow related to the driver of the flow (see e.g. Hoekstra and Romme, 1992). This approach to decoupling thinking has been well established in the operations and supply chain literature, which has been reflected in manufacturing based concepts such as the customer order decoupling point (CODP), order penetration point (OPP), push-pull boundary, postponement and leagility (e.g. see Sharman, 1984; Giesberts and van der Tang, 1992; Hoekstra and Romme, 1992; Pagh and Cooper, 1998, Naylor *et al*., 1999; Chopra and Meindl, 2004; Kellar *et al.,* 2016).

Similar to the notion of separating the high customer contact front-office processes from the standardised back-office processes, manufacturing decoupling thinking explores opportunities for improving efficiency and effectiveness by separating the production flow into sub flows with different specific properties (Wikner, 2014). The decoupling of the flow can be based on several perspectives but the most common is the distinction between a forecast driven (or make to stock) portion and a customer order driven (or make to order) portion. More recently, Wikner and Noroozi (2016) employed the perspectives of control mode and object type as extensions of the driver perspective.

Despite the different origins and perspectives of decoupling in manufacturing and service contexts, there are similarities between manufacturing and service decoupling which have led us to question whether manufacturing and service decoupling thinking have some complementary potential. For example, can service operations benefit from the more elaborate structural manufacturing decoupling thinking? Would manufacturing operations benefit from a more comprehensive view of the advantages decoupling thinking can provide when services are included? We are also motivated by our observation that, with the boundaries between service and goods becoming more blurred (e.g. Vargo and Lusch, 2004; Sampson and Froehle, 2006), there is an overlapping of concepts and frameworks occurring related to services and manufacturing. This is increasingly being witnessed in the emerging literature on servitisation of manufacturing, product-service systems, and provision of solutions or value packages (Vandermerwe and Rada, 1988; Maull *et al.*, 2014). Our purpose here is therefore to explore how manufacturing decoupling thinking can be operationalised in service operations in a more generic sense. The intention is to provide a more integrative perspective on manufacturing and service operations compared to the recent work on applying manufacturing CODP in the service sector (e.g. Chopra and Lariviere, 2005; Rahimnia and Moghadasian, 2010; Guven-Uslu *et al*., 2014). This part of the literature has treated service operations as a manufacturing system, e.g. thinking of patient flow as material flow. However, this tends to oversimply the unique nature of uncertainty and variation in service operations (Frei, 2006). Furthermore, unlike manufacturing operations, service operations do not necessarily have the luxury of building inventory to complement capacity and time buffers to cope with variability (e.g. in demand). This poses a significant challenge in the conceptualisation of CODP in service operations (Chopra and Lariviere, 2005). Building on previous studies on decoupling thinking in manufacturing operations, we develop a flow based framework for service decoupling thinking.

The remainder of this paper proceeds as follows: First, the relevant literature is reviewed regarding decoupling thinking to support the purpose of this research. Thereafter a flow based framework including content and process is outlined for service decoupling thinking. A healthcare example is then provided to illustrate how the framework can be applied. The paper concludes with managerial implications and suggestions for further research opportunities derived from this research.

2 Literature review

The review of the literature provides a background on the decoupling thinking literature including the fundamentals of manufacturing decoupling, the front- and back-office distinction, and new emerging decoupling thinking in services. As a result we provide a basis for further exploring how a more integrative decoupling thinking approach can be used to inform our understanding of service systems design in a wider context.

2.1 Manufacturing decoupling fundamentals

Originally the concept of manufacturing decoupling was known to separate (i.e. decouple) the performance of two consecutive operations or activities. Later this was refocused to emphasise the decision-making where a strategic decoupling point separates different preconditions for decision making, such as the flow driver being a forecast or customer order (Wikner and Rudberg, 2005; Wikner, 2014). Hoekstra and Romme (1992) used decoupling point as a label for what was later referred to as the ‘customer order decoupling point’ (CODP) (Giesberts and van der Tang, 1992; van Donk and van Doorne, 2016), also known as the order penetration point (Sharman, 1984). The key concept here is that the CODP is a point (Olhager and Östlund, 1990; Pagh and Cooper, 1998; Chopra and Meindl, 2004; Liu *et al*., 2015; Calle *et al.* 2016; Liu *et al*., 2016) where the organisation or the supply chain switches from producing to a forecast (i.e. push or speculation and standardised) and starts producing directly to a customer order (i.e. pull or postponement and sometimes customised). This type of driver based decoupling point is in the literature also referred to as customer order point (Olhager and Östlund 1990) and push-pull boundary (Chopra and Meindl 2004). Nevertheless, the logic remains the same. From here on CODP is used in relation to the flow driver but not covering the differentiation between standardised and customised. The CODP is only referred to as decoupling the flow in terms of what drives the process (i.e. the flow driver).

With the focus on the customer as the driver of the process, two aspects of the CODP are further highlighted in the literature:

(1) *Buffer point*: Intuitively it would be advantageous to have the bottleneck of the production process upstream of the CODP (Olhager, 2003). This way the bottleneck does not have to deal with volatile demand and a variety of different products. The level of inventory (e.g. as safety stock) and capacity upstream of the CODP (acting as a buffer point) can be determined based on aggregated demand (e.g. Hoekstra and Romme, 1992; Pagh and Cooper, 1998). Following this logic manufacturing postponement strategies have been proposed in order to reduce the dependence on forecasting, which lead to better resource planning and allocation (Pagh and Cooper, 1998). These strategies also reduce risk by pooling the variance of the demand, which is aligned with the concept of centralisation of inventories (Eppen, 1979).

(2) *Customisation or Differentiation Point*: As CODP is the point where a product is made for a specific customer order, it is often described as a customisation or product differentiation point (Hoekstra and Romme, 1992; Pagh and Cooper, 1998; Vanteddua and Chinnamb, 2014). In this context, the different positions of the CODP are closely related to the determination of production strategies (e.g. make to stock, assemble to order, make to order and engineer to order) and the level of postponement (Pagh and Cooper, 1998; Olhager, 2003). For example, Olhager (2003) states if the customisation offered is extensive and is at the early stages of production a make-to-order policy is necessary. Alternatively, if customisation occurs at a late stage in production then an assemble-to-order policy may be more appropriate. The distinction between what drives the process and the differentiation of the product can, and many times should, be explicit from a conceptual point of view (e.g. see Olhager and Östlund, 1990; García-Dastugue and Lambert, 2007; Forza *et al*., 2008). The point that separates standardised from customised has, in the same manner as CODP, become known by various terms. Here we use the term customer adaptation decoupling point (CADP) (see e.g. Wikner and Bäckstrand, 2012; Wikner, 2014).

As outlined above, the application of the CODP is in line with the concept of push-pull and its potential boundaries. For clarification, in this paper the push policy is defined as being based on anticipated demand, which means that it corresponds to being forecast driven as used above. The pull policy is defined as being based on a trigger which is generated when a customer order is received (Spearman and Zazanis 1990) (i.e. it is customer order driven). The push-pull hybrid policy represents processes upstream of the CODP which are managed by the push policy, and processes downstream of the decoupling point which are managed by the pull policy (Pyke and Cohen, 1990). This defines the CODP as being a separator between forecast driven and customer order driven activities.

2.2 Decoupling and services

Decoupling has mainly been considered within the context of manufacturing, yet it has an important relevance to services. This section reviews the literature on the front- and back-office distinction, and in addition covers the new emerging decoupling thinking in service systems design including the application of the manufacturing CODP.

2.2.1 Front-office and back-office distinction

The consensus in the literature is that central to a service delivery system design is the explicit consideration of the impact of customer contact (Chase, 1978; Frei 2006). One of the most important goals in service management is to make sure that customer contact activities take into account the when, where and why it needs to operate efficiently without neglecting customer satisfaction (Palmer and Cole, 1995; Zomerdijk and de Vries, 2007). Customer contact, whether the physical presence of the customer in the service delivery system (Chase, 1978) or the virtual presence (e.g. via electronic media such as phone, mail and the Internet) (Zomerdijk and de Vries, 2007), introduces uncertainties and variation (Chase, 1978; Chase and Tansik, 1983; Frei, 2006). While it is deemed to be the dominant constraint on the efficiency and conformance quality of service operations, customer contact also provides valuable sales opportunities e.g. directly responding to customer needs and cross-selling other products. A crucial decision based on the focus of customer contact is the relative allocation of service tasks to the front- and back-offices, with the former responsible for the high-contact elements of work and the latter taking care of the low-contact elements. This decoupling thinking aims to achieve both external effectiveness at the customer interface with the front-office and internal efficiency of operations at the back-office.

Indeed, the decoupling of back-office activities has been a predominant strategy in operations management literature (Metters and Vargas, 2000), as the back-office processes are removed from customer view and can be designed for efficiency. A natural way to improve efficiency is to identify and shift additional activities to the back-office (Chase and Tanski, 1983). This is also in line with the view of Thompson (1967) that the technical core activities (such as back-office work) can be sealed off from the environment (e.g. the randomness of customer behaviour). Low customer contact back-office processes can be rationalised by taking a production-line approach in manufacturing such as mass production (Levitt, 1972), lean production (Bowen and Youngdahl, 1998) and centralisation for economies of scale (Metters and Vargas, 2000). Accordingly, the back-office is often referred to as a “service factory” (Kellogg and Nie, 1995). Rather than lower costs, front- and back-office distinction can also lead to a better fit between job descriptions and worker personality types (Metters and Vargas, 2000; Chase and Tansik, 1983). These two types of activities require public relations and interpersonal skills for high-contact purposes and technical and analytical attributes for low-contact purposes (Chase, 1978). The decoupling of back-office work from front-office can also be linked to other strategic operational objectives such as higher service quality (Metters and Vargas, 2000; Zomerdijk and de Vries, 2007).

Front- and back-office distinction has been well represented in the service design tools. One such tool is service blueprinting, which was introduced by Shostack (1984). It has helped companies like IBM and Aramark to identify possible failure points, improve existing services and develop new services. Service blueprinting is used to visualise process actions and interactions at and around the interface between the customer and service provider. In particular, it draws a line of visibility separating service processes steps that can be seen and experienced by customers and those that cannot (see Figure 1). When and how to move the line of visibility is a crucial issue in service delivery system design (Fließ and Kleinaltenkamp, 2004). Moving the line of visibility, for example, may help to inform the customer about the different steps of the service process and to give him/her insight into the service operation (Fließ and Kleinaltenkamp, 2004). In practice, the line of visibility has been deliberately moved to let customers see what was previously hidden from them, particularly in industries plagued by poor workmanship and shoddy business practices (Harvey, 1998).

2.2.2 Emerging new decoupling thinking in service systems design

While the configuration of front- and back-office work is probably the most common way of conceptualising the impact of customer contact on a service delivery system (Zomerdijk and de Vries, 2007), recently the research on service decoupling thinking appears to be more aligned with manufacturing decoupling thinking. Rather than purely being based on customer contact, it further builds on the vital role of customer integration in service processes. Representing one of the earliest attempts, Fließ and Kleinaltenkamp (2004) introduce manufacturing decoupling thinking to service processes. They divide the service production into two stages: (1) potential stage: activities within this stage only require the service provider's internal production factors, and are thus characterised as customer-independent activities; and (2) process stage (integrating the customer in the service operations): activities within this stage can only be carried out after having been started by the customer or his/her external factors, and thus called customer-induced activities. However, unlike the CODP penetrating into the production system at a point of time, Fließ and Kleinaltenkamp (2004) propose to draw a line of order penetration spanning across the whole service process, to separate the customer-independent activities from the customer-induced activities. Even if this approach uses a different way of illustrating the concept it corresponds well with the CODP as defined here. As with service back-office operations and the forecast driven portion of manufacturing activities, for the purpose of improving efficiency, they suggest that customer independent activities (forecast driven) can be rationalised, standardised, automated, and/or outsourced, and that more activities are moved from the customer-induced area (customer order driven) of the blueprint into the customer-independent area of the service provider.

Moeller (2008) uses this distinction between customer-independent and customer-induced activities in a framework for provision of value that is divided into four stages related to facilities, transformation, and usage. Stage 1 (facility) establishes all company resources that are necessary for service provision. Stage 2 is divided into two sub stages related to customer-independent (stage 2a) and customer-induced (stage 2b). Stage 3 finally involves usage of the output of the provider.

In alignment with manufacturing decoupling thinking, an emerging stream of research is more explicit in relation to the location of interface of decoupled processes. To improve efficiency and effectiveness, for example, manufacturing decoupling thinking has been applied in a healthcare context where patient flow has been simply thought of as material flow (Towill and Christopher, 2005; Rahimnia and Moghadasian, 2010; Guven-Uslu *et al*., 2014). Implicit to this is that the customer arrival point is viewed as the CODP to services. However, this appears to ignore the view that those traditional differentiating service characteristics, such as intangibility, heterogeneity, inseparability and perishability (IHIP), would mean the CODP is placed right at the start of the service operation (i.e. position 0 in Figure 1), namely a service operation is purely a customer order driven (pull) system. For example, inseparability of service production and consumption, and inability to store services indicate that a service normally starts with the arrival of customers. Along a similar line, service is deemed to be essentially a just-in-time system in the literature (e.g. see Sampson, 2000). To clearly delineate service processes from non-service processes Sampson and Froehle (2006) presented a Unified Services Theory (UST). The UST holds that a service production process relies on customer inputs, and customers act as suppliers for all service processes. It also follows from the UST that the precise service juxtaposition is forecast driven (or push) manufacturing. In other words, the customer-order driven (or pull) proportion (i.e. delayed differentiation activities until after a CODP where a customer order is received) in custom manufacturing, or mass customisation is a service (Sampson and Froehle, 2006).

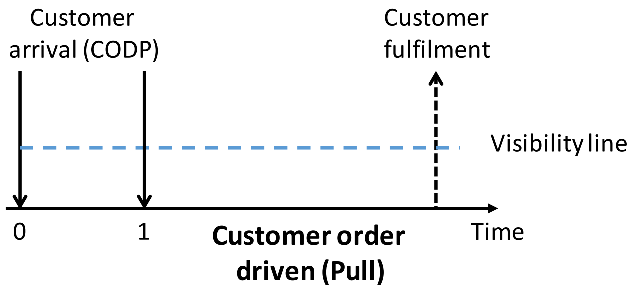

Figure 1. The CODP in a service context

In recognition of these differentiating service characteristics, Chopra and Lariviere (2005) defined service inventory as the work that can be performed and stored prior to the customer's arrival. They also proposed the customer arrival point as the equivalent of the manufacturing push-pull boundary (i.e. the CODP where it is possible to hold inventory of physical products). Hence, the placement of the CODP in service defines the portion of the work that has been performed and stored before the customer arrives. "By wisely choosing what kind of (service) inventory to hold" (Chopra and Lariviere, 2005, p. 56), the CODP can then be moved closer to the market by identifying additional process steps to be completed before customer arrival, and thus reducing the amount of work done in the customer driven phase. In this way, increasing service inventory provides a novel way to lower costs, increase service quality, shorten waiting times, and/or offer greater customisation (Chopra and Lariviere, 2005). As illustrated in Figure 1, this decoupling thinking means shifting the CODP from position 0 to position 1, thus increasing the forecast driven (push, or speculation) portion of the total process. Building on this thinking, Yang *et al.* (2010) demonstrate how postponement can be applied through a re-positioning of the service CODP and moving of the line of visibility to achieve improved service process performance. This is also closely related to the service decoupling point (Wikner, 2012) that highlights similarities between customer-order driven manufacturing and services.

To summarise, decoupling thinking in services frequently aims to increase the forecast driven (push or speculation) portion of a process in order to improve efficiency and quality consistency. From reviewing the literature, it is interesting to note that early decoupling thinking such as the division of labour and the PWP are basically only resource oriented. The back- and front-office is a combination of process and resource based in the sense that it is decoupling the process in terms of customer contact but with emphasis on the efficient use of the resources. The manufacturing decoupling thinking decouples the process from a customer perspective related to the driver. While useful in terms of configuring service operations, the existing literature on applying the manufacturing CODP to services tends to simplify customer input (Sampson and Froehle 2006) and customer integration (Fließ and Kleinaltenkamp 2004) as a one-off order placement activity, which may not always be true in service operations. Therefore, there is a need to further consider the characteristics of service such as the nature of customer contact/involvement while applying manufacturing decoupling thinking into services. A useful framework has also emerged from our review of the literature which can be developed based on the integration of the decoupling thinking concept from manufacturing operations that covers drivers as well as customization, with the concept of front-/back-office from service operations. This goes beyond the positioning of CODP in service operations (Chopra and Lariviere, 2005) by looking for further standardisation opportunities in relation to customer arrival.

As discussed in the previous section, the CODP represents a point of reference for the possible customisation or product differentiation of the offering and this is where the customer adaptation decoupling point (CADP) can be positioned at the earliest (see e.g. Wikner, 2014). This is defined from the viewpoint of a manufacturer, who, by sharp contrast to a service provider, has almost complete control of production processes including the quality and availability of input materials. In manufacturing, customer involvement is generally a one-time event to specify needs and requirements on the basis of which products can be customised for a specific order. In view of the different nature of customer involvement in service operations, we distinguish between CODP and CADP as illustrated in Figure 2. These two points lead to three different types of activities: forecast driven and standardised (FD-ST), customer order driven and standardised (CD-ST) and customer order driven and customised (CD-CU).

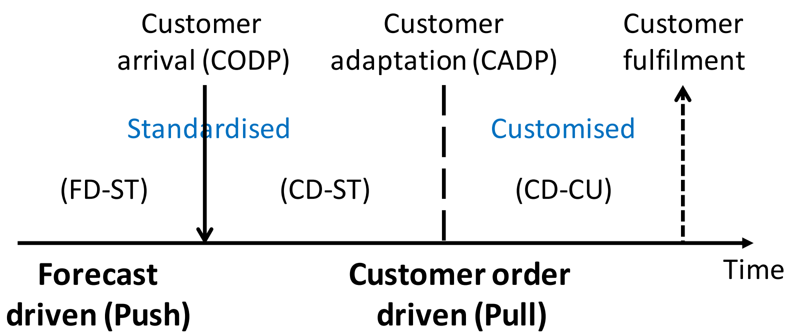

Figure 2. Conceptual framework

3 A flow based framework for service decoupling thinking

Material flow concepts emphasise the importance of a systems approach by highlighting an integrated dynamic control mechanism, with appropriate interfacing and handovers between the core processes of an enterprise (Towill, 1997; Böhme *et al*., 2013; Abaunza *et al*., 2015). This focus on flow is critical for a successful smooth operation as reflected in a "Swift and Even Flow" (Schmenner and Swink, 1998; Fredendall *et al*., 2009). Decoupling thinking is related to discontinuities of the flow in the sense that continuous flow is not possible throughout the complete system. At some points in the flow the context of the flow changes in a way that inhibits a smooth continuous flow. This point is referred to as a decoupling point, or sometimes as a transition point (Wikner and Noroozi, 2016). The discontinuity can be related to many different aspects of the flow such as the driver, the product as a package of goods and services, the transformation resources employed, the control, and the customer contact, all related to the service decoupling thinking outlined above.

The research approach employed to develop the framework is based on a deductive approach where the three decoupling thinking perspectives of flow driver, flow differentiator and customer contact are identified in the literature above as a point of departure. The three perspectives are then combined into a framework where the content refers to the three decoupling perspectives. In addition a process is outlined for how to use the constructs of the content, i.e. building blocks representing different types of decoupling. As an illustration of the potential use of the framework a case from the health care sector is used.

3.1 The content of service decoupling thinking

The general content for both manufacturing and service systems design is based on a set of decision categories that can be combined in different ways (see e.g. Wikner, 2014; Wikner and Noroozi, 2016). The characteristics of these different possible combinations are yet to be fully investigated. For the purpose of service decoupling thinking, five decision categories are used for application on a service operations case. The first two decision categories are necessary for defining the flow to be investigated in terms of the system boundary. The customer represents the downstream boundary and the upstream boundary is simply identified as the Boundary (see Figure 3). The positioning of the Boundary may however be implicit in the sense that the beginning of the flow is "obvious". Besides these two decision categories we have selected three additional decision categories based on the requirement of service operations modelling in line with the literature review above. The first two of these three decision categories (related to CODP and CADP) are detailed based on strategic lead-times and the last category (related to customer contact decoupling point, CCDP) is positioned based on the decided interaction with the customer:

- Flow driver (forecast or customer order) – CODP
  The customer accepts to wait during the delivery lead time (D) between the arrival of the customer (order) and the finalising of the provisioning related to customer fulfilment.
- Flow differentiation (standardised or customised) – CADP
  Since speculation on customised products is a high risk endeavour a strong recommendation is to position CADP downstream from CODP. In this case the requested customisation takes the adapt lead time (A) to perform, and consequently $A \leq D$.
- Customer contact (back-office or front-office) – CCDP
  The CCDP can be combined in different ways with CODP and CADP. It is however not positioned based on strategic lead times, such as D or A, but rather on the decided level of customer interaction.

The combination of these five decision categories (related to Customer, Boundary, CODP, CADP, and CCDP) is illustrated in Figure 3. Since the decision category Customer contact (with separation of back-office and front-office) is included the resulting modules are mainly related to service operations. In Figure 3 a time perspective is also included horizontally in line with the above definitions of CODP and CADP. The supply lead time (S) indicates the time it takes to perform the longest sequence of activities related to the system and in each individual case it is the relation between these three lead times (D, A, and S) (see e.g. Wikner, 2014) that can be used to identify the modules relevant for a particular case. As a consequence the boundary is positioned at the far left of Figure 3 (as in Figure 1 and Figure 2). In this context it is important to note that the concept of standardisation in services is an area that provides additional challenges compared to the context of goods. Standardisation of goods also implies that the process is standardised in order to provide the same goods consistently. In relation to services it is usually the process itself that is in focus when discussing standardisation. It is however important to note that the actual definition of standard is different. A standardised process might provide customised services but here standardisation implies that the service, as perceived by the customer, is not customised but rather the same, independent of the customer. By combining the three decision categories it is possible to identify six fundamental modules as shown in Figure 3. These modules represent the core components of the framework for decoupling thinking in services and can be used in combination for designing a flow model where the characteristics of the respective modules are combined. Each module is referenced with the format Customer contact - Flow driver - Flow differentiator:

(1) The first module (FO-CD-CU) is the basic service module since the front-office is suitable for performing customised activities for a particular customer.
(2) The second module (BO-CD-CU) is when the back-office performs customised activities for a particular customer. In this case customisation can occur independently of the interaction with the customer (Sampson and Froehle, 2006), such as an audit firm providing a unique audit process based on the financial records of a client (corresponding to customer order driven manufacturing of customised items).
(3) The third module (BO-FD-ST) represents standard activities performed on speculation about future customers and back-office with no customer contact. As indicated above a service is sometimes defined as a process performed to customer order and in that case this module would be classified as a goods transforming process rather than a service process (corresponding to forecast driven manufacturing of standardised items).
(4) The fourth module (FO-FD-ST) is at first sight counterintuitive since the activities are performed on speculation but with customer contact. In some cases such as when a trigger at the customer’s work place initiates an activity which means that the provider visits the customer to perform a “service”. But, as in the case of the third module, in many cases, this would not be considered as a service from a theoretical perspective since the activity is performed to forecast even if it involves customer contact. This case has similarities to planned maintenance where a production engineer is performing work triggered by a schedule and not the internal customer. Interestingly this means that a front-office activity can be performed to forecast and hence not classified as a service.
(5) The fifth module (FO-CD-ST) is intriguing since it requires a more relaxed definition of standardisation than what is sometimes used in goods provisioning. Even if the service is considered as a standard service, the customer contact usually involves some basic level of customisation, at least at the personal contact level. From a more general perspective it might of course be possible to completely standardise the interaction with some kind of “robot” application.
(6) The sixth module (BO-CD-ST) is fairly common since it represents a standard procedure performed to a customer request but without customer contact. This is a typical hotspot for automation of services (corresponding to customer order driven manufacturing of standardised items).

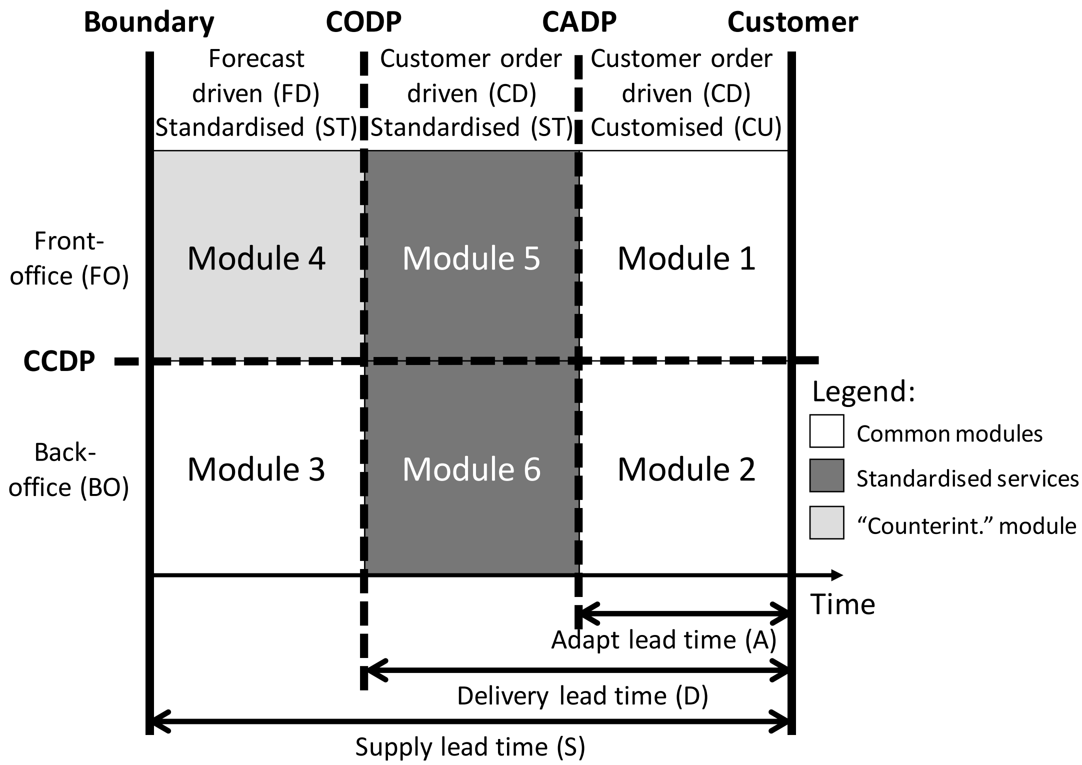

Figure 3. Combining the five decision categories of decoupling thinking content in a service context.

Although the importance of standardisation and customisation in service systems design has been well documented (e.g. see Maister and Lovelock, 1982, Silvestro *et al*., 1992), decoupling thinking informs the design of service delivery systems with a balanced degree of standardisation and customisation. In Figure 3 the two central modules 5 and 6 further develop decoupling thinking. Customer involvement in the service process tends to induce high variability and difficulties in defining specifications. However, service requirements and output need to be distinguished from the service process. While service requirements may be fuzzy and output is heterogeneous (implicit in the unique nature of service), the service process that acquires customer inputs and delivers services is not necessarily unpredictable and hence can be standardised to a certain degree.

3.2 The process of service decoupling thinking

The content of service decoupling thinking, as outlined above, provides details of the method but little information about how to operationalise it. The content of service decoupling thinking covers a wide range of different decision categories but they all share a common foundation in a flow based approach. Using the flow perspective as a point of departure it is possible to identify a generic process, or method, for decoupling thinking which covers the application of the decision categories. The process is summarised in five steps as an ongoing process for continuous improvement in decoupling thinking: The first step Evaluate initiates the process and states the purpose. At this stage the discontinuities related to the decision categories, that are deemed key to the investigation are identified. The second step Estimate outlines the key properties of the identified discontinuities. The third step Design is where the flow is designed based on the identified discontinuities and using the different types of decoupling that are considered as important. The Design step is fundamental since it involves the positioning of decoupling points for each decision category and the analysis of their individual and combined properties (e.g. Wikner, 2014). When the flow is designed it is ready to be used in practice and this is covered by step four Manage/control. Since the preconditions of the flow design may change, the last step Realign is used when the balance between demand and supply must be updated by revising the design of the flow. The process outlined here is not targeting any particular type of industry or company. Instead it is generic and can be applied in all flow related contexts involving process analysis related to individual or combined goods and services.

4 Healthcare based application of decoupling thinking in services

This section illustrates the application of the flow based framework for service decoupling thinking (developed in the previous section) in a healthcare setting. We select this particular sector for the following reasons: (1) With the increasing concerns over medical errors, patient safety and escalating healthcare delivery costs, and the influence of electoral and ideological considerations on healthcare policies and practise (Towill and Christopher, 2005; McFadden *et al*., 2006; Boyer and Pronovost, 2010; Taylor and Taylor, 2014), healthcare has presented an important context and priority for services research. Indeed, as global healthcare systems continue to struggle with increasing demands for their services many organisations are looking for ways in which to improve the design of their systems and work more effectively across the healthcare economy; (2). A careful front- and back-office configuration has deemed to be an important element in developing efficient client-centred healthcare services (Broekhuis *et al*., 2009); and (3). The recent literature has also explored opportunities to improve efficiency and effectiveness by separating the patient flow into sub flows (or different patient pipelines) with different specific properties (Towill and Christopher, 2005; Rahimnia and Moghadasian 2010; Guven-Uslu *et al*.; 2014). Our illustration and analysis is based on a healthcare-related case published by Rahimnia and Moghadasian (2010), which examines the decoupling of patient services. The case focuses mainly on back-office and front-office activities and the use of lean and agility in the design of three treatment flows. Here we have a wider discussion on decoupling by incorporating the five-step process (for continuous improvement) and some key flow based decision categories.

The case is a specialised hospital which focuses its service delivery on the treatment of traumas and injuries. Three patient flows (sometimes referred to as patient pipelines, pathways or value streams) were identified as rupture, fracture and serious injuries. To aid our discussion on decoupling thinking it is appropriate here to briefly outline the characteristics of each of the flows:

- *Rupture flow* –includes patients who are conscious and do not need to be hospitalised. The time to treatment is short and the variety and variability is high. It is a high volume flow but less intense than the serious injury flow, but still requires a quick response.
- *Fracture flow* – patients are conscious and may need to be hospitalised for a short period. The time to treatment is short and the variety and variability is high. The volume of patients is high but again this flow is less intense than

serious injuries and requires a quick response. Once a fracture has been suspected the patient is sent to radiology for an exact diagnosis.

- *Serious injuries flow* – consists of low volumes but high variety and variability. Patients may need resuscitation so response times are critical. Once stabilised patients go through a treatment process which will differ according to the patient's situation.

Figure 4 provides a simple illustration to show the common entry point for patients at the emergency department (corresponding to the CODP) and after the initial diagnosis (corresponding to the CADP) the patient is assigned to one of the three flows.

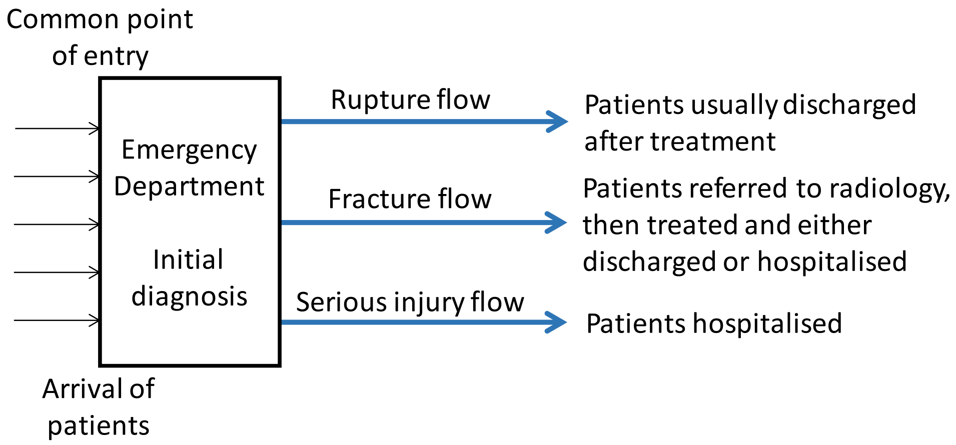

Figure 4. Three flows of hospital treatment

We develop Rahimnia and Moghadasian's (2010) discussion by using this healthcare example to understand the process of decoupling thinking applied in a service context, in particular decoupling thinking content and flow discontinuities.

Step1- Evaluate The first step requires the organisation to assess and evaluate the context and the concerned flows. In terms of the context both the internal (e.g. skill mix, availability of resources and finances, targets and performance measurement) and external (e.g. population health, cultural mix, availability of funding, developments in technology) environment needs to be considered. This step is also concerned with improving the linkage and communication between service providers (hospital) and customers (patients). A key decision in this step is the extension of the system under study. The endpoint is usually the customer but the extension upstream is a matter of decision. In a case where only order fulfilment is to be designed then only the flow from the customer arrival needs to be included. On the other hand, if the preparation for new customers is to be included then the system must also cover some of the forecast driven parts, i.e. upstream of the point where customers enter the system. For the case example the emergency department (ED) is the first point of contact for the patient for all three flows but we also include some of the preparatory activities. Other service providers are likely to be involved such as the emergency services, General Practice, Care homes and Social Services. In the event of a serious injury it is possible that the emergency services will notify ED of the suspected diagnosis and expected time of arrival. This will enable the ED team to prepare for arrival and ensure appropriate skills and equipment are available (e.g. notify staff who are elsewhere in the hospital or on standby for emergencies and clear theatre space for emergency procedures). Essentially there is a need to evaluate current provision to estimate how to improve the design of pathways/flows to ensure that the patient receives high quality, seamless healthcare (Parnaby and Towill, 2008). Evaluation in healthcare needs to be evidenced-based and premised on reliable information/performance management systems. As the patient has not arrived at ED when these preparatory activities are taking place these would be seen as back-office activities.

Step 2 - Estimate The second step starts to consider the design of the service and requires the estimation of key information relating to lead time, cost of supply and value for customer (patient). Understanding what 'customers' value is central to flow design along with understanding capacity and demand, estimating the resources and conditions required to deliver an efficient and effective service. For healthcare the uncertainty of demand levels often makes this stage complex. In our case example all three flows were classified high in terms of variety and variability and short in terms of response times (Rahimnia and Moghadasian, 2010). The cost of supply can be minimised and the value propositions improved by ensuring patients are seen without delay, by appropriately qualified staff and appropriate diagnostics being conducted. Triaging of patients in relation to the severity of condition will provide information in relation to how quickly patients need to be seen.

Step 3 – Design This step focuses on the design of the flow and the positioning of decoupling points, and as a consequence identification of the modules as shown in Figure 3. For our healthcare example we deem the following decisions to be pertinent to the design:

(1) *Customer (Patient)* – For all three flows the ideal situation for the patient is to be treated without delay and the right diagnosis and treatment to be given. This will ensure patients follow the correct flow and ensure they receive high quality and safe care. The customer entry is at the CODP and can be defined as the arrival of the patient at the emergency department to the point that the patient is either discharged or admitted to a ward (Modules 1 or 5). Any treatment the patient receives before arriving at the emergency department (ED) will be referred to as the forward flow (from a strict integrated process perspective the activities performed before arrival are also customer driven). In some cases, preparations are made based on information about the patient but before the actual arrival to ED (Modules 2 and 6). If, for any reason, the patient does not receive the right care and returns to emergency department this would be known as reverse flow, but nevertheless still a customer driven flow.

(2) *Boundary* – Prior to the patient arriving at the emergency department (between the upstream boundary and the CODP from an ED perspective) it is largely a non-controllable situation for the hospital and its staff. As noted above, it relies on good communication with any community or emergency services involved in the patient's care prior to arrival at the ED. Once the patient arrives the situation becomes controllable and diagnostics can commence (which we define as front-office activities corresponding to Modules 1 or 5). Availability of any information prior to the arrival of the patient assists the ED to ensure that back-office activities can be undertaken and the appropriate equipment and staff made available to treat the patient. In our terminology this means that the activities are performed back-office but downstream of the CODP, i.e. the activities are customer driven (Module 2 or 6). In addition substantial preparations are usually made before the patient is known, i.e. on forecast, related to e.g. purchasing of materials and general preparation of the ED (Modules 3 and 4).

(3) *Driver* – The arrival of the patient at the ED (CODP) is the trigger that drives the start of the patient's care (unless sufficient information has been received before the patient arrives) (Modules 1 and 5). Information is provided by the patient, relatives, GP and/or ambulance staff. Clearly the challenge for any healthcare provider is to balance the demand for services with the capacity to treat patients. The use of population and historical data will enable trends and some predictability of referrals or arrivals of patients to be identified (Module 3). In the UK many emergency departments have introduced separate streams for major and minor injuries in order to decrease the number of patients enduring long waits (Cooke *et al*. 2002). As noted above, on arrival at ED patients are usually triaged (CADP) to ascertain the urgency of treatment needed (e.g. whether a minor or major case) but the pathway or flow may not be fully determined at this stage.

(4) *Differentiation* – The diagnosis of the injury (CADP) identifies the appropriate flow (rupture, fracture or serious injury) for the patient. At this point a decision is made to identify the route the patient needs to take and the information about the skills and urgency needed to treat the patient. For patients with a fracture they are first sent to radiology for an exact diagnosis. Although fractures might differ across patients the process of radiology is standardised (Module 5). Beyond radiology the treatment is customised depending on the patient's situation (Module 1). Whilst waiting for a known patient to arrive, and hence before any direct contact with the patient, the preparation can be either standard routines (Module 6) or specific customised actions in preparation for the arrival of the patient (Module 2).

(5) *Customer (patient) contact* – Rahimnia and Moghadasian (2010) make a clear distinction between back-office and front-office activities. The former being identified as provision of medicines, tests, equipment and materials, which Chase (1996) suggests operates similar to a factory and therefore with the application of lean principles can deliver cost efficiencies (Module 3). In an emergency context it is rare to have contact with patients based on a schedule provided by the ED (Module 4). The latter is dependent on the quick, safe and appropriate action of the frontline staff which has to be flexible and adaptable to the needs of each patient (Modules 1 and 5). Broekhuis et al. (2009) also suggest healthcare front-office activities are a costly way of providing services and propose the need for coupling back-office and front-office activities in one job. In our case this might relate to giving prior notice of the arrival of a patient then the appropriate seniority and skill base can be alerted. This could be seen as a back office activity stored until the arrival of the patient and the need for front-activities to be activated (Modules 2 and 6). Clearly the storage of such specialized skills would need to be managed carefully to ensure appropriate utilization of resources.

Step 4 - Manage/Control The management and control of product and service flows is reliant on the decoupled flow. The planning and control of the three flows is highly dependent on the availability of information (e.g. medical records, tests) and resources (e.g. staff, equipment, medical supplies, beds). Performance measurement mechanisms can be employed which monitor for example the level of demand for the flows, patient waiting times (in relation to the CODP and CADP), length of stay (in the flow) and readmissions (reverse flow).

Step 5 - Realign Once decoupling points are identified it is important these are regularly reviewed and realigned in response to the dynamic and changing environment that organisations operate. This is particularly pertinent for healthcare organisations as they need to respond to various challenges including budgetary and resource constraints, policy and regulatory requirements, advancements in medicine/technology, ageing populations, complex medical conditions, improvement interventions and cultural diversity. For our case example, any internal (e.g. staff, equipment, management, finances) and external (e.g. patient population, cultural mix, healthcare policies, new treatments/technologies) changes would prompt the need to realign the design of the flows and the use of decoupling thinking process. The dissemination of the design of improved patient flows is important to the wider healthcare system.

This case has demonstrated how decoupling thinking in terms of the content and the five step process can be employed to provide real support in service design. The process encourages a systematic approach that evaluates the internal and external environment along with reviewing the content and process of decoupling thinking in service operations.

5 Conclusions

As the importance of services continues to build, we have witnessed a growing interest in the transfer of best practices between the manufacturing and service sectors. We have carried out a review of service decoupling thinking within manufacturing and services. We then developed a flow based framework for service decoupling thinking including both content and process. This leads to an interesting, valuable and novel insight into the exploration of decoupling thinking in service systems design, which also contributes to the new emerging literature on applying manufacturing decoupling thinking to service. Using a healthcare example, we illustrate the process of decoupling thinking and the usefulness of understanding decoupling content and flow discontinuities.

The managerial implications of our flow based framework are twofold and similar to those provided for manufacturing operations by Wikner and Noroozi (2016). First, the framework provides a modularized approach to service system design where a clear distinction can be made in terms of the type of customer contact, flow driver and flow differentiation. Based on the modules involved in a particular design a management approach can be outlined for operating the system. Second, the use of the six modules provides an opportunity for benchmarking between different service systems which are based on this modularized approach. In addition, the benchmarking can be extended to manufacturing systems since the framework is partly based on concepts originating from manufacturing operations.

Building on the decoupling thinking framework we note some points of interest that warrant further exploration. This paper has broadened the conceptual scope of decoupling thinking and its service implications. It is a starting point for further work into the development of a unified framework for the use of decoupling in both manufacturing and service systems design. As such most of the paper is deliberately discussed at a high level of abstraction (i.e. considering all or most service in one generic category). Here we have applied the flow based framework to a healthcare setting, however there are other sector-specific characteristics of services which need to be addressed in further research, Finally, while front- and back-office differentiation has a strong conceptual foundation (i.e. customer contact), the literature recognises that the separation of activities between front- and back-office is not clear cut, and that various contingencies affect the configuration and interactions between front- and back-office processes. The conceptual framework presented in Figure 3 thus needs to be empirically tested for its applicability and generalisability to different service environments and potentially also to manufacturing environments.

6 Abbreviations

A	Adapt lead time
BO	Back-office
CADP	Customer adaptation decoupling point
CCDP	Customer contact decoupling point
CODP	Customer order decoupling point
CD	Customer order driven
CU	Customised
D	Delivery lead time
ED	Emergency department
FO	Front-office
FD	Forecast driven
IHIP	Intangibility, heterogeneity, inseparability and perishability
OPP	Order penetration point
PWP	Plant-within-a-plant
S	Supply lead time
ST	Standardised
UST	Unified services theory

7 References

Abaunza, Felipe, Valerie Chavez-Demoulin, Ari-Pekka Hameri and Tapio Niemi. 2015. "Do flow principles of operations management apply to computing centres?" *Production Planning & Control* 26(4): 249-264.

Böhme, Tillmann, Paul Childerhouse, Eric Deakins, Denis Towill, and Sharon Williams. 2013. "Methodology challenges associated with benchmarking healthcare supply chains." *Production Planning and Control* 24(10-11): 1002-1014.

Broekhuis, Manda, Carolien de Blok, and Bert Meijboom. 2009. "Improving client-centred care and services: the role of front/back-office configurations." *Journal of Advanced Nursing,* 65 (5): 971-980.

Bowen, David and William Youngdahl. 1998. *"*'Lean' Service: In Defence of a Product-line Approach." *International Journal of Service Industry Management*, 9(3): 207-225.

Boyer, Kenneth, and Peter Pronovost. 2010. "What medicine can teach operations: what operations can teach medicine." *Journal of Operations Management* 28 (5): 367-371.

Calle, Marcos, Pedorl L. Gonzalez-R and Henri Pierreval. 2016. "Impact of the customer demand on order fulfilment strategies based on floating decoupling point: a simulation analysis." *International Journal of Production Research* 54 (24): 7359-7373.

Chase, Richard 1978. "Where does the Customer Fit in a Service Operation?" Harvard Business Review 56(6): 137-142.

Chase, Richard 1996. “The Mall is My Factory: Lessons from a Service Junkie.” Production and Operations Management 5(4): 298-308.
Chase, Richard and David Tansik. 1983. “The Customer Contact Model for Organisation Design.” Management Science 29 (9): 1037-1050.
Chopra, Sunil and Martin A. Lariviere. 2005. “Managing Service Inventory to Improve Performance.” MIT Sloan Management Review 47(1): 56–63.
Chopra, Sunil and Peter Meindl. 2004. Supply Chain Management: Strategy, Planning, and Operation, Upper Saddle River, NJ: Prentice Hall.
Cooke, Matthew, Sarah Wilson and Sally-Anne Person. 2002. “The effect of a separate stream for minor injuries on accident and emergency department waiting times”, *Emergency Medicine Journal,* 19: 28-30.
Deming, Edwards. 1982. *Out of the Crisis.* Cambridge, MA: MIT Press.
Eppen, Gary. 1979. “Effects of Centralization on Expected Costs in a Multi-location Newsboy Problem.” *Management Science* 25: 498–501.
Fließ, Sabine and Michael Kleinaltenkamp. 2004. “Blueprinting the Service Company: Managing Service Processes Efficiently.” *Journal of Business Research* 57: 392-404.
Forza, Cipriano, Fabrizio Salvador and Alessio Trentin. 2008. “Form postponement effects on operational performance: a typological theory.” *International Journal of Operations & Production Management* 28(11): 1067-1094.
Frei, Frances X. 2006. *“*Breaking the Trade-off between Efficiency and Service.” *Harvard Business Review 84(11):* 92-101.
Fredendall, Lawrence, Janet Craig, Pat Fowler, and Uzay Damali. 2009. “Barriers to swift, even flow in the internal supply chain of perioperative surgical services department: a case study.” *Decision Sciences* 40(2): 327–349
García-Dastugue, Sebastián and Douglas Lambert. 2007. “*Interorganizational Time-Based Postponement in the Supply Chain.” Journal of Business Logistics 28(1):* 57-81.
Giesberts, Paul and Laurens Van der Tang. 1992. “Dynamics of the Customer Order Decoupling Point: Impact on Information Systems for Production Control.” *Production Planning and Control* 3(3): 300–313.
Guven-Uslu, Pinar, Hing Chan, Sadia Ijaz, Ozlem Bak, Barry Whitlow and Vikas Kumar. 2014. *“*In-depth Study of ‘Decoupling Point’ as a Reference Model: an Application for Health Service Supply Chain.” *Production Planning & Control* 25(13-14):1107-1117.
Harvey, Jean. 1998. “Service Quality: a Tutorial.” *Journal of Operations Management* 16(5): 583-597.
Hoekstra, Sjoerd and Jac Romme. 1992*. Integral Logistic Structures: Developing Customer-oriented Goods Flow*, New York: Industrial Press.
Kellar, Gregory M., George G. Polak and Xinhui Zhang. 2016. “Synchronization, cross-docking, and decoupling in supply chain networks.” *International Journal of Production Research* 54(9): 2585-2599.Kellogg, Deborah and Winter Nie. 1995. “A Framework for Strategic Service Management.” *Journal of Operations Management* 13(4): 323-337.
Levitt, Theodore. 1972. “Production-line Approach to Service.” *Harvard Business Review* 50(5): 41-52.
Liu, Weihua, Yuming Mo, Yi Yang and Zi Ye. 2015. “Decoupling point on multiple customer demands in logistics service supply chain.” *Production Planning & Control*, 26 (3): 178-202.

Liu, Weihua, Zhicheng, Liang, Zi Ye and Liang Liu. 2016. “The optimal decision of customer order decoupling point for order insertion scheduling in logistics service supply chain.” *International Journal of Production Economics*, 175: 50-60.

Maister, David and Christopher Lovelock. 1982. “Managing Facilitator Services.” *Sloan Management Review* 23: 19–31.

Maull, Roger, Andi Smart, and Liang Liang 2014. “A Process Model of Product Service Supply Chains.” *Production Planning & Control* 25 (13-14): 1091-1106.

McFadden, Kathleen, Gregory Stock and Charles Gowen III, 2006. “Implementation of patient safety initiatives in US hospitals.” *International Journal of Operations & Production Management* 26: 326–347

Metters, Richard and Vicente Vargas. 2000. “A Typology of Decoupling Strategies in Mixed Services.” *Journal of Operations Management* 18(6): 663-682.

Moeller, Sabine. 2008. “Customer Integration—A Key to an Implementation Perspective of Service Provision.” *Journal of Service Research* 11: 197-210.

Naylor, Ben, Mohamed Naim and Danny Berry. 1999. “Leagility: Integrating the Lean and Agile Manufacturing Paradigms in the Total Supply Cain.” *International Journal of Production Economics* 62 (1-2): 107-118.

Olhager, Jan. 2003 “Strategic Positioning of the Order Penetration Point.” *International Journal of Production Economics* 85(3): 319-329.

Olhager, Jan and Björn Östlund. 1990. “An Integrated Push–pull Manufacturing Strategy.” *European Journal of Operational Research* 45(2–3): 135–142.

Pagh, Janus and Martha Cooper. 1998. “Supply Chain Postponement and Speculation Strategies: How to Choose the Right Strategy.” *Journal of Business Logistics* 19: 13-33.

Palmer, Adrian and Catherine Cole. 1995. *Services Marketing: Principles and Practice.* Englewood Cliffs, NJ: Prentice-Hall.

Parnaby, John., and Dennis Towill. 2008. “Seamless healthcare delivery system.” *International Journal of Health Care Quality Assurance* 21(3): 249-273.

Pawar, Kulwant, Ahmad Beltagui and Johann Riedel, (2009). “The PSO triangle: designing product, service and organisation to create value.” *International Journal of Operations & Production Management* 29(5): 468-493.

Ponsignon, Frederic, P. Andi Smart and Roger Maull (2011). “Service delivery system design: characteristics and contingencies.” *International Journal of Operations & Production Management* 31(3): 324 – 349.

Pyke, David and Morris Cohen. 1990. “Push and Pull in Manufacturing and Distribution Systems.” *Journal of Operations Management* 9(1): 24-43.

Rahimnia, Farubirz and Mahdi Moghadasian. 2010. “Supply Chain Leagility in Professional Services.” *Supply Chain Management: An International Journal* 15(1): 80-91.

Sampson, Scott. E. 2000. “Customer-supplier Duality and Bidirectional Supply Chains in Service Organizations”, *International Journal of Service Industry Management* 11 (4): 348-364.

Sampson, Scott. E. and Craig M. Froehle. 2006. “Foundations and Implications of a Proposed Unified Services Theory.” *Production and Operations Management* 15: 329-343.

Schmenner, Roger and Morgan Swink. 1998. “On theory in operations management.” *Journal of Operations Management* 17(1): 97–113.

Sharman, Graham. 1984. “The Rediscovery of Logistics.” *Harvard Business Review* 62: 71–79.

Shostack, Lynne. 1984. "Designing Services that Deliver" *Harvard Business Review*, 62(1): 133-139.
Silvestro, Rhian, Lin Fitzgerald, Robert Johnston and Chris Voss. 1992. "Towards a Classification of Service Processes." *International Journal of Service Industry Management* 3(3): 62–75.
Skinner, Wickham. 1974. "The Focused Factory." *Harvard Business Review* 52 (3): 113-121.
Smith, Adam. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations.* London: Strahan & Cadeli.
Spearman, Mark and Michael Zazanis. 1992. "Push and Pull Production Systems: Issues and Comparisons." *Operational Research* 40: 521–532.
Taylor, Frederick Winslow. 1911. *The Principles of Scientific Management*. New York: Harper and Brothers.
Taylor, Margaret, and Andrew Talyor. 2014, "Performance Measurement in the Third Sector: The Development of a Stakeholder-focussed Research Agenda." *Production Planning & Control* 25 (16): 1370-1385.
Thompson, James. 1967. *Organizations in Action.* New York: McGraw-Hill.
Towill, Denis. 1997. "*Forridge - principles of good practice in material flow.*" *Production Planning and Control* 8(7): 622-632.
Towill, Denis and Martin Christopher. 2005. "An evolutionary approach to the architecture of effective healthcare delivery systems." *Journal of Health Organisation and Management* 19(2): 130-147.
Vandermerwe, Sandra and Juan Rada. 1988. "Servitization of Business: Adding Value by Adding Services." *European Management Journal,* 6 (4): 314-324.
van Donk, Dirk Pieter and Ron van Doorne. 2016. "The Impact of the Customer Order Decoupling Point on Type and Level of Supply Chain Integration." *International Journal of Production Research,* 54 (9): 2572–2584.
Vanteddua, Gangaraju and Ratna Babu Chinnamb. 2014. "Supply Chain Focus Dependent Sensitivity of the Point of Product Differentiation." *International Journal of Production Research,* 52 (17): 4984-5001.
Vargo, Stephen and Rober Lusch. 2004. "The Four Service Marketing Myths – Remnants of a Goods-based, Manufacturing Model." *Journal of Service Research* 6 (4): 324–335.
Wikner, Joakim. 2012. "A Service Decoupling Point Framework for Logistics, Manufacturing, and Service Operations."*, International journal of services sciences,* 4(3), 330-357.
Wikner, Joakim. 2014. "On Decoupling Points and Decoupling Zones." *Production & Manufacturing Research* 2: 167-215.
Wikner, Joakim and Jenny Bäckstrand. 2012. "Decoupling points and product uniqueness impact on supplier relations". 19th EurOMA Conference.
Wikner, Joakim and Sayeh Noroozi. 2016. "A modularised typology for flow design based on decoupling points – a holistic view on process industries and discrete manufacturing industries", *Production Planning & Control*. 27(16): 1344–1355.
Wikner, Joakim and Martin Rudberg. 2005. "Introducing a Customer Order Decoupling Zone in Logistics Decision-making." *International Journal of Logistics Research and Applications* 8(3): 211-224.
Womack, James and Daniel Jones. 1996, "Beyond Toyota: How to Root Out Waste and Pursue Perfection." *Harvard Business Review* 74(5): 140-153.

Yang, Biao, Ying Yang, and Sharon Williams, 2010. “Service Postponement: Translating Manufacturing Postponement to Service Operations.”, *Journal of Manufacturing Technology Management* 21: 470-483.
Zomerdijk, Leonieke and Jan de Vries. 2007. “Structuring Front Office and Back Office Work in Service Delivery Systems.” *International Journal of Operations and Production Management* 27(1): 108 – 131.