

## PRODUÇÃO ENXUTA

### INTRODUÇÃO

Este capítulo examina uma abordagem para gerenciar a produção e a cadeia de suprimento, normalmente chamada de “enxuta”. É tanto filosofia quanto um método de planejamento e controle, bem como uma abordagem de melhoramento. A produção enxuta visa atender instantaneamente a demanda, com qualidade perfeita, sem desperdício e com baixo custo. Isso envolve fornecer produtos e serviços em sincronização perfeita com a demanda por eles. Esses princípios foram uma ruptura radical na prática tradicional de administração da produção, mas agora se tornaram ortodoxos para promover a sincronização do fluxo ao longo dos processos, da produção e das redes de suprimento. E, embora esse assunto já tenha sido tratado estritamente como um fenômeno da fabricação (talvez isso não seja novidade, devido ao papel pioneiro da Toyota na gestão enxuta), os princípios da produção enxuta se aplicam a todos os setores, incluindo finanças, saúde, TI, varejo, construção, agricultura e o setor público. A [Figura 15.1](#) posiciona a produção enxuta no modelo global de administração da produção.



#### Questões-chave

- O que é produção enxuta?
- Como a lógica enxuta elimina o desperdício?
- Como a lógica enxuta se aplica ao longo da rede de suprimento?
- Como a lógica enxuta se compara a outras abordagens?

**Figura 15.1** Este capítulo examina a produção enxuta.

## O QUE É PRODUÇÃO ENXUTA?

O foco da produção enxuta é alcançar um fluxo de materiais, informações ou clientes que ofereça exatamente o que os clientes desejam (qualidade perfeita), em quantidades exatas (nem muito nem pouco), exatamente quando for necessário (não antes nem depois), exatamente onde for necessário (no local certo) e com o menor custo possível. Esse é um conceito muito semelhante ao “*just-in-time*” (JIT), o “Toyota Production System” (TPS), “produção sem estoque” e “sincronização enxuta”. Ele resulta em itens fluindo rápida e regularmente através de processos, operações e redes de suprimento.

Neste capítulo, oferecemos exemplos de organizações em diversos setores que aplicam os aspectos da filosofia enxuta às suas operações e redes de suprimento. Muitos dos exemplos da filosofia e da sincronização enxuta no setor de serviço são diretamente análogos aos encontrados no setor manufatureiro porque os itens físicos são movidos ou processados de alguma forma. Por exemplo, os supermercados, geralmente, repõem seu estoque na loja apenas quando os clientes retiram produtos suficientes das prateleiras. O movimento de bens do depósito para a prateleira é acionado apenas pelo sinal de demanda da “prateleira vazia”. Construtoras cada vez mais criam uma regra de pedir materiais para seus canteiros apenas no dia anterior a quando o material for necessário. Isso reduz o acúmulo e as chances de roubo. Os dois são exemplos da aplicação de princípios de *controle puxado*. Outros exemplos dos conceitos e métodos enxutos se aplicam até mesmo quando quase todos os elementos do serviço são intangíveis. Por exemplo, novas tecnologias de publicação permitem que os professores montem material de estudo impresso e por *e-learning* ajustado às necessidades dos cursos individuais, ou mesmo a alunos individuais. Aqui, vemos os princípios enxutos da flexibilidade e lotes pequenos permitindo a customização e

a entrega rápida.

## OPERAÇÕES NA PRÁTICA

### Refeições “enxutas” de Jamie<sup>1</sup>

A maioria das pessoas não tem tempo para dedicar muito tempo cozinhando – possivelmente, o motivo pelo qual o *chef* celebridade Jamie Oliver escreveu o livro *Refeições de 30 minutos do Jamie*, cuja filosofia é que cozinhar um delicioso jantar deve ser tão rápido e mais barato do que comprar e aquecer uma comida pronta. O livro apresenta 50 menus prontos, com três a quatro pratos por menu, projetados para não demorar mais de 30 minutos no preparo. Para alcançar esse desempenho Jamie, talvez inadvertidamente, aplicou os princípios e métodos de produção enxuta à atividade cotidiana de cozinhar.

Imagine que sua família está chegando para jantar e você quer surpreender com uma nova refeição indiana de múltiplos pratos com frango, arroz, salada de um lado e, claro, uma sobremesa. Tradicionalmente, você procuraria quatro receitas diferentes, uma para cada prato. Como todas as receitas vêm de lugares diferentes, você precisa descobrir a quantidade de alimentos para comprar, fazer a matemática no caso de ingredientes compartilhados em todos os pratos, como alocar as vasilhas, panelas e outros equipamentos para os diferentes ingredientes e, o mais importante, você precisa descobrir em que ordem preparar as coisas, especialmente se quiser todos os seus pratos prontos ao mesmo tempo. A abordagem de Jamie reduz significativamente toda essa complexidade, garantindo que os pratos sejam preparados exatamente quando o próximo passo no processo precisar deles, independentemente do prato que for. Em outras palavras, os pratos não são cozidos em sequência, um após o outro, mas são preparados e concluídos simultaneamente.

Se identificarmos todas as tarefas relacionadas à preparação da salada (por exemplo, cortar os vegetais) com a letra A, cozinhar o arroz (por exemplo, misturar) com a letra B, cozinhar o frango com a letra C e finalmente fazer a sobremesa com a letra D, então, pela forma tradicional de cozinhar, nossa programação de tarefas se pareceria com algo assim: AAAA BBBBBBBB CCCCCC DDDD. Isso resulta em lotes, tempo de espera e pratos prontos antes que o jantar pudesse ser servido. Por outro lado, a culinária de 30 minutos de Jamie Oliver envolve tarefas de programação em uma sequência como ABCDACBADCABDC, onde tarefas únicas relacionadas a pratos diferentes seguem sem problemas, pois o *chef* corta um ingrediente de salada, depois mistura o arroz e, em seguida, corta mais ingredientes de salada enquanto o frango está sendo assado no forno e uma parte da sobremesa está sendo preparada. Desse modo, todos os pratos ficam prontos ao mesmo tempo, na hora certa, e nada é preparado antes de ser necessário, evitando qualquer tipo de desperdício. Uma técnica de programação nivelada assim é chamada *heijunka* (mistura de modelos) na abordagem enxuta.

Além disso, a culinária enxuta de Jamie se baseia em tempos de preparação reduzidos. No início de cada receita, o equipamento necessário para preparar o menu é apresentado sob o título “para começar”. Outras preparações necessárias, como o aquecimento do forno, também são especificadas. Ter todo o equipamento pronto desde o início economiza tempo no processo, e é, de acordo com Jamie, um pré-requisito para a preparação em 30 minutos. O uso de equipamentos simples, que são adequados para vários propósitos diferentes, também torna o processo mais rápido à medida que as trocas são reduzidas. O objetivo é tirar o máximo proveito do tempo disponível, eliminando a “chatice” na culinária (atividade sem valor agregado, no linguajar da administração da produção) e deixando apenas o que é estritamente “bom e rápido de cozinhar”, sem comprometer a qualidade.

---

## Três perspectivas da produção enxuta

Definir a produção enxuta não é uma pergunta direta porque, de várias maneiras, pode ser vista como três coisas relacionadas, porém distintas: uma filosofia, um método de planejamento e controle com prescrições úteis de como gerenciar operações cotidianas e um conjunto de ferramentas de melhoria.

- **Produção enxuta é uma filosofia de como conduzir a produção.** É um conjunto coerente de princípios fundamentados no fluxo regular ao longo dos processos ao fazer bem as coisas simples, gradualmente melhorando-as ou atendendo às necessidades dos clientes exatamente e (acima de tudo) eliminando o desperdício em cada etapa. Três questões-chave definem a filosofia enxuta: o envolvimento dos funcionários na operação, o impulso para melhoria contínua e a eliminação do desperdício. Outros capítulos examinam as duas primeiras questões; portanto, dedicamos a maior parte do restante deste capítulo à ideia central da eliminação do desperdício.
- **Produção enxuta é um método de planejar e controlar a produção.** Muitas ideias da abordagem enxuta referem-se a como os itens (materiais, informações, clientes) fluem ao longo da produção; e, mais especificamente, como os gerentes de produção podem gerenciar esse fluxo. Por essa razão, a abordagem enxuta pode ser vista como um método de planejamento e controle. Todavia, é a atividade de planejamento e controle que é feita em busca dos propósitos filosóficos da abordagem enxuta. O fluxo descoordenado causa imprevisibilidade, e a imprevisibilidade causa desperdício porque as pessoas mantêm estoque, capacidade física ou tempo para se proteger contra ele. Assim, o planejamento e controle enxutos usam diversos métodos para atingir um fluxo sincronizado e reduzir o desperdício. Acima de tudo, usa o controle “puxado” que foi anteriormente descrito no [Capítulo 10](#) (em contraste com o MRP, descrito no [Capítulo 14](#), que usa controle “empurrado”). Isso é geralmente atingido usando algum tipo de sistema *kanban* (descrito mais adiante). Além disso, os outros métodos de planejamento e controle enxutos que promovem um fluxo regular incluem programação e entrega niveladas e mistura de modelos (novamente descritos mais adiante neste capítulo).
- **Produção enxuta é um conjunto de ferramentas que melhora o desempenho da produção.** O “coração” da filosofia enxuta é uma coleção de ferramentas e técnicas de melhoria que são os meios para cortar o desperdício. Há muitas técnicas que podem ser denominadas “técnicas enxutas” e, novamente, muitas delas seguem de forma natural e lógica a partir da filosofia enxuta em geral. O que é também importante entender é como a introdução da abordagem enxuta como filosofia ajudou a mudar o foco da administração da produção, geralmente para ver a melhoria como seu principal propósito. Além disso, a ascensão das ideias enxutas influenciou técnicas que agora se tornaram dominantes na administração da produção. Algumas dessas ferramentas e técnicas são bem conhecidas

fora da esfera da abordagem enxuta e estão incluídas em outros capítulos deste livro.



### Princípio de administração da produção

A abordagem enxuta pode ser vista como uma filosofia de como conduzir a produção, um método de planejamento e controle da produção e um conjunto de ferramentas que melhoram o desempenho da produção.

## OPERAÇÕES NA PRÁTICA

### Pixar adota a produção enxuta

Parece que os princípios enxutos (ou alguns deles) podem ser aplicados até mesmo aos processos mais improváveis. Nenhum é menos provável que os estúdios de animação Pixar, o estúdio de animação por computador premiado pela academia e os criadores de filmes que resultaram em uma série sem precedentes de sucessos de crítica e de bilheteria, incluindo *Toy Story* (1, 2 e 3), *Vida de Inseto*, *Monstros e Companhia*, *Procurando Nemo*, *Os Incríveis*, *Ratatouille*, *WALL-E* e *Up*. Desde a sua incorporação, a Pixar foi responsável por muitos avanços importantes na aplicação da computação gráfica (CG) na produção de filmes. Assim, a empresa atraiu alguns dos melhores talentos técnicos, criativos e de produção do mundo na área. E tal talento “baseado no conhecimento” é notoriamente difícil de administrar – certamente, não são os tipos de processos geralmente considerados como apropriados para a produção enxuta. A gestão da criatividade envolve um dilema difícil entre encorajar a liberdade de produzir novas ideias e garantir que elas funcionem dentro de uma estrutura geral eficaz.

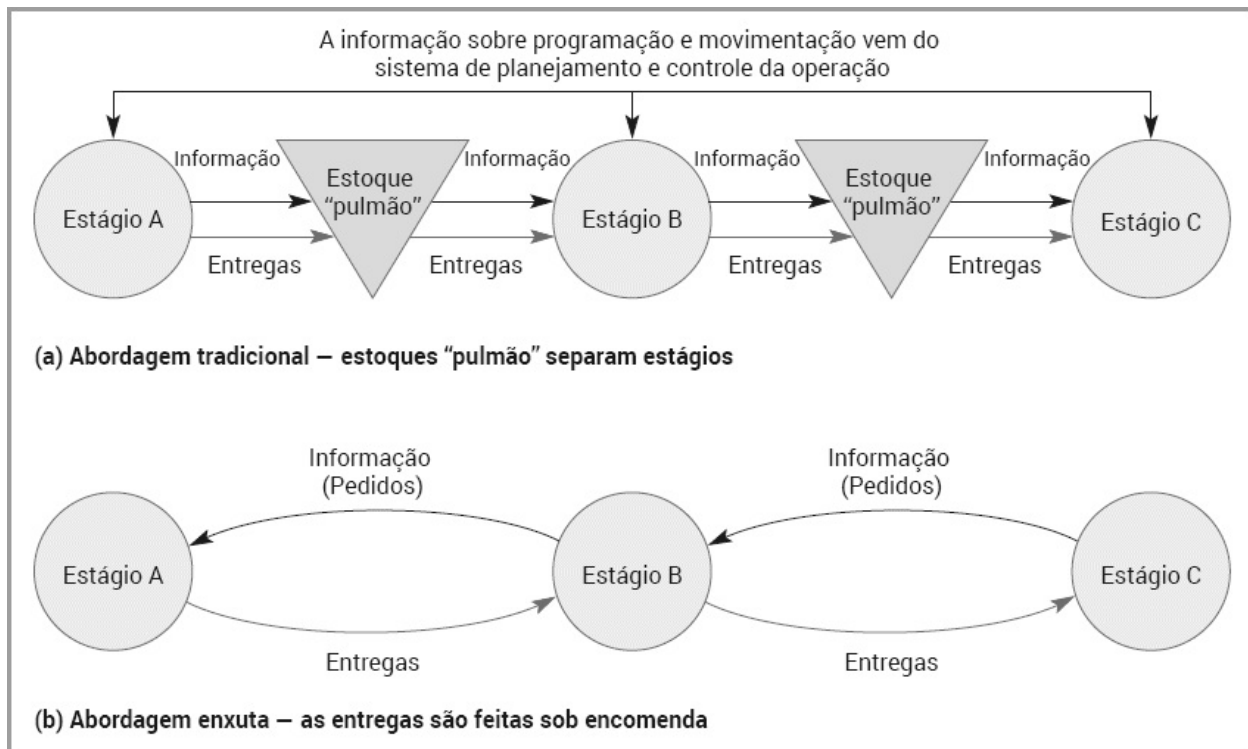
No entanto, a Pixar obteve inspiração da Toyota e da forma como ela usa a produção enxuta – em particular, a maneira pela qual a Toyota incentivou sugestões e críticas contínuas de seus trabalhadores da linha de produção para melhorar seu desempenho. A Pixar percebeu que poderia fazer o mesmo com a produção de personagens de desenho animado. A adoção de *feedback* constante evidencia os problemas antes de se tornarem crises, e oferece inspiração e desafio às equipes criativas. A Pixar também dedica muito esforço para persuadir seu pessoal criativo a trabalharem juntos. Em empresas semelhantes, as pessoas podem colaborar em projetos específicos, mas não são tão boas em se concentrar no que está acontecendo em outras partes do negócio. A Pixar, no entanto, tenta cultivar um senso de responsabilidade coletiva. A equipe até mesmo mostra o trabalho inacabado entre si nas reuniões diárias, acostumando-se então a oferecer e receber críticas construtivas.

## Como a produção enxuta considera o fluxo

A melhor forma de entender como a sincronização enxuta difere das abordagens mais tradicionais para gerenciar o fluxo é comparar os dois processos simples da [Figura 15.2](#). A abordagem tradicional assume que cada estágio do processo colocará sua produção em um estoque que “isola” esse estágio do estágio adiante (a jusante, ou *downstream*) no processo. O estágio seguinte irá (por fim) suprir-se dos componentes desse estoque, processá-los e enviá-los ao próximo estoque “pulmão”. Esses estoques existem para desacoplar cada estágio dos estágios vizinhos, tornando cada um deles relativamente independente, de modo que, por exemplo, se o

estágio A interrompe sua produção por algum motivo, o estágio B pode continuar trabalhando, ao menos por algum tempo. Quanto maior o estoque “pulmão”, maior o grau de isolamento entre os estágios. Esse isolamento é obtido ao custo da estocagem e do lento tempo de atravessamento porque os produtos, clientes ou informações ficarão algum tempo esperando entre os estágios no processo.

Contudo, o principal argumento contra essa abordagem tradicional recai sobre a própria condição que ela visa promover, ou seja, o isolamento de um estágio em relação ao outro. Quando um problema ocorre em um estágio, o problema não ficará imediatamente aparente em qualquer ponto do sistema. A responsabilidade pela solução do problema estará largamente centrada nas pessoas desse estágio, e as consequências do problema estarão impedidas de propagar a todo o sistema. Entretanto, compare isso com o processo enxuto puro ilustrado na [Figura 15.2](#). Aqui, os itens são processados e depois transferidos diretamente para o próximo estágio de modo sincronizado *just-in-time*, para serem novamente processados. Os problemas em qualquer estágio têm efeito muito diferente em tal sistema. Agora, se o estágio A parar de processar, o estágio B perceberá imediatamente e o estágio C logo após. O problema do estágio A está agora rapidamente exposto ao processo global, que é imediatamente afetado pelo problema. Isso significa que a responsabilidade pela solução do problema não está mais confinada aos funcionários do estágio A. Está agora compartilhada por todos, melhorando consideravelmente as chances de o problema ser resolvido, nem que seja porque agora ele é muito importante para ser ignorado. Em outras palavras, ao evitar a acumulação de itens entre os estágios, a operação aumentou as chances de a eficiência intrínseca da fábrica ser melhorada. As abordagens não sincronizadas procuram encorajar a eficiência ao proteger cada parte do processo da interrupção. A abordagem enxuta assume a visão oposta. A exposição do sistema (embora não repentinamente, como em nosso exemplo simplificado) aos problemas pode torná-los mais evidentes e mudar a “estrutura de motivação” de todo o sistema em direção à solução dos problemas. A abordagem enxuta vê as acumulações de estoque, seja ele de produtos, clientes ou informação, como um “manto” estendido sobre o sistema de produção, evitando que os problemas sejam descobertos.



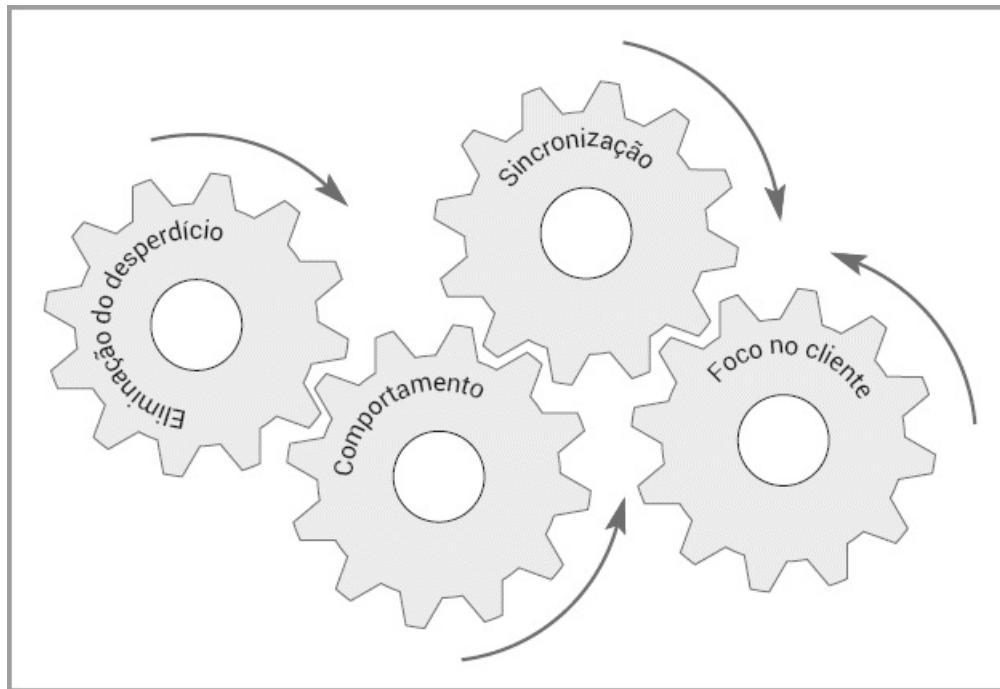
**Figura 15.2** Fluxo tradicional (a) e na abordagem enxuta (b) entre os estágios.



### Princípio de administração da produção

O estoque “pulmão”, utilizado para isolar estágios ou processos, detecta a motivação com o objetivo de implantar melhorias.

Pense agora na lógica por trás da explanação anterior. Há quatro ideias inter-relacionadas que se misturam para formar uma cadeia lógica. Primeiro, entre cada estágio, é o estágio “cliente” a jusante (*downstream*) que sinaliza a necessidade de ação. É o cliente que, com efeito, “puxa” os itens ao longo do processo. O ponto de partida da filosofia enxuta é um foco no cliente. Segundo, essa “puxada” encoraja o fluxo dos itens ao longo do processo de modo sincronizado (em vez de mantê-los no estoque). Terceiro, o fluxo sincronizado suave e a resultante redução no estoque afeta a motivação para melhorar, porque os estágios não mais estão desacoplados. Quarto, o aumento da motivação a melhorar expõe o desperdício e encoraja sua eliminação. Essas quatro ideias relacionadas estão ilustradas na [Figura 15.3](#), que serão mais discutidas no restante deste capítulo.



**Figura 15.3** Os quatro elementos da abordagem enxuta.

## Como a produção enxuta considera o estoque

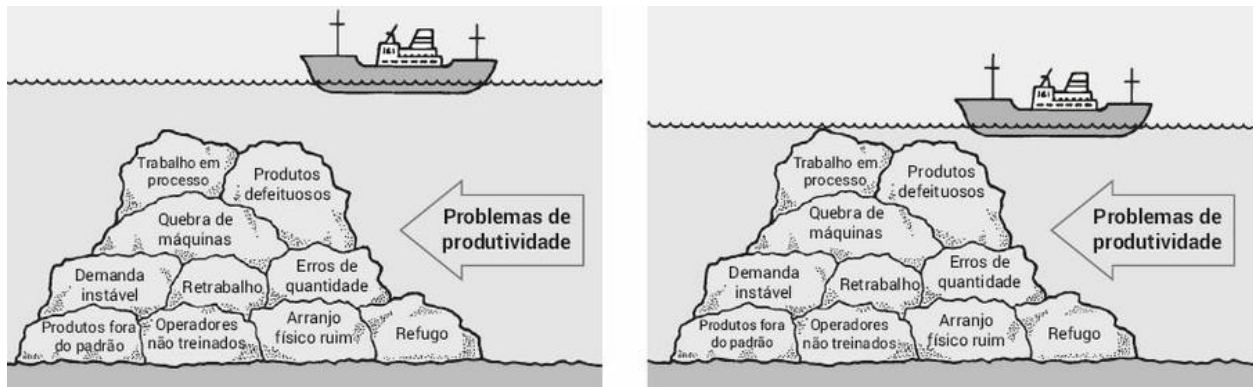
A ideia dos efeitos obscuros do estoque é frequentemente ilustrada em forma gráfica, como na [Figura 15.4](#). Os muitos problemas da operação são mostrados como rochas no leito de um rio, que não podem ser vistas em razão da profundidade da água. Nessa analogia, a água representa o estoque na produção. Todavia, muito embora as rochas não possam ser vistas, retardam o progresso do fluxo do rio e causam turbulência. A redução gradual da profundidade da água (estoque) expõe o pior dos problemas que podem ser resolvidos, após o que a água baixa ainda mais, expondo mais problemas, e assim por diante. O mesmo argumento também se aplicará ao fluxo entre os processos ou as operações completas. Por exemplo, os estágios A, B e C da [Figura 15.2](#) podem ser uma operação do fornecedor, uma operação focal e uma operação do cliente, respectivamente.

## Como a produção enxuta considera a utilização de capacidade

A sincronização enxuta tem muitos benefícios, mas eles ocorrem ao custo da pior utilização da capacidade física. Retornemos ao processo mostrado na [Figura 15.2](#). Quando paradas ocorrem no sistema tradicional, o “pulmão” de estoque permite que cada estágio continue trabalhando e, assim, atinja a utilização de alta capacidade. No entanto, a alta utilização não necessariamente leva o processo como um todo a produzir mais. Frequentemente, a “produção” extra está nos estoques “pulmões” ou nas filas de clientes. Em um processo enxuto, qualquer parada afetará



todo o processo. Isso, necessariamente, levará à menor utilização de capacidade física, pelo menos a curto prazo. Em organizações que dão alto valor à utilização da capacidade física, isso pode ser particularmente difícil de aceitar. Entretanto, não faz sentido produzir apenas por conta própria. De fato, trabalhar apenas para manter a alta utilização não é indicado e passa a ser contraproducente, porque o estoque extra produzido serve simplesmente para tornar as melhorias menos prováveis. A [Figura 15.5](#) ilustra as duas abordagens de utilização da capacidade física.



**Figura 15.4** A redução do nível de estoque (água) permite que a administração da produção (o navio) veja os problemas da operação (rochas) e trabalhe para reduzi-los.



### Princípio de administração da produção

Focar em produção enxuta pode reduzir inicialmente a utilização de recursos.

**Figura 15.5** As diferentes visões da utilização de capacidade na (a) abordagem tradicional e na (b) abordagem enxuta para operações.

## Como a produção enxuta considera o papel das pessoas

Frequentemente, os proponentes da operação enxuta reforçam a importância do envolvimento de todos os funcionários na abordagem enxuta. Desse modo, é similar a outros conceitos baseados em melhoria, como “qualidade total”, que é discutida posteriormente em detalhes (veja o [Capítulo 17](#)). Entretanto, a abordagem enxuta no gerenciamento de pessoas é muito influenciada por sua origem japonesa, o que é evidente pela terminologia e pelos próprios conceitos. Assim, por exemplo, os defensores da operação enxuta original chamavam essa abordagem de sistema de “respeito aos seres humanos”. Encoraja (e, frequentemente, requer) a solução de problemas baseada em equipe, o enriquecimento de cargo (ao incluir tarefas de manutenção e de preparação nos cargos dos operadores), a rotação de tarefas e as habilidades múltiplas. A intenção é encorajar um alto grau de responsabilidade pessoal, engajamento e “posse” da tarefa. De modo semelhante, as denominadas “práticas básicas de trabalho” são, às vezes, usadas para implementar o princípio do “envolvimento de todos”. Essas práticas incluem:

- **Disciplina** – Padrões de trabalho que são cruciais para a segurança dos funcionários, para o meio ambiente, bem como para a qualidade, devem ser seguidos por todos e em todo o tempo.
- **Flexibilidade** – Deverá ser possível expandir as responsabilidades de acordo com a capacitação das pessoas. Isso se aplica tanto a gerentes quanto ao pessoal do chão de fábrica. Devem ser removidas as barreiras para a flexibilidade, como estruturas de avaliação e práticas restritivas.
- **Igualdade** – Políticas de pessoal injustas e divergentes devem ser descartadas. Muitas empresas implementam a mensagem de tratamento igualitário por meio de uniformes da empresa, estruturas de pagamento consistentes, que não diferenciam entre o pessoal de tempo integral e os que recebem por hora trabalhada, além de escritórios abertos.
- **Autonomia** – Delegar responsabilidade a pessoas envolvidas nas atividades diretas, de modo que a tarefa da gerência torna-se apoiar processos. A delegação inclui dar ao pessoal a responsabilidade por interromper processo no caso de problemas, programar o trabalho, colher dados de monitoramento de desempenho e solução de problemas gerais.
- **Desenvolvimento de pessoal** – Com o tempo, o objetivo é criar mais membros da força de trabalho da empresa que possam suportar o rigor de serem mais competitivos.
- **Qualidade de vida no trabalho** – Pode incluir, por exemplo, o envolvimento dos funcionários na tomada de decisão, na segurança do trabalho e no prazer de trabalhar nas instalações da empresa.
- **Criatividade** – Este é um dos elementos indispensáveis da motivação. A criatividade neste contexto significa não apenas realizar um trabalho, mas também melhorar o modo como ele é feito e adaptar a melhoria ao processo.

- **Envolvimento total das pessoas** – Os funcionários assumem mais responsabilidade ao usar suas capacitações para beneficiar a empresa como um todo. Eles deverão participar de atividades, como seleção de novos contratados, envolvimento direto com fornecedores e clientes sobre as programações, problemas de qualidade e informação sobre entregas, gastos do orçamento com melhorias, e planejamento e revisão do trabalho diário mediante reuniões de comunicação.

## OPERAÇÕES NA PRÁTICA

### DNA enxuto da Toyota

Vista como principal praticante e criadora da abordagem enxuta, a Toyota Motor Company vem sincronizando, de forma progressiva e simultânea, todos os seus processos para atingir alta qualidade, tempos rápidos de atravessamento e excepcional produtividade. Tem feito isso ao desenvolver um conjunto de práticas que moldaram em grande parte o que hoje denominamos de “produção enxuta” ou *just-in-time* (JIT), mas que a Toyota denomina “Sistema Toyota de Produção” (STP). O STP possui dois temas: “*just-in-time*” e “*jidoka*”. *Just-in-time* é definido como movimentação rápida e coordenada de componentes ao longo do sistema de produção e rede de suprimento para atender à demanda do cliente. É operacionalizado por meio de *heijunka* (nivelamento e suavização do fluxo de itens), *kanban* (sinalização para o processo anterior de que mais componentes são necessários) e *nagare* (posicionamento de processos para alcançar o fluxo de componentes mais regular ao longo da produção). *Jidoka* é descrito como “humanização da interface entre operador e máquina”. A filosofia da Toyota é de que a máquina está disponível para servir ao propósito do operador. O operador deve ser deixado livre para exercer seu julgamento. *Jidoka* é operacionalizado por meio de mecanismos “à prova de falhas” (ou *jidoka* mecânico), autoridade para interromper a linha de montagem (*jidoka* humano) e controle visual (*status* visual dos processos de produção e visibilidade dos padrões do processo).

A Toyota acredita que tanto o *just-in-time* como o *jidoka* deveriam ser aplicados fortemente para a eliminação de desperdício, definindo desperdício como “tudo além da mínima quantidade de equipamento, itens, partes e trabalhadores que são absolutamente essenciais à produção”. Fujio Cho, da Toyota, identificou sete tipos de desperdício que precisam ser eliminados de todos os processos de produção. São eles: desperdício de produção em excesso, desperdício de tempo de espera, desperdício de transporte, desperdício de estoque, desperdício de processamento, desperdício de movimentação e desperdício de produtos defeituosos. Além disso, autoridades da Toyota afirmam que seu ponto forte está no entendimento das diferenças entre instrumentos e práticas usados nas operações da Toyota e na filosofia geral de sua abordagem enxuta. Essa é a razão pela qual alguns identificaram o aparente paradoxo do STP da Toyota, “a saber, que atividades, conexões e fluxos de produção em uma de suas fábricas seguem roteiros rígidos, ao mesmo tempo em que as operações são altamente flexíveis e adaptáveis. As atividades e os processos estão sendo constantemente desafiados e empurrados para um nível mais alto de desempenho, possibilitando à empresa inovar e aprimorar continuamente.” Embora alguns dos que adotam os princípios da abordagem enxuta possam pensar que “são enxutos”, a Toyota simplesmente muda seu objetivo constantemente para desafiar a melhoria. Dessa forma, notamos uma distinção importante entre aqueles que veem a abordagem enxuta como um ponto final específico, a ser alcançado através de uma série de ferramentas de melhoria, e aqueles como a Toyota, que tratam a abordagem enxuta como uma filosofia que define um modo de realização de negócios.

---



### Comentário crítico

Nem todos os comentaristas veem as práticas de gestão de pessoas influenciadas pela abordagem enxuta como totalmente positivas. Provavelmente, a maior crítica da abordagem enxuta seja o efeito potencialmente prejudicial, sobre os trabalhadores, que poderia ter esse foco constante na melhoria. Além disso, a ênfase na redução do desperdício costuma ignorar outros indicadores importantes, como a responsabilidade social corporativa e a qualidade de vida no trabalho. Embora menos autocrática do que algumas práticas de gestão japonesas de antigamente, a abordagem enxuta certamente não está alinhada a algumas das filosofias de projeto do trabalho que dão maior ênfase à contribuição e ao comprometimento, como descritos no **Capítulo 9**. Por fim, como exploramos em nosso capítulo sobre risco e recuperação, em face da maior incerteza, as operações enxutas podem enfrentar níveis mais altos de exposição a eventos de falha.

## Como a produção enxuta considera a melhoria

Os objetivos da abordagem enxuta são frequentemente expressos como ideais, como em nossa definição anterior: “atender à demanda no momento exato, na qualidade perfeita e sem desperdício”. Ainda que o desempenho atual de qualquer organização possa estar longe desses ideais, uma crença fundamental dessa abordagem é a de que é possível aproximar-se deles ao longo do tempo. Sem essas crenças para guiar o progresso, os proponentes da abordagem enxuta afirmam que a melhoria tem mais chance de ser transitória do que contínua. É por isso que o conceito de melhoria contínua é parte tão importante da filosofia enxuta. Se seus objetivos são estabelecidos em termos de ideais, os quais as organizações individuais nunca podem alcançar, a ênfase então deve estar na forma com a qual uma organização aproxima-se desse estado ideal. A palavra japonesa para melhoria contínua é *kaizen* e ela é uma parte essencial da filosofia enxuta.

## COMO A PRODUÇÃO ENXUTA ELIMINA O DESPERDÍCIO?

Comprovadamente, a parte mais significativa da filosofia enxuta é seu foco na eliminação de todas as formas de desperdício. Desperdício pode ser definido como qualquer atividade que não agrega valor. Por exemplo, estudos frequentemente mostram que não mais do que 5% do tempo total de atravessamento é realmente gasto diretamente agregando valor. Isso significa que, durante 95% de seu tempo, uma operação está agregando custo ao serviço ou produto, mas não agregando valor. Tais cálculos podem alertar, mesmo as operações relativamente eficientes, para o enorme desperdício que está escondido em todas as operações. Esse mesmo fenômeno aplica-se tanto aos processos de serviço como aos de fabricação. Solicitações relativamente simples, como tirar uma carteira de habilitação, podem demorar apenas alguns minutos para processar, embora levem dias (ou semanas) para se chegar ao seu resultado final.



### Causas de desperdício – *muda, mura, muri*

Na filosofia enxuta, termos japoneses são frequentemente usados para descrever ideias essenciais. A eliminação de desperdício é certamente uma ideia essencialmente enxuta. Os termos *muda*, *mura* e *muri* são palavras japonesas que transmitem a ideia de três causas dos desperdícios que devem ser reduzidas ou eliminadas.

- *Muda*. São as atividades em um processo que são desperdiçadoras porque não agregam valor à operação ou ao cliente. As principais causas dessas atividades desperdiçadoras são, provavelmente, os objetivos mal comunicados (incluindo o não entendimento das exigências do cliente) ou o uso ineficiente dos recursos. A implicação disso é que, para qualquer atividade ser efetiva, deve ser adequadamente registrada e comunicada a quem quer que a esteja desempenhando.
- *Mura*. Significa “falta de consistência” ou irregularidades que resultam em sobrecarga periódica de pessoal ou equipamentos. Assim, por exemplo, se as atividades não forem adequadamente documentadas para que diferentes pessoas em ocasiões diferentes desempenhem a mesma tarefa de modo diferente, não surpreende que o resultado da atividade possa ser diferente. Os efeitos negativos disso são semelhantes à falta de confiabilidade (veja o [Capítulo 2](#)).
- *Muri*. Significa o absurdo ou a irracionalidade. Está baseado na ideia de que as exigências desnecessárias ou irracionais colocadas em um processo resultarão em maus resultados. A implicação disso é que habilidades apropriadas, planejamento efetivo, estimativa precisa de tempo e programação evitarão esse desperdício de sobrecarga “*muri*”. Em outras palavras, o desperdício pode ser causado pela falta de condução de tarefas de planejamento das operações básicas, como priorizar atividades (sequenciamento), entender o tempo necessário (programação) e recursos (carregamento) para desempenhar atividades. Todos esses assuntos são discutidos no [Capítulo 10](#).

Essas três causas de desperdício estão obviamente relacionadas. Quando um processo é inconsistente (*mura*), pode levar à superutilização de equipamentos e pessoas (*muri*), que, por sua vez, causarão todos os tipos de atividades que não agregam valor (*muda*).

### Tipos de desperdício

*Muda*, *mura* e *muri* são três causas de desperdício. Agora, vejamos os tipos de desperdício,

que têm aplicação em muitos tipos diferentes de operações – tanto de serviço quanto de produção – e que formam a essência da filosofia enxuta. Aqui, consolidamos isso em quatro categorias globais.

### *Desperdício pelo fluxo irregular*

A sincronização perfeita significa o fluxo regular e uniforme através de processos, operações e redes de fornecimento. As barreiras que impedem o fluxo simplificado incluem o seguinte:

- *Tempo de espera.* A eficiência do equipamento e da mão de obra são duas medidas populares amplamente usadas para medir o tempo de espera de equipamento e de pessoas, respectivamente. Menos óbvio é o tempo de espera por produtos, clientes ou informações, dissimulados pelos operadores, que são mantidos ocupados produzindo peças ou componentes desnecessários naquele momento.
- *Transporte.* Movimentar itens ou clientes em torno da operação, junto com a manipulação dupla e tripla de componentes, não agregam valor. Mudanças do arranjo físico que aproximam os processos e melhoram os métodos de transporte e a organização do local de trabalho podem reduzir o desperdício.
- *Ineficiências do processo.* O próprio processo pode ser uma fonte de desperdício. Algumas operações podem existir somente devido ao mau projeto do componente ou da má manutenção e, assim, podem ser eliminadas.
- *Estoque.* Independentemente do tipo (produto, cliente, informação), qualquer estoque deve tornar-se alvo de eliminação. Entretanto, é apenas combatendo as causas do estoque ou filas, como fluxo irregular, que ele pode ser reduzido.
- *Movimentação desperdiçada.* Um operador pode parecer ocupado, mas, às vezes, nenhum valor está sendo agregado ao trabalho. A simplificação do trabalho é uma fonte rica de redução do desperdício da movimentação.

### *Desperdício pelo fornecimento inexato*

A sincronização perfeita também significa fornecer exatamente o que é desejado, exatamente quando é necessário. Qualquer falta ou excesso de fornecimento e qualquer entrega antecipada ou atrasada resultará em desperdício, algo que já exploramos no capítulo de gestão da capacidade física em particular. Barreiras para alcançar uma combinação exata entre suprimento e demanda incluem o seguinte:

- *Superprodução ou subprodução.* Produzir mais ou menos do que for imediatamente necessário para o próximo estágio, processo ou operação. (Esta é a maior fonte de

desperdício, segundo a Toyota.)

- *Entrega antecipada ou atrasada.* Os itens só deverão chegar exatamente quando forem necessários. A entrega antecipada é um desperdício, assim como a entrega atrasada.
- *Estoque.* Mais uma vez, todos os estoques deverão se tornar um alvo para eliminação. Porém, somente enfrentando as causas do estoque, como o suprimento inexato, é que ele poderá ser reduzido.

### *Desperdício pela resposta inflexível*

As necessidades do cliente podem variar, em termos do que eles querem, quanto eles querem e quando eles querem. No entanto, os processos geralmente acham mais conveniente mudar o que eles fazem com pouca frequência, porque cada alteração implica algum tipo de custo. É por isso que os hospitais agendam clínicas especializadas apenas em momentos específicos e por que as máquinas muitas vezes fazem um lote de produtos semelhantes juntos. No entanto, responder às demandas dos clientes precisa e instantaneamente requer um alto grau de flexibilidade do processo. Sintomas de flexibilidade inadequada incluem o seguinte:

- *Grandes lotes.* Enviar lotes de itens através de um processo inevitavelmente aumenta o estoque à medida que o lote se movimenta pelo processo inteiro.
- *Atrasos entre atividades.* Quanto maior o tempo (e o custo) para mudar de uma atividade para outra, mais difícil é sincronizar o fluxo para corresponder instantaneamente à demanda do cliente.
- *Mais variação no mix de atividades do que na demanda do cliente.* Se o *mix* de atividades nos diferentes períodos de tempo varia mais do que a demanda do cliente, então deve estar ocorrendo algum “agrupamento” (ou formando-se algum lote) de atividades.

### *Desperdício pela variabilidade*

A sincronização implica níveis exatos de qualidade. Se houver variabilidade nos níveis de qualidade, os clientes não se considerarão adequadamente supridos. A variabilidade, portanto, é uma barreira importante para alcançar o suprimento sincronizado. Os sintomas da baixa variabilidade incluem o seguinte:

- *Pouca confiabilidade do equipamento.* Equipamento pouco confiável normalmente indica uma falta de conformidade com os níveis de qualidade. Também significa que haverá irregularidade no suprimento dos clientes. De qualquer forma, isso impede a sincronização do suprimento.
- *Produtos ou serviços defeituosos.* O desperdício causado pela má qualidade é significativo na maior parte das operações. Os erros em produto ou serviço causam perda de tempo para



clientes e processos, até que sejam corrigidos.

## Procurando o desperdício (e as oportunidades *kaizen*) – *gemba*

*Gemba* (às vezes, também denominado “*gamba*”), quando mais ou menos traduzido do japonês, significa “o local real” onde algo acontece. É um termo frequentemente usado na filosofia enxuta, ou na questão da melhoria em geral, para transmitir a ideia de que, se você realmente deseja entender algo, vá onde isso realmente ocorre. Somente assim é que a verdadeira apreciação da realidade das oportunidades de melhoria pode ser obtida. Os defensores da melhoria enxuta costumam defender o uso da ideia de “ir ao local” para se conseguir maior visibilidade dos problemas. Assim, eles querem dizer que os gerentes devem regularmente visitar o local onde o trabalho é realizado para tentar encontrar desperdício. (Isso é semelhante à ideia ocidental da “gestão aproximando e conversando com as pessoas”.) O conceito de *gemba* também é usado no desenvolvimento de novo serviço ou produto em que os projetistas devem ir aonde o serviço acontece ou aonde o produto é usado para desenvolver suas ideias.



### Princípio de administração da produção

Não há substituto para ver o modo como os processos operam realmente na prática.

---

## Eliminação do desperdício mediante o fluxo regular

O fluxo regular de materiais, informações e pessoas na operação é uma ideia central da sincronização enxuta. Os longos roteiros de processo fornecem oportunidades para atraso e formação de estoque, não agregam valor e retardam o tempo de atravessamento. Portanto, a primeira contribuição que qualquer operação pode dar para agilizar o fluxo é reconsiderar o arranjo físico básico de seus processos. Primeiramente, reconfigurar o arranjo físico de um processo para ajudar a sincronização enxuta envolve deslocar para baixo a “diagonal natural” do projeto do processo que foi discutido no [Capítulo 6](#). Em termos gerais, isso significa passar de arranjos físicos funcionais para arranjos físicos baseados em células ou de arranjos físicos baseados em células para arranjos físicos em linha. De uma forma ou de outra, é necessário passar para um arranjo físico que proporcione maior sistematização e controle ao fluxo do processo. Em nível mais detalhado, as técnicas de arranjos físicos típicos incluem: colocar as estações de trabalho próximas, de modo que o estoque de materiais ou clientes não possa ser formado porque não há espaço para isso, e arranjar as estações de trabalho de tal forma que todos que contribuem para uma atividade comum possam ver uns aos outros e possam fornecer auxílio mútuo. Por exemplo, no Virginia Mason Medical Center, em Seattle, EUA, um importante proponente da sincronização enxuta nos serviços de saúde, muitas das salas de espera foram



bastante reduzidas em sua capacidade, ou totalmente removidas. Isso força um foco no fluxo do processo inteiro, pois os pacientes literalmente não possuem um local para serem “armazenados”.

### *Examinando todos os elementos do tempo de atravessamento*

Frequentemente, o tempo de atravessamento é assumido como uma medida substituta para o desperdício em um processo. Quanto mais tempo os itens processados forem mantidos em estoque, movimentados, conferidos ou sujeitos a algo mais que não agregue valor, mais tempo eles demoram para atravessar o processo. Assim, examinar o que ocorre aos itens em um processo é um excelente método de identificar fontes de desperdício.

O mapeamento do fluxo de valor (*value stream mapping*, também conhecido como mapeamento do sistema de “ponta a ponta”) é uma abordagem simples mas eficaz para entender o fluxo de materiais e informações à medida que um produto ou serviço é agregado enquanto passa por um processo, operação ou rede de suprimento. Ele mapeia visualmente a “produção” de um produto ou serviço do início ao fim. Ao fazer isso, registra não apenas as atividades diretas de criação de produtos e serviços, mas também os sistemas de informação “indiretos” que apoiam o processo direto. É denominado mapeamento de “fluxo de valor” porque foca as atividades de agregação de valor e distingue entre as atividades que agregam e não agregam valor. É semelhante ao mapeamento do processo (veja o [Capítulo 6](#)), mas diferente em quatro maneiras:

- Usa uma variedade mais ampla de informações do que a maioria dos mapas de processo.
- Geralmente, em um nível mais alto (de 5 a 10 atividades) do que a maioria dos mapas de processo.
- Frequentemente, tem um escopo mais amplo, envolvendo toda a cadeia de suprimento.
- Pode ser usado para identificar onde focar as futuras atividades de melhoria.

Uma perspectiva de fluxo de valor envolve trabalhar e melhorar o “panorama geral” em vez de apenas otimizar os processos individuais. O mapeamento do fluxo de valor é visto por muitos praticantes como o ponto de partida para ajudar a reconhecer o desperdício e identificar suas causas. É uma técnica de quatro etapas, que identifica o desperdício e sugere as formas pelas quais as atividades podem ser alinhadas. Primeiramente, envolve identificar o fluxo de valor (o processo, operação ou cadeia de suprimento) a mapear. Segundo, envolve mapear fisicamente um processo, depois, acima disso, mapear o fluxo de informações que possibilita a ocorrência do processo. Esse é o chamado mapa da “situação atual”. Terceiro, os problemas são diagnosticados, e mudanças, sugeridas, fazendo um mapa da “situação futura” que representa o processo, a operação ou a cadeia de suprimento melhorados. Finalmente, as mudanças são implementadas. A [Figura 15.6](#) mostra um mapa de fluxo de valor para um serviço de instalação

de ar-condicionado. O próprio processo do serviço é dividido em cinco estágios relativamente grandes, e vários itens de dados para cada estágio são marcados no gráfico. Aqui, os tipos de dados coletados variam, mas todos os tipos de mapas de fluxo de valor comparam o tempo total de atravessamento com a quantidade de tempo de valor agregado dentro do processo maior. Neste caso, apenas oito das 258 horas do processo estão agregando valor.

### *Adotando a gestão à vista*

A gestão à vista é uma das técnicas enxutas projetadas para tornar o estado atual e planejado da operação ou processo transparente a todos, de modo que as pessoas (que trabalham ou não no processo) possam perceber muito rapidamente o que está ocorrendo. Geralmente emprega algum tipo de sinal visual, como um quadro de aviso, uma tela de computador ou simplesmente luzes ou outros sinais que mostram o que está ocorrendo. Embora um dispositivo aparentemente trivial e simples, a gestão visual oferece vários benefícios, como ajudar a:

- atuar como um foco comum para reuniões de equipe;
- demonstrar métodos para práticas de trabalho seguras e efetivas;
- comunicar a todos como o desempenho está sendo avaliado;
- avaliar em relance o *status* atual da operação;
- entender as prioridades das tarefas e do trabalho;
- avaliar seu desempenho e de outras pessoas;
- identificar o fluxo do trabalho – o que ocorreu e o que está sendo feito;
- identificar quando algo não ocorre conforme planejado;
- mostrar quais devem ser os padrões combinados;
- fornecer *feedback* em tempo real sobre o desempenho a todos os envolvidos;
- reduzir a dependência das reuniões formais.

**Figura 15.6** Mapa de fluxo de valor para um serviço de instalação de ar-condicionado industrial.

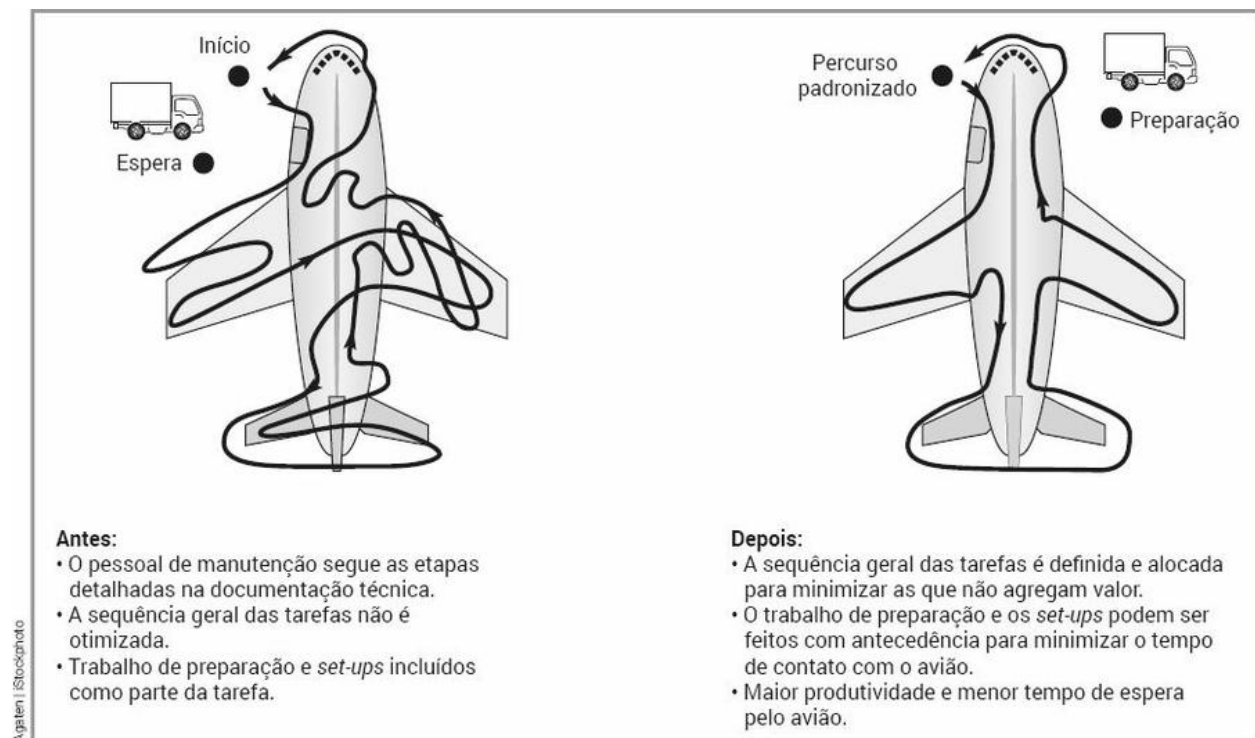
## OPERAÇÕES NA PRÁTICA

### Redução do desperdício na manutenção de aeronaves<sup>2</sup>

A manutenção de aviões é algo muito importante. As aeronaves têm uma tendência angustiante a cair, a menos que sejam verificadas, reparadas e, geralmente, submetidas à manutenção regular. Assim, o objetivo dominante das operações que mantêm aviões deve ser a qualidade das atividades de manutenção. Porém, esse não é o único objetivo. Melhorar o prazo de conclusão da manutenção pode reduzir o número de aeronaves que uma companhia aérea precisa possuir, porque elas não ficam fora de ação durante muito tempo. Além disso, quanto mais eficiente o processo de manutenção, mais rentável será a atividade e mais provável que uma grande companhia aérea, com operações de manutenção bem definidas, possa criar faturamento adicional fazendo manutenção para outras linhas aéreas. A [Figura 15.7](#) mostra como uma operação de manutenção de uma companhia aérea aplicou princípios enxutos para atingir todos esses objetivos. Os objetivos da análise enxuta eram preservar ou mesmo melhorar os níveis de qualidade, enquanto ao mesmo tempo melhorar o custo de manutenção dos aviões e aumentar a disponibilidade das aeronaves ao reduzir o tempo de parada.

A análise enxuta focava a identificação do desperdício no processo de manutenção. Duas constatações surgiram dessa prática. Primeiro, a sequência das atividades na própria aeronave estava sendo estabelecida pelas tarefas identificadas nos manuais técnicos fornecidos pelos fabricantes das turbinas, fuselagem e sistemas de controle e outros fornecedores. Ninguém

havia considerado todas as atividades individuais juntas para definir uma sequência que economizaria tempo e esforço do pessoal de manutenção. A sequência geral das atividades foi definida e alocada com a preparação do trabalho estruturado das ferramentas, materiais e equipamentos. A [Figura 15.7](#) mostra o roteiro seguido pelos funcionários de manutenção antes e após a análise enxuta. Segundo, os funcionários de manutenção estariam frequentemente esperando até a fuselagem ficar disponível. Todavia, alguns *set-ups* e trabalho preparatório não precisavam ser feitos enquanto a aeronave estivesse presente. Assim, por que não direcionar os funcionários a fazer essas tarefas no momento em que estariam esperando a aeronave estar disponível? O resultado dessas mudanças foi uma substancial melhoria em custo e disponibilidade. Ademais, a preparação do trabalho foi conduzida de maneira rotineira e mais rigorosa, e os funcionários estavam mais motivados porque muitas pequenas frustrações e barreiras ao trabalho eficiente foram removidas.



**Figura 15.7** Procedimentos de manutenção de avião sujeitos à análise de redução do desperdício.

Como um exemplo de como funciona a gestão à vista, o departamento financeiro de uma empresa varejista baseada na *web* estava tendo problemas. Os níveis de serviço estavam baixos, e as reclamações, altas, enquanto o departamento tentava lidar com os pagamentos de clientes, as faturas de fornecedores e os pedidos de informação de seu centro de distribuição; tudo isso enquanto a demanda estava aumentando. Concordeu-se que os processos do departamento eram caóticos e mal gerenciados, com pouco entendimento das prioridades ou como cada funcionário estava contribuindo. Para remediar essa situação, primeiro o gerente responsável pelo departamento tentou esclarecer o processo, definiu os papéis dos indivíduos e das equipes e passou a praticar a gestão à vista. Coletivamente, os funcionários mapearam os processos e estabeleceram objetivos de desempenho. Esses objetivos foram mostrados em um grande quadro que todos os funcionários do escritório podiam ver. No final de cada dia, os supervisores dos processos atualizavam o quadro com o desempenho diário de cada processo. Também indicadas no quadro estavam as representações visuais de vários projetos de melhoria sendo executados pelas equipes. Todas as

manhãs, os funcionários encontravam-se no que era chamada “reunião da manhã” para discutir o desempenho do dia anterior, identificar como podia ser melhorado, revisar o progresso dos projetos de melhoria em andamento e planejar o trabalho do dia seguinte.

---

O exemplo anterior ilustra as três principais funções da gestão visual:

- Atuar como um mecanismo de comunicação.
- Encorajar o comprometimento às metas combinadas.
- Facilitar a cooperação entre os membros da equipe.

Uma técnica importante que é usada para assegurar a visibilidade do fluxo é o uso de sinais simples, mas altamente visuais, indicando que um problema ocorreu, acompanhado de uma autoridade operacional para interromper o processo. Por exemplo, em uma linha de montagem, se um funcionário detecta algum tipo de problema de qualidade, ele pode ativar um sinal que acende uma luz (chamada Andon) em um ponto elevado de uma estação de trabalho e interrompe a linha de montagem. Embora isso possa parecer redução da eficiência da linha, a ideia é que essa perda de eficiência a curto prazo seja menor do que as perdas acumuladas de permitir que os defeitos continuem no processo. A menos que os problemas sejam resolvidos imediatamente, podem nunca ser corrigidos.

## OPERAÇÕES NA PRÁTICA

### Andon na Amazon<sup>3</sup>

Os princípios de uma corda ou luz Andon foram aplicados na Amazon. Todos os dias, os agentes de serviços da Amazon recebem chamadas de clientes que não estão satisfeitos com algum aspecto do produto que lhes foi entregue. Os agentes do cliente que lidam com essas queixas agora estão autorizados a avaliar a medida em que tais queixas podem ser sistêmicas. Nos casos em que suspeitam que seja um defeito repetitivo, os agentes de serviço podem “parar a linha de produção” para um produto específico. Isso envolve tirar o produto do *site* enquanto o problema está sendo totalmente investigado. De acordo com a Amazon, a visibilidade aprimorada do sistema eliminou dezenas de milhares de defeitos por ano e também deu aos agentes de serviços uma forte sensação de poder lidar de modo eficaz com as reclamações dos clientes. Agora, um agente pode não apenas reembolsar o cliente individual, mas também informar ao cliente que outros não receberão produtos até que o problema tenha sido devidamente investigado. A empresa também afirma que cerca de 98 por cento das vezes em que a corda Andon é puxada, realmente há um problema sistêmico, destacando o valor de confiar nos agentes de serviços para tomar decisões sensatas sobre quando parar e quando não parar a linha de produção.

---

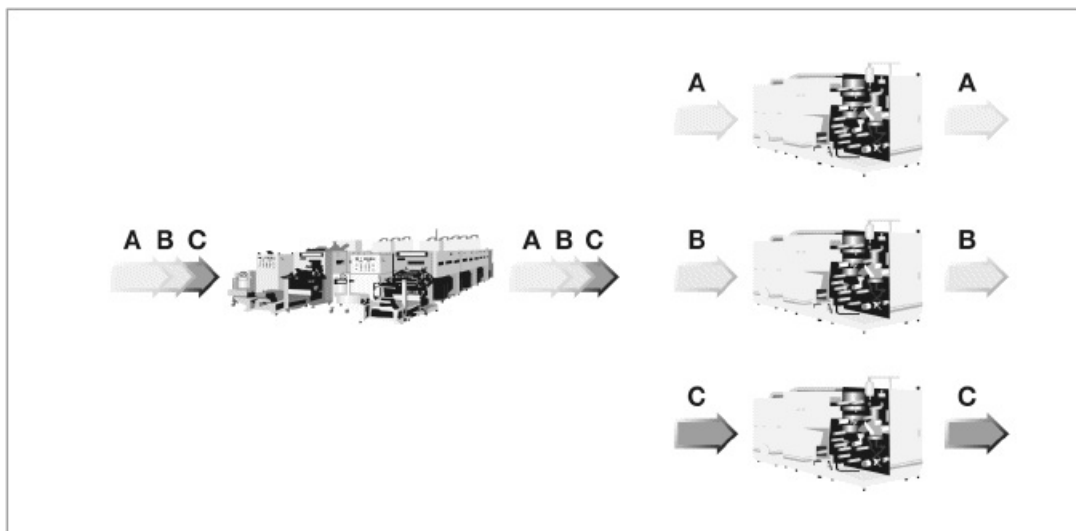
## *Usando tecnologia de processo simples e de pequena escala*

Pode também haver possibilidades de encorajar o fluxo otimizado e regular mediante o uso

de tecnologias de pequena escala. Isto é, usar várias pequenas unidades de tecnologia de processo (por exemplo, máquinas), em vez de uma grande unidade. Máquinas pequenas têm várias vantagens sobre as grandes. Primeiro, podem processar simultaneamente produtos e serviços diferentes. Por exemplo, na [Figura 15.8](#), uma grande máquina produz o lote A, seguido pelo lote B e, depois, o lote C. Entretanto, se três máquinas menores forem usadas, cada uma pode produzir simultaneamente os lotes A, B ou C. O sistema também é mais robusto. Se uma grande máquina quebrar, todo o sistema para de operar. Se uma das três pequenas máquinas quebrar, o sistema ainda pode continuar operando com dois terços de eficácia. Máquinas pequenas são também facilmente movidas, de modo que a flexibilidade do arranjo físico é aumentada, e os riscos de cometer erros nas decisões de investimento são reduzidos. Entretanto, o investimento na capacidade física pode aumentar no total porque instalações paralelas são necessárias, de modo que a utilização pode ser menor (veja os argumentos anteriores).

### **Eliminação do desperdício tornando o suprimento e a demanda exatamente iguais**

O valor do suprimento de serviços ou produtos é sempre dependente do tempo. Algo que é entregue antes ou depois frequentemente tem menos valor do que algo entregue exatamente quando é necessário. Por exemplo, as empresas de entrega de encomendas cobram mais pela garantia de entrega mais rápida. Isso ocorre porque nossa necessidade real de recebimento da encomenda é, frequentemente, a mais rápida possível. Quanto mais próxima da entrega instantânea, maior valor ela representa para nós e mais estaremos dispostos a pagar por ela. De fato, a entrega da informação mais cedo do que o necessário pode ser ainda mais prejudicial do que sua entrega atrasada, porque resulta em estoques de informação que servem para confundir o fluxo ao longo do processo.



**Figura 15.8** O uso de várias máquinas pequenas em vez de uma máquina grande permite processamento simultâneo, além de ser mais robusto e mais flexível.



### Princípio de administração da produção

Entregar apenas e exatamente o que é necessário e quando necessário regulariza o fluxo e expõe o desperdício.

---

## Controle puxado

Frequentemente, a ação de igualar o suprimento e a demanda é mais bem obtida, sempre que possível, usando-se o “controle puxado” (veja no [Capítulo 10](#)). Em um exemplo simples, consideremos como alguns restaurantes *fast-food* cozinham e montam o alimento e o colocam na área de aquecimento apenas após o cliente ter feito seu pedido a um funcionário. A produção é acionada apenas pela demanda real do cliente. Em outro exemplo, o departamento de imposto de renda da Austrália estava habituado a receber formulários pelo correio, os encaminhar ao departamento relevante e, após processá-los, os enviar ao departamento seguinte. Isso ocasionava pilhas de formulários não processados obstruindo os processos, causando problemas de rastreamento e extravio, de seleção e priorização e, o pior de tudo, longo tempo de atravessamento. Agora, os funcionários só abrem os envelopes com os formulários quando as etapas iniciais puderem processá-los. Cada seção solicita mais trabalho apenas após ter processado o trabalho anterior. A essência do controle puxado é deixar o estágio a jusante (*downstream*) em um processo, operação ou rede de suprimento puxar os itens (produtos, clientes ou informação) ao longo do sistema em vez de serem “empurrados” pelo estágio de suprimento.

## Kanbans

O uso de *kanbans* é um método de operacionalizar o controle puxado. *Kanban* é a palavra japonesa para cartão ou sinal. Às vezes, é o denominado “transportador invisível” que controla a transferência de itens entre os estágios de uma operação. Em sua forma mais simples, é um cartão usado no estágio cliente para instruir seu estágio fornecedor a enviar mais itens. Os *kanbans* podem também assumir outras formas. Em algumas empresas japonesas, são marcadores de plástico sólido ou até bolas de pingue-pongue coloridas. Qualquer que for o tipo de *kanban* usado, o princípio é sempre o mesmo: o recebimento de um *kanban* aciona o movimento, a produção ou o suprimento de uma unidade ou um recipiente padrão de unidades. Se dois *kanbans* forem recebidos, esses acionam a movimentação, a produção ou o suprimento de duas unidades ou recipientes padrão de unidades, e assim por diante. Os *kanbans* são o único meio pelo qual a movimentação, a produção ou o suprimento pode ser autorizado. Algumas

empresas usam “quadrados *kanban*”. Estes são espaços marcados no chão da fábrica ou na bancada, desenhados para ajustar uma ou mais peças ou recipientes com trabalho. Apenas a existência de um quadrado vazio aciona a produção no estágio que preencherá um quadrado. De modo semelhante, os quadros brancos de *kanban* são cada vez mais usados para “puxar” a atividade através do processo de serviço (veja a seção “Operações na Prática”). O *kanban* é visto como servindo a três propósitos:

- É uma instrução para o processo precedente enviar mais trabalho.
- É uma ferramenta de controle visual para mostrar áreas de excesso de produção e falta de sincronização.
- É uma ferramenta para *kaizen* (melhoria contínua). As normas da Toyota declaram que “o número de *kanbans* deve ser reduzido ao longo do tempo”.

## OPERAÇÕES NA PRÁTICA

### Controle *kanban* para os *web designers* da Torchbox<sup>4</sup>

A Torchbox é uma empresa independente de *web design* e desenvolvimento, situada em Oxfordshire (para maiores detalhes, veja o caso do [Capítulo 1](#): “Torchbox: *web designers* ganhadores de prêmio”). A Torchbox fornece soluções de alta qualidade, de melhor custo-benefício e éticas para seus clientes. Embora seja uma empresa que confia na capacidade criativa de seus funcionários, isso não significa que as técnicas enxutas não têm lugar em suas operações. Ao contrário, ela faz uso pleno da abordagem *kanban* para controlar seu trabalho à medida que progride ao longo do processo de *web design*. “Sabemos que o controle *kanban* originou-se nas fábricas de automóveis, como a Toyota, mas nossas equipes de desenvolvimento podem também se beneficiar de seus princípios básicos”, diz Edward Kay, chefe de produção da Torchbox. “É um modo de programar o trabalho baseado no que precisa ser produzido e quais recursos estão disponíveis para isso. Na Torchbox, usamos um quadro magnético branco e grande (chamado ‘quadro *kanban*’) para acompanhar os elementos concluídos em cada estágio do processo de design; desde a descoberta, passando por desenvolvimento, demonstração, desdobramento e finalização do projeto (chamado estágio ‘concluído’). Cada elemento possui seu próprio cartão que se move fisicamente no quadro, preso no local com um ímã. Você não pode ter mais elementos em andamento do que o número de ímãs que possui para prendê-los no quadro, de modo que o princípio é imposto como uma limitação física. Quando um elemento entra na coluna ‘concluído’, outro pode ser puxado para a coluna ‘descoberta’. Há um processo de puxar, no qual o trabalho concluído permite ao projetista iniciar algo novo.”

No início de cada dia, a equipe faz uma reunião de pé na frente do quadro *kanban*, na qual cada membro discorre rapidamente sobre o que fez no dia anterior e o que fará no dia atual. Cada desenvolvedor possui alguns cartões que serão posicionados nos elementos em que estão trabalhando. Isso ajuda a vincular a “imagem global”, que mostra como os elementos de um projeto estão se desenvolvendo, a uma “imagem pequena”, sobre em que cada desenvolvedor está trabalhando a cada dia e ajuda as equipes a assegurar que todos os trabalhos feitos estão sendo acompanhados ao longo do quadro *kanban*.

“Um dos maiores benefícios de usar *kanban*”, afirma Edward Kay, “é que, em razão de estarmos visualizando as etapas que um elemento percorre para ser concluído, estamos em condições de verificar onde estão os gargalos que causam morosidade no



*trabalho. Em razão de podermos perceber onde o processo está sendo retardado, temos condições de melhorar continuamente nossos processos para assegurar que o trabalho esteja sendo feito do modo mais eficiente possível. Se constatarmos que as características de um projeto estão atrasadas no estágio de projeto, podemos trazer mais desenvolvedores para resolver o gargalo. O uso de kanban com o desenvolvimento orientado para a característica ajuda-nos constantemente a oferecer valor a nossos clientes. Essa abordagem de maior mensuração e controle para lidar e controlar o trabalho novo ajuda a assegurar que, a cada hora, nosso trabalho produz o que realmente vale uma hora. Finalmente, isso tem tudo a ver com a entrega pontual de grandes produtos, conforme o orçamento, e kanban é uma ótima ferramenta para ajudar a atingir isso."*

---

## **Eliminação de desperdício mediante processos flexíveis**

Responder de modo exato e instantâneo à demanda dos clientes implica que os recursos das operações precisam ser suficientemente flexíveis para mudar o que fazem e quanto fazem sem incorrer em custos elevados ou longos atrasos. De fato, os processos flexíveis (frequentemente com tecnologias flexíveis) podem intensificar significativamente o fluxo regular e sincronizado. Por exemplo, um escritório de advocacia costumava demorar dez dias para emitir as faturas a seus clientes. Isso significava que os clientes não eram cobrados até dez dias após o serviço ter sido prestado. Agora usam um sistema que, diariamente, atualiza a conta de cada cliente. Assim, quando uma fatura é enviada, inclui todos os trabalhos prestados até o dia anterior de sua emissão. O princípio aqui é que a inflexibilidade do processo também atrasa o fluxo de caixa.

### *Redução dos tempos de troca*<sup>5</sup>

Para muitos processos, aumentar a flexibilidade significa reduzir o tempo de troca de uma atividade para a seguinte. Geralmente, esse tempo pode ser reduzido, às vezes, radicalmente. Por exemplo, comparemos o tempo decorrido em uma troca de pneu de seu carro com o tempo de uma equipe de Fórmula 1. A redução do tempo de troca pode ser atingida por diversos métodos, como os seguintes:

- *Meça e analise as atividades de troca.* Às vezes, medir simplesmente os tempos de troca atuais, gravando-os e analisando exatamente quais atividades desempenhadas podem ajudar a melhorar esses tempos.
- *Separe as atividades internas das externas.* Atividades “externas” são simplesmente as atividades que podem ser conduzidas enquanto o processo está em andamento. Por exemplo, os processos podem ficar prontos para o próximo cliente ou tarefa enquanto estiver esperando pela tarefa seguinte (veja o exemplo, descrito anteriormente, da manutenção de aeronave). Atividades “internas” são aquelas que não podem ser realizadas enquanto o processo estiver em andamento (como entrevistar o cliente enquanto estiver concluindo um serviço solicitado pelo cliente anterior). Ao identificar e separar as

atividades internas e externas, a intenção é fazer o máximo possível enquanto o processo/etapa estiver em andamento.

- *Converta atividades internas em externas.* A outra abordagem comum para a redução do tempo de troca é converter o trabalho que era previamente executado durante a troca pelo trabalho que é desempenhado externamente durante o mesmo período. Há três métodos importantes para se obter a transferência do trabalho de troca interno para o trabalho externo:
- Prepare antecipadamente as atividades ou o equipamento em vez de ter que fazê-lo durante os períodos de troca.
- Torne o processo de troca intrinsecamente flexível e capaz de desempenhar todas as atividades necessárias sem qualquer atraso.
- Agilize quaisquer mudanças necessárias de equipamento, informação ou funcionário, por exemplo, usando dispositivos simples.
- *Pratique rotinas de troca.* Não é de se surpreender que a prática constante de rotinas de troca e a curva de aprendizagem associada afetem as tendências para reduzir os tempos de troca.



### Princípio de administração da produção

A flexibilidade de troca (*set-up*) reduz o desperdício e suaviza o fluxo.

---

As trocas rápidas de equipamentos são particularmente importantes para as companhias aéreas porque elas não podem ganhar dinheiro com avião ocioso no solo. No setor aéreo, isso é denominado “alta utilização do avião”. Para muitas companhias pequenas, a maior barreira à alta utilização é que seus mercados não são grandes o suficiente para justificar voos de passageiros durante o dia e a noite. Assim, para evitar que os aviões fiquem ociosos à noite, eles precisam ser utilizados de outra forma. Essa foi a razão por trás da ideia do avião Boeing 737 QC (*Quick Change* – mudança rápida). Com ele, as linhas aéreas têm a flexibilidade de usá-lo de dia para voos de passageiros e, com o tempo de troca (*set-up*) menor do que uma hora, podem usá-lo como avião de carga durante a noite. Os engenheiros da Boeing projetaram suportes que fixam fileiras inteiras de assentos que podem deslizar-se suavemente para dentro e para fora do avião, permitindo que 12 assentos sejam deslocados de uma só vez. Quando usado como avião de carga, os assentos são simplesmente deslizados para fora e substituídos por contêineres de carga projetados para se encaixarem na curvatura da fuselagem e evitar danos ao interior. Antes de reinstalar os assentos, as paredes laterais da aeronave são completamente lavadas, de modo que, uma vez recolocados os assentos, os passageiros não conseguem notar a diferença entre um avião QC e um 737 normal. Algumas linhas aéreas valorizam particularmente a flexibilidade do avião. Isso permite a elas prestar serviços frequentes confiáveis aos mercados de passageiros e de carga.

Assim, a aeronave que pode transportar passageiros durante o dia pode ser usada para o embarque de carga durante a noite.

## Eliminando o desperdício mediante minimização de variabilidade

Uma das maiores causas da variabilidade que interromperá o fluxo e evitará a sincronização enxuta é a variação da qualidade dos itens. É por isso que uma discussão de sincronização enxuta deve sempre incluir uma avaliação de como a qualidade de conformidade é assegurada dentro dos processos, algo que foi referido anteriormente como *mura*. Em particular, os princípios de controle estatístico do processo (CEP) podem ser usados para se entender a variabilidade da qualidade. O [Capítulo 17](#) e seu suplemento sobre CEP examinam esse assunto, de modo que nesta seção focaremos outras causas da variabilidade. A primeira dessas causas é a variabilidade do *mix* de produtos e serviços que se move ao longo dos processos, operações ou redes de suprimento.

### Nivelando as programações de produto ou serviço

A programação nivelada (ou *heijunka*) significa manter o *mix* e o volume do fluxo entre os estágios, ainda que por mais tempo. Por exemplo, em vez de produzir 500 peças em um lote, que atenderia às necessidades para os próximos três meses, a programação nivelada exigiria que o processo fabricasse regularmente itens que satisfizessem aquela demanda. Assim, o princípio da programação nivelada é muito direto. Entretanto, as exigências para colocá-la em prática são muito rigorosas, embora os benefícios que resultam disso possam ser substanciais. O movimento da programação convencional para a programação nivelada é ilustrado na [Figura 15.9](#). Convencionalmente, se um *mix* de produtos for requerido em determinado período de tempo (geralmente, um mês), o tamanho do lote seria calculado por produto, e os lotes produzidos, em alguma sequência. A [Figura 15.9\(a\)](#) mostra três produtos que são fabricados no período de 20 dias em uma operação.

Quantidade requerida do item A = 3.000

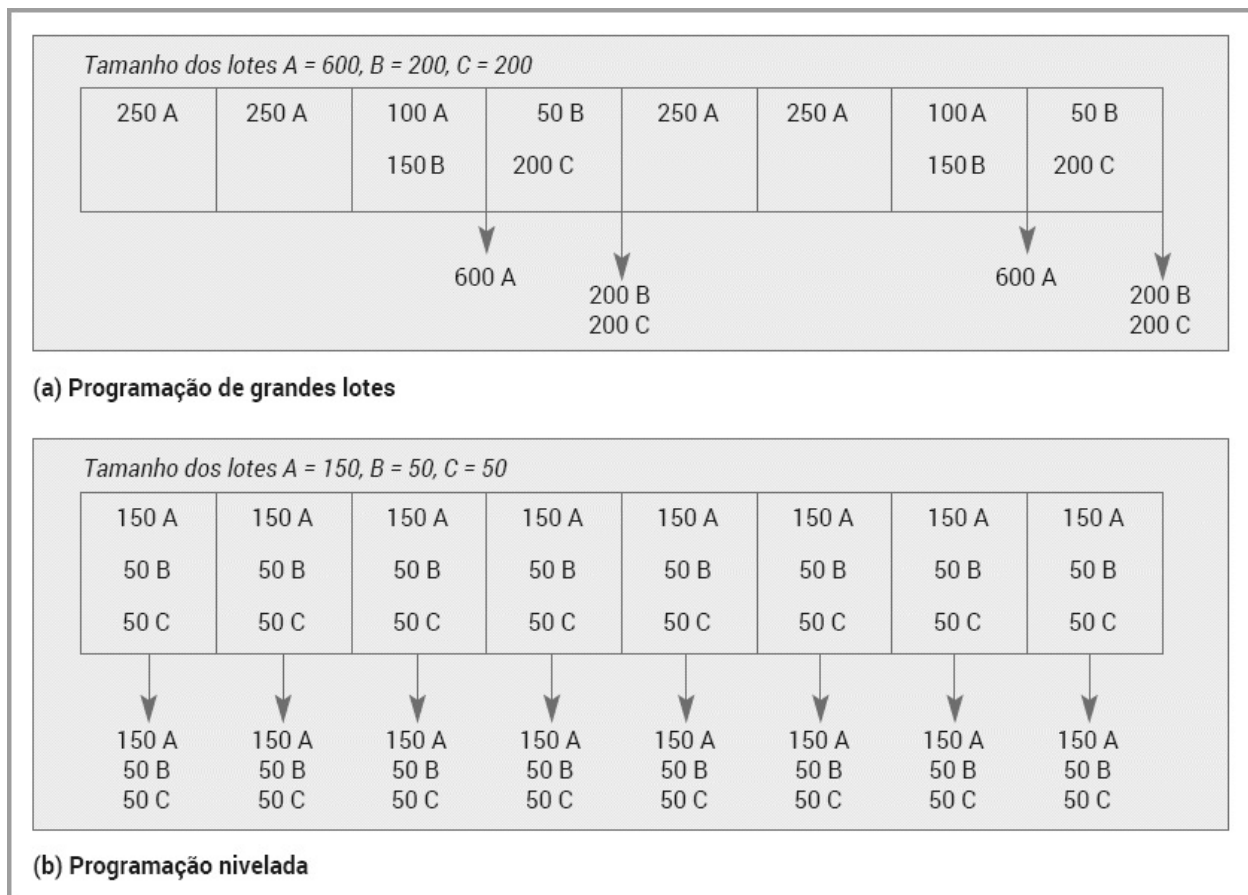
Quantidade requerida do item B = 1.000

Quantidade requerida do item C = 1.000

Tamanho do lote do item A = 600

Tamanho do lote do item B = 200

Tamanho do lote do item C = 200



**Figura 15.9** A programação nivelada equaliza o *mix* de produtos fabricados a cada dia.

A partir do 1º dia, a unidade começa a produzir o item A. Durante o 3º dia, o lote de 600 itens A é finalizado e despachado ao estágio seguinte. O lote do item B é iniciado, mas não finalizado até o 4º dia. O restante do 4º dia é gasto fabricando o lote do item C e ambos os lotes são despachados no final desse dia. Assim, o ciclo se repete. A consequência de usar lotes maiores é, primeiro, que volumes relativamente grandes de estoque são acumulados dentro e entre as unidades e, segundo, a maioria dos dias é diferente um do outro em termos do que é esperado produzir (em circunstâncias mais complexas, dois dias nunca seriam iguais).

Agora, suponhamos que a flexibilidade da unidade pudesse ser aumentada ao ponto em que os tamanhos dos lotes para os itens fossem reduzidos a um quarto de seus níveis anteriores, sem perda da capacidade (veja a [Figura 15.9\(b\)](#)):

Tamanho do lote do item A = 150

Tamanho do lote do item B = 50

Tamanho do lote do item C = 50

Um lote de cada item pode agora ser concluído em um único dia, no final do qual os três lotes são encaminhados ao estágio seguinte. Os lotes de estoque menores estão movendo-se entre

cada estágio, o que reduzirá o nível geral do trabalho em processo de operação. Entretanto, igualmente significativo é o efeito sobre a regularidade e o ritmo de produção na unidade. Agora, qualquer dia do mês é o mesmo em termos de que necessidades são processadas. Isso torna o planejamento e controle de cada estágio da operação muito mais fáceis. Por exemplo, se no 1º dia do mês o lote diário de A foi concluído às 11 h e todos os lotes forem concluídos com sucesso no mesmo dia, no dia seguinte a unidade saberá que, se concluir novamente todos os lotes de A às 11 h, cumprirá a programação. Quando todos os dias são diferentes, a simples pergunta: “Estamos dentro da programação para concluir nosso processamento hoje?” requer alguma investigação antes que possa ser respondida. Entretanto, quando todos os dias são iguais, todas as pessoas da unidade podem informar se a produção está conforme o previsto ao olhar para o relógio. O controle torna-se visível e transparente a todos, e as vantagens de programações regulares e diárias podem ser passadas aos fornecedores a montante (*upstream*).



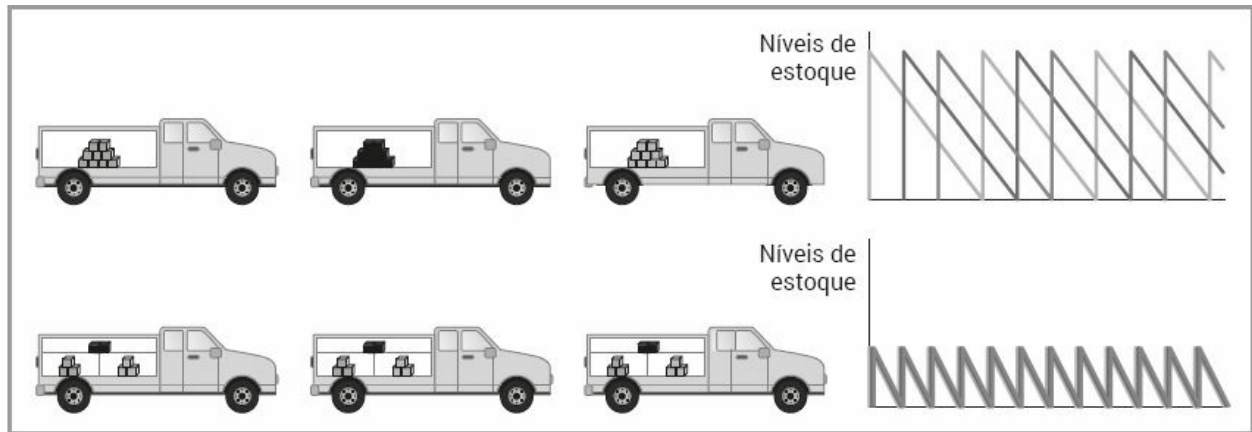
#### Princípio de administração da produção

A variabilidade na qualidade, na quantidade ou no *timing* do produto/serviço atua contra o fluxo regular e a eliminação do desperdício.

---

### *Nivelando as programações de entrega*

Um conceito semelhante à programação nivelada pode ser aplicado a muitos processos de transporte. Por exemplo, uma rede de lojas de conveniências pode precisar fazer entregas de todos os tipos diferentes de produtos que vende diariamente. Tradicionalmente, pode ter que despachar um caminhão carregado com um produto específico para suprir todas as lojas, de modo que cada uma delas receba o volume apropriado do produto necessário por uma semana. Isso é equivalente aos grandes lotes discutidos no exemplo anterior. Uma alternativa seria despachar quantidades menores de todos os produtos em um único caminhão com maior frequência. Assim, cada loja receberia entregas menores mais frequentemente, os níveis de estoque seriam menores e o sistema poderia responder às tendências da demanda mais prontamente porque mais entregas significa maior oportunidade de alterar as quantidades entregues em uma loja. Isso está ilustrado na [Figura 15.10](#).

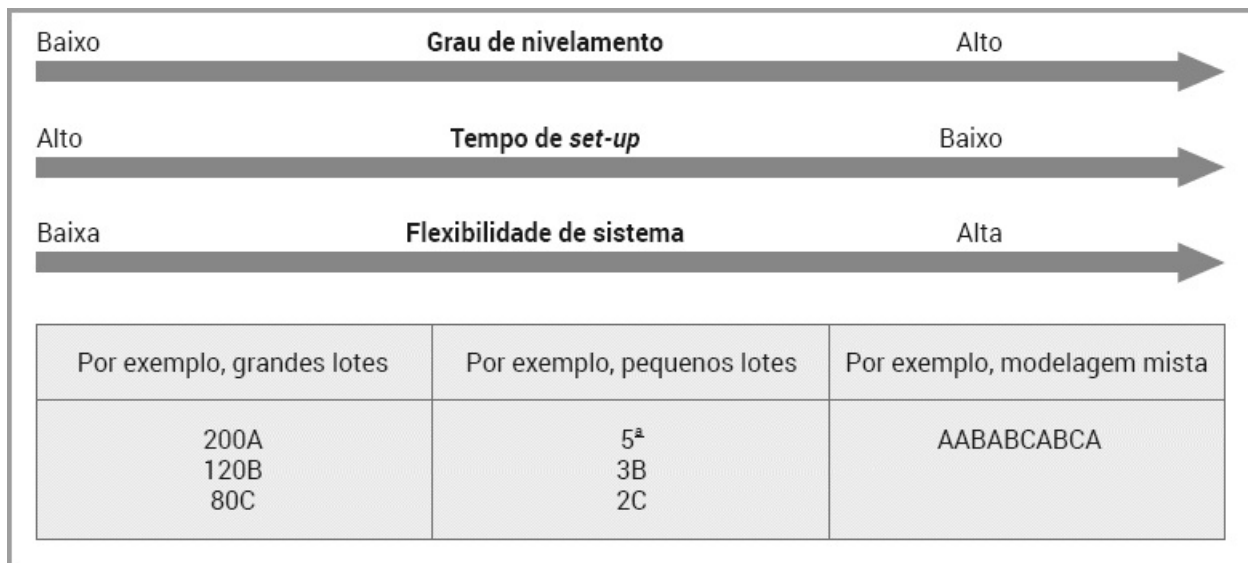


Agaten | iStockphoto

**Figura 15.10** Entregar quantidades menores mais frequentemente pode reduzir os níveis de estoque.

### *Adotando a mistura de modelos*

O princípio de programação nivelada pode ir mais longe para proporcionar um *mix* repetido de *outputs* – conhecido como mistura de modelos. Suponhamos que os processos podem se tornar tão flexíveis que atinjam o ideal de tamanho padrão de um “lote”. A sequência de itens individuais emergindo do processo pode ser progressivamente reduzida como ilustrado na [Figura 15.11](#). Isso produziria um fluxo regular de cada item fluindo continuamente a partir do processo. Entretanto, a sequência dos itens nem sempre ocorre tão conveniente como na [Figura 15.11](#). Os tempos de processamento unitário por item não são geralmente idênticos e as proporções dos volumes requeridos são menos convenientes. Por exemplo, se um método for exigido para processar os itens A, B e C na proporção 8:5:4, ele pode produzir 800 de A, 500 de B, 400 de C, ou 80A, 50B e 40C. Seria ideal sequenciar os itens o mais regularmente possível para produzir repetidamente na ordem... BACABACABACABACAB... repetindo... repetindo... etc. Fazendo isso, atinge-se fluxo relativamente regular (mas confia-se na flexibilidade significativa do processo).



**Figura 15.11** Programação nivelada e modelagem mista: a modelagem mista torna-se possível à medida que o tamanho do lote aproxima-se de 1.

### *Simplificando as coisas – o 5S*

A terminologia 5S<sup>6</sup> veio originalmente do Japão, e embora a tradução seja aproximada, ela é, geralmente, adotada para representar o seguinte:

1. **Seppure** (*seiri*) – Elimine o que não é necessário e mantenha o necessário.
2. **Organize** (*seiton*) – Posicione as coisas de tal forma que sejam facilmente alcançadas sempre que necessário.
3. **Limpe** (*seiso*) – Mantenha tudo limpo e arrumado; nenhum lixo ou sujeira na área de trabalho.
4. **Standardize** (*seiketsu*) – Mantenha sempre a ordem e a limpeza – arrumação permanente.
5. **Sustente** (*shitsuke*) – Desenvolva comprometimento e orgulho em manter padrões.

Os 5S podem ser pensados como um simples método de arrumação da casa para organizar áreas de trabalho que enfatizem ordem visual, organização, limpeza e padronização. Isso ajuda a eliminar todos os tipos de desperdício relacionados à incerteza, espera, busca de informações relevantes, criação de variação etc. Ao eliminar o que for desnecessário e ao deixar tudo limpo e previsível, a desordem é reduzida, os itens necessários estão sempre nos mesmos lugares e o trabalho fica mais fácil e mais rápido.

### *Adotando a manutenção produtiva total (MPT ou em sigla inglesa, TPM)*

A manutenção produtiva total visa eliminar a variabilidade dos processos de produção



causada pelo efeito das quebras. A MPT será tratada posteriormente no capítulo sobre a gestão de risco e recuperação ([Capítulo 18](#)).

## PRODUÇÃO ENXUTA APLICADA AO LONGO DA REDE DE SUPRIMENTO

Embora a maioria dos conceitos e técnicas discutidos neste capítulo seja dedicada a gerenciar os estágios *dentro* dos processos e os processos *dentro* de uma operação, os mesmos princípios podem aplicar-se em toda a cadeia de suprimento. Nesse contexto, os estágios em um processo estão em todas as empresas, operações ou processos entre os quais os produtos fluem. Quando qualquer empresa passa a utilizar a abordagem enxuta, eventualmente irá contra as restrições impostas pela falta de sincronização das outras operações de sua cadeia de suprimento. Assim, a obtenção de ganhos maiores deve envolver a tentativa de propagar a prática enxuta entre seus parceiros na cadeia. Garantir que as redes de suprimento inteiras adotem a filosofia enxuta certamente é uma tarefa muito mais complexa do que entre os estágios individuais dentro de um processo. Provavelmente um *mix* bem mais complexo de produtos e serviços está sendo fornecido e toda a rede está sujeita a um conjunto menos previsível de eventos potencialmente destrutivos. Tornar uma cadeia de suprimento enxuta significa mais do que tornar cada operação enxuta. Um conjunto de operações enxutas localizadas raramente leva a uma cadeia globalmente enxuta. Ao contrário, é necessário aplicar a filosofia de sincronização enxuta à cadeia de suprimento como um todo. Todavia, as vantagens de cadeias verdadeiramente enxutas podem ser significativas.

Essencialmente, os princípios da abordagem enxuta são os mesmos para uma cadeia de suprimento, como também são para um processo. O atravessamento rápido de uma rede de suprimento inteira é ainda mais valioso e reduzirá custos por toda a rede. Níveis menores de estoque tornarão mais fácil atingir a sincronização enxuta. O desperdício é mais evidente (e ainda maior) no nível da rede de suprimento, e reduzi-lo é ainda uma tarefa que vale a pena. O fluxo alinhado, a correspondência exata entre suprimento e demanda, o aumento da flexibilidade e a redução da variabilidade ainda são tarefas que beneficiarão a totalidade da rede. Os princípios do controle puxado podem funcionar entre operações completas, da mesma forma que podem funcionar entre os estágios de um único processo. De fato, os princípios e as técnicas da abordagem enxuta são basicamente os mesmos, não importa que nível de análise esteja sendo usado. E como a abordagem enxuta está sendo implementada em maior escala, os benefícios também serão proporcionalmente maiores.

Um dos pontos fracos da abordagem enxuta é a dificuldade de ser atingida quando as condições estão sujeitas a distúrbios inesperados. Isso é especialmente um problema com a aplicação dos princípios da sincronização enxuta no contexto de toda a rede de suprimento. Embora possa ocorrer flutuações inesperadas e distúrbios nas operações, a administração local tem um grau razoável de controle que pode exercer para reduzi-los. Fora da operação, dentro da



rede de suprimento, isso é bem mais difícil. Todavia, assume-se geralmente que, embora a tarefa seja mais difícil e mais demorada para ser alcançada, o propósito da abordagem enxuta é tão valioso para a rede de suprimento como um todo quanto o é para uma operação individual.



### Princípio de administração da produção

As vantagens da sincronização enxuta aplicam-se no nível de processo, operação e rede de suprimento.

## Cadeias de suprimento enxutas são como sistemas de controle de tráfego aéreo

O conceito de rede de suprimento enxuta vem sendo vinculado a um sistema de tráfego aéreo, em que se tenta dar “visibilidade e controle em tempo real” contínuos para todos os elementos da rede. Esse é o segredo de como os aeroportos mais movimentados do mundo lidam diariamente com milhares de partidas e chegadas. Todos os aviões recebem um número de identificação que aparece em um mapa de radar. Os aviões que se aproximam de um aeroporto são detectados pelo radar e contatados por rádio. A torre de controle posiciona precisamente uma aeronave em um padrão de aproximação que ela coordena. O radar detecta quaisquer pequenos ajustes necessários, que são comunicados à aeronave. Essa visibilidade e controle em tempo real podem otimizar o atravessamento do aeroporto, mantendo segurança e confiabilidade extremamente altas.

Compare isso a como a maioria das redes de suprimento é coordenada. As informações são capturadas apenas periodicamente, provavelmente uma vez por dia, e quaisquer ajustes na logística dos níveis de produção das várias operações da cadeia de suprimento são efetuados, e os planos, rearranjados. Imaginemos o que aconteceria se o aeroporto fosse operado desse modo, com apenas uma “varrida do radar” uma vez por dia. Coordenar um avião com tolerância suficiente para organizar decolagens e aterrissagem a cada dois minutos estaria fora de questão. O avião estaria em risco ou, alternativamente, se o espaço entre os voos fosse ampliado para manter a segurança, o tempo produtivo seria drasticamente reduzido. Todavia, é assim que a maioria das redes de suprimento opera tradicionalmente. Elas fazem uma “varrida” diária em seus sistemas ERP (veja no [Capítulo 14](#)). Essa visibilidade limitada significa que as operações devem espaçar seus trabalhos para evitar “colisões” (isto é, pedidos perdidos de clientes), reduzindo, assim, a produção, ou devem “fazer voo cego”, colocando a confiabilidade em perigo.



### Comentário crítico

Lembre-se da seção anterior sobre a vulnerabilidade das cadeias de suprimento (veja o [Capítulo 13](#)), quando se

argumentou que os princípios enxutos podem ser levados ao extremo. Quando as ideias *just-in-time* passaram a ter impacto sobre as práticas de produção ocidentais, algumas autoridades defenderam a redução a zero dos estoques entre os processos. Embora, a longo prazo, isso dê um ultimato na motivação para os gerentes de produção assegurarem a eficiência e a confiabilidade de cada estágio do processo, não admite a possibilidade de alguns processos serem intrinsecamente menos do que totalmente confiáveis. Uma visão alternativa é permitir estoques (embora pequenos) em torno dos estágios do processo com incerteza acima da média. Isso, pelo menos, permite alguma proteção para o restante do sistema. As mesmas ideias aplicam-se à entrega *just-in-time* entre as fábricas. A brusca interrupção das redes de suprimento, como pelos efeitos do *tsunami* no Japão, levou muitas fábricas a interromper a produção durante algum tempo em razão da escassez dos componentes mais importantes.

## A PRODUÇÃO ENXUTA COMPARADA COM OUTRAS ABORDAGENS

Tanto quanto uma filosofia ampla ou um método prático do planejamento e controle da produção, a abordagem enxuta não é a única utilizada na prática. Há outras abordagens que podem ser usadas para sustentar a melhoria e o planejamento e controle da produção. No [Capítulo 16](#), descreveremos posteriormente como as operações enxutas se comparam a outras abordagens de melhoria. Aqui, examinamos brevemente duas alternativas à sincronização enxuta como um método de planejamento e controle: a teoria das restrições (TOC) e o *material requirements planning* (MRP) (que examinamos no suplemento ao [Capítulo 14](#)).

### Abordagem enxuta e teoria das restrições

Uma ideia central da abordagem enxuta é o fluxo regular de itens ao longo de processos, operações e redes de suprimento. Qualquer gargalo interromperá esse progresso regular. Assim, é importante reconhecer a significância das restrições de capacidade ao processo de planejamento e controle. Essa é a ideia por trás da teoria das restrições (TOC – *Theory Of Constraints*) que tem sido desenvolvida para focar a atenção nas restrições de capacidade ou partes da operação que formam gargalos. Identificando a localização das restrições, trabalhando para removê-las e depois procurando a nova restrição, uma operação está sempre focando a parte que determina criticamente o andamento da produção. A abordagem que usa essa ideia é denominada tecnologia de produção otimizada (OPT – *Optimized Production Technology*). Seu desenvolvimento e o marketing como produto de um *software* proprietário foram originados por Eliyahu Goldratt.<sup>7</sup> OPT é uma técnica e ferramenta baseada em computador que ajuda a programar os sistemas de produção no andamento ditado pelo recurso mais fortemente carregado – isto é, os gargalos. Se a taxa de atividade de qualquer parte do sistema exceder o gargalo, itens serão produzidos e não poderão ser usados. Se a taxa de trabalho situar-se abaixo do ritmo do gargalo, todo o sistema estará subutilizado. Há princípios básicos do OPT que demonstram esse foco nos gargalos:

1. Fluxo balanceado, e não capacidade física. É mais importante reduzir o tempo de atravessamento em vez de atingir um balanceamento da capacidade entre estágios e processos.
2. O nível de utilização de um não gargalo é determinado por alguma outra restrição no sistema, não por sua própria capacidade. Isso se aplica aos estágios de um processo, aos processos de uma operação e às operações de uma rede de suprimento.
3. A utilização e a ativação de um recurso não são a mesma coisa. Conforme a TOC, um recurso está sendo utilizado apenas se contribui para o processo inteiro ou para a operação que cria mais *output*. Um processo ou estágio pode ser ativado no sentido que está trabalhando, mas pode apenas estar criando estoque ou desempenhando outra atividade que não agrega valor.
4. Uma hora perdida (ou não usada) em um gargalo é uma hora perdida para sempre em todo o sistema. O gargalo limita o *output* de todo o processo ou operação, de modo que a subutilização de um gargalo afeta todo o processo ou operação.
5. Uma hora economizada em um não gargalo é uma miragem. Os não gargalos têm capacidade disponível. Por que se preocupar tornando-os ainda menos utilizados?
6. Os gargalos governam o atravessamento e o estoque no sistema. Se os gargalos governam o fluxo, governam também o tempo de atravessamento, que, por sua vez, governa o estoque.
7. Você não precisa transferir lotes das mesmas quantidades à medida que os produz. Provavelmente, o fluxo será melhorado ao dividir grandes lotes de produção em lotes menores para se moverem ao longo do processo.
8. O tamanho do lote do processo deve ser variável, não fixo. Novamente, a partir do modelo do lote econômico de compra (LEC), as circunstâncias que controlam o tamanho do lote podem variar entre diferentes produtos.
9. As flutuações nos processos sequenciais conectados e dependentes acrescentam umas às outras em vez de nivelar-se. Assim, se dois processos ou estágios paralelos são capazes de uma taxa média de *output*, em sequência, nunca serão capazes de atingir a mesma média de *output*.
10. As programações devem ser estabelecidas ao examinar, simultaneamente, todas as restrições. Em razão dos gargalos e das restrições dentro de sistemas complexos, é difícil desenvolver programações conforme um sistema simples de regras. Ao contrário, todas as restrições precisam ser consideradas juntas.

A OPT usa a terminologia do “tambor, pulmão e corda” para explicar sua abordagem de planejamento e controle – explicamos essa ideia no [Capítulo 10](#). O centro do trabalho gargalo torna-se um “tambor”, marcando o andamento para o restante da fábrica. Essa “batida do tambor” determina a programação nas áreas não gargalos, puxando o trabalho (a corda) alinhado com a capacidade do gargalo, não com a capacidade do centro do trabalho em questão.

A um gargalo, nunca deverá ser permitido funcionar abaixo da capacidade plena; assim, o “pulmão” de estoque deve ser colocado antes do gargalo, para assegurar que nunca faltará trabalho.

## As cinco etapas da teoria das restrições

Como método prático de sincronizar o fluxo, a TOC enfatiza as cinco etapas seguintes:<sup>8</sup>

1. *Identifique a restrição do sistema* – a parte de um sistema que constitui seu elo mais fraco; pode ser uma restrição física, ou mesmo uma tomada de decisão, ou uma restrição política.
2. *Decida como explorar a restrição* – obtenha a capacidade que for possível da restrição, de preferência, sem mudanças onerosas. Por exemplo, reduza ou elimine qualquer tempo não produtivo no gargalo.
3. *Subordine tudo à restrição* – os elementos que não são restrição do processo são ajustados a um nível no qual a restrição possa operar com eficácia máxima. Após isso, o processo global é avaliado para determinar se a restrição mudou para algum outro ponto do processo. Se a restrição foi eliminada, siga para a etapa 5.
4. *Eleve a restrição* – “elevar” a restrição significa eliminá-la. Essa etapa é considerada somente se as etapas 2 e 3 não foram bem-sucedidas. As grandes mudanças no sistema existente são consideradas nesta etapa.
5. Reinicie a partir da etapa 1.

A [Tabela 15.1](#) mostra algumas das diferenças entre a teoria das restrições e a sincronização enxuta. Comprovadamente, a principal contribuição da TOC para o fluxo regular e sincronizado é sua inclusão da ideia de que os efeitos das restrições do gargalo (a) devem ser priorizadas e (b) podem “perdoar” estoque, se isso significar maximizar a utilização do gargalo. Nem (diferente, por exemplo, do ERP/MRP) requer necessariamente grande investimento em nova tecnologia de informação. Além disso, por tentar melhorar o fluxo de itens ao longo de um processo, pode liberar estoque que, por sua vez, libera o investimento de capital. As afirmações sobre o *payback* financeiro da OPT são frequentemente baseadas nessa liberação de capital e no atravessamento rápido.

**Tabela 15.1** Teoria das restrições comparada com a sincronização enxuta<sup>9</sup>

|                     | Teoria das restrições   | Sincronização enxuta   |
|---------------------|---|--|
| Objetivos globais   | Aumentar a rentabilidade ao aumentar o fluxo de um processo ou operação | Aumentar a rentabilidade ao agregar valor a partir da perspectiva do cliente |
| Medidas de eficácia | ■ Ganho ( <i>throughput</i> )   | ■ Custo  |

|                       |  |   |
|-----------------------|--|---|
|                       | <ul style="list-style-type: none"> <li>■ Estoque</li> <li>■ Despesa operacional</li> </ul> | <ul style="list-style-type: none"> <li>■ Tempo de atravessamento</li> <li>■ Eficiência do valor agregado</li> </ul> |
| Obter melhoria ao ... | Focar as restrições (os “elos mais fracos”) no processo                                    | Eliminar o desperdício e agregar valor ao considerar o processo, a operação ou a rede de suprimento completos       |
| Como implementar      | Um processo contínuo de cinco etapas (veja anteriormente) atuando localmente               | Melhoria contínua enfatizando toda a rede de suprimento   |

## Abordagem enxuta e MRP

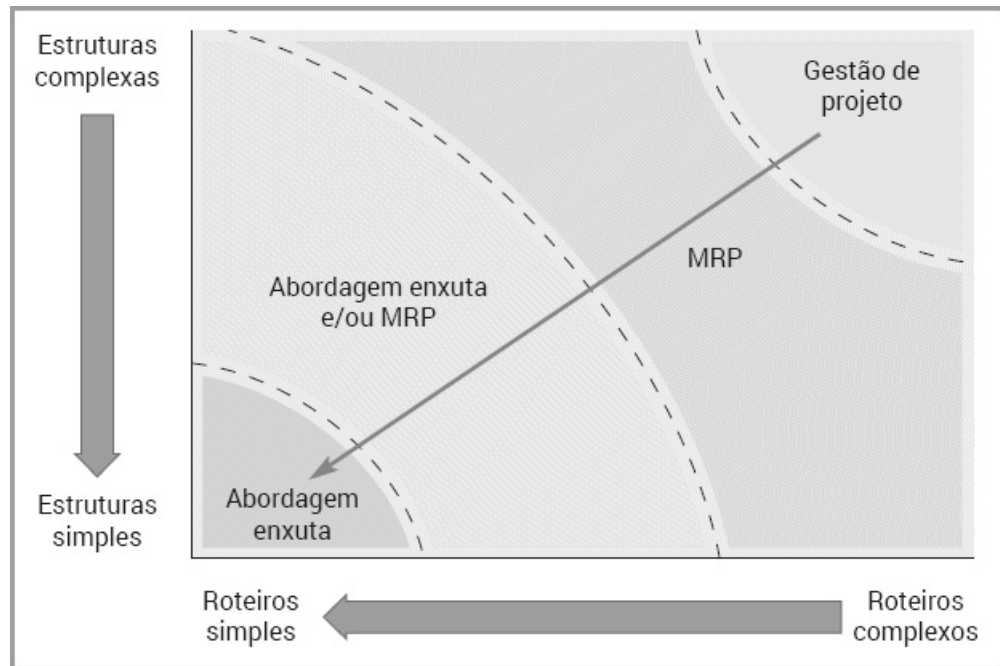
As filosofias operacionais da sincronização enxuta e o MRP parecem ser fundamentalmente opostas. A sincronização enxuta encoraja um sistema “puxado” de planejamento e controle, enquanto o MRP é um sistema “empurrado”. A abordagem enxuta tem propósitos mais amplos do que a atividade de planejamento e controle das operações, enquanto o MRP é basicamente um “mecanismo de cálculo” do planejamento e controle. Todavia, as duas abordagens podem reforçar uma à outra na mesma operação, desde que as respectivas vantagens sejam preservadas. A ironia é que sincronização enxuta e o MRP têm objetivos similares. A programação JIT visa conectar a rede interna e os processos de suprimento externos por meio de esteiras invisíveis, de modo que as partes apenas se movem em resposta a sinais coordenados e sincronizados, derivados da demanda do cliente final. O MRP procura atender à demanda projetada do cliente ao direcionar os itens que forem produzidos à medida que necessários para atender a essa demanda. Entretanto, há diferenças. O MRP é programado pela programação mestre de produção que identifica a demanda futura do produto acabado. Modela um ambiente de *lead time* fixo, usando o poder da computação para calcular quanto e quando cada parte deve ser fabricada. Sua produção está de acordo com os planos de necessidades escalonadas no tempo, centralmente calculada e coordenada. As partes são fabricadas em resposta às instruções centrais. Os distúrbios do dia a dia, como registros de estoque imprecisos, debilitam a autoridade do MRP e podem tornar os planos impraticáveis. Embora o MRP seja excelente para planejamento, é fraco no controle. Por outro lado, a sincronização enxuta visa atender instantaneamente à demanda mediante sistemas de controle simples, baseados no *kanban*. Se o tempo de atravessamento total ( $P$ ) for inferior ao *lead time* da demanda ( $D$ ), os sistemas de sincronização enxuta devem ser capazes de atender a essa demanda. Porém, se a razão  $P:D$  for maior do que 1, alguma produção especulativa será necessária. Repentinamente, se a demanda for maior do que o esperado para certos produtos, o sistema JIT pode não ter condições de atender. A programação “puxada” é um conceito reativo que funciona melhor quando a demanda independente estiver nivelada, e a demanda dependente, sincronizada. Enquanto a sincronização enxuta pode ser boa no controle,

ela é fraca no planejamento.

O MRP também é melhor para lidar com a complexidade, conforme medida pelo número de itens que são processados. Pode tratar de exigências detalhadas, mesmo que sejam “estranhas”. A programação “puxada” da sincronização enxuta é menos capaz de responder instantaneamente às mudanças na demanda, à medida que aumentam a quantidade, as opções e as cores das peças. Assim, os sistemas de produção da sincronização enxuta favorecem os projetos baseados em estruturas de produto mais simples, com grande semelhança entre as partes. Tais disciplinas desafiam a complexidade desnecessária, de modo que mais partes podem ser compradas sob o controle da programação “puxada”.

### **Quando usar a abordagem enxuta, MRP e sistemas combinados**

A [Figura 15.12](#) distingue entre a complexidade das estruturas de um produto e a complexidade dos roteiros de fluxo ao longo das quais devem passar.<sup>10</sup> Estruturas de produto simples que têm roteiros com alta repetibilidade são as principais candidatas ao controle “puxado”. A sincronização enxuta pode facilmente lidar com seus requisitos relativamente diretos. À medida que as estruturas e os roteiros tornam-se mais complexos, o poder da computação é necessário para romper as estruturas do produto e, assim, emitir pedidos aos fornecedores. Em muitos ambientes é possível usar a programação “puxada” para controlar a maioria dos materiais internos. Novamente, os principais candidatos ao controle “puxado” são os materiais usados regularmente a cada semana ou a cada mês. Seu número pode ser aumentado pela padronização do projeto, como indicado pela direção da seta da [Figura 15.12](#). À medida que as estruturas e rotinas tornam-se ainda mais complexas, e os usos das partes tornam-se mais irregulares, as oportunidades de usar a programação “puxada” diminuem. Estruturas muito complexas requerem métodos de planejamento em rede (*veja o Capítulo 19*), para planejamento e controle.



**Figura 15.12** Complexidade como determinante de um sistema apropriado de planejamento e controle.

**Fonte:** Adaptado de VOSS, C.A.; HARRISON, A. Strategies for implementing Jit, In: VOSS, C.A. (ed.) *Just-in-Time Manufacture*, IFs/Springer-Verlag, 1987. Reproduzida com permissão.

## RESPOSTAS RESUMIDAS ÀS QUESTÕES-CHAVE

### O que é produção enxuta?

- Abordagem enxuta é uma abordagem para operações que tentam atender instantaneamente à demanda com qualidade perfeita e sem desperdício. É uma abordagem que difere das práticas de produção tradicionais, de modo que reforça a eliminação de desperdício e agiliza o atravessamento, o que contribui para estoques baixos.
- A habilidade da entrega *just-in-time* não apenas economiza capital de giro (mediante a redução dos níveis de estoque), mas também tem impacto significativo sobre a habilidade de uma operação melhorar sua eficiência intrínseca.
- A filosofia da abordagem enxuta pode ser resumida no que diz respeito a três elementos que se sobrepõem: (a) eliminação do desperdício em todas as suas formas, (b) inclusão de todos os funcionários da operação em sua melhoria e (c) a ideia de que qualquer melhoria deve ser contínua.
- A maioria das ideias da abordagem enxuta é diretamente aplicável às operações de serviço.

### Como a produção enxuta elimina o desperdício?

- A parte mais significativa da filosofia enxuta é seu foco na eliminação de todas as formas de desperdício, definido como qualquer atividade que não agrega valor.



- A abordagem enxuta identifica sete tipos de desperdício que, juntos, formam quatro barreiras para atingir a abordagem enxuta. São elas: desperdício do fluxo irregular (não simplificado), desperdício do suprimento inexato, desperdício da resposta inflexível e desperdício decorrente da variabilidade.

#### **Como a produção enxuta se aplica ao longo da rede de suprimento?**

- A maioria dos conceitos e técnicas da abordagem enxuta, embora geralmente descrita como aplicada aos processos e operações individuais, também se aplica às redes de suprimento inteiras.
- O conceito de cadeia de suprimento enxuta tem sido vinculado a um sistema de controle de tráfego aéreo, em que tenta fornecer “visibilidade e controle em tempo real” a todos os elementos da rede.

#### **Como a sincronização enxuta se compara a outras abordagens?**

- Há outras abordagens que tentam desempenhar a mesma função da abordagem enxuta. Duas alternativas para a abordagem enxuta como método de planejamento e controle são a teoria das restrições (TOC) e o *material requirement planning* (MRP).
- Embora a TOC e o MRP possam parecer abordagens diferentes, podem ser combinadas.
- A forma pela qual podem ser combinados depende da complexidade das estruturas de produto, complexidade do roteiro de produto, características do volume-variedade da operação e nível de controle requerido.

## **ESTUDO DE CASO**

### **Hospital Saint Bridget<sup>11</sup>**

Quando Denize Ahlgren chegou ao Saint Bridget, um dos principais hospitais da área de Göteborg, ela sabia que havia ganhado uma reputação pelo modo de pensar em como os cuidados com a saúde poderiam ser organizados para oferecer níveis superiores de atendimento público a um custo menor para o contribuinte. Na verdade, essa foi uma das razões pelas quais ela assumiu o cargo de chefe de administração (COA). Em particular, Denize estava lendo sobre a iniciativa “Quality Care” (QC) do Saint Bridget. “*Sim, QC obviamente é importante*”, explicou o Dr. Pär Solberg, que, além de suas atividades clínicas, também encabeçou a iniciativa QC, “*mas não pense que se trata apenas de ‘qualidade’. Nós não apenas gastamos dinheiro para melhorar a qualidade dos tratamentos; também queremos melhorar a eficiência. Qualquer dinheiro economizado pela melhoria da eficiência pode então ser investido na melhoria dos resultados clínicos.*”

#### **“Tudo começou com a qualidade”**

Embora administrado por uma empresa privada, Saint Bridget é um pouco diferente de qualquer outro hospital sueco. Para seus pacientes, o tratamento é gratuito, após uma cobrança mínima que é universal na Suécia. O Saint Bridget obtém praticamente todas as suas receitas do governo. No entanto, em termos de como se organiza, está na vanguarda da implementação de ideias mais comuns nos negócios privados. “*Tudo começou com nossos esforços há alguns anos para*



*sermos sistemáticos na maneira como medimos a qualidade", disse Pär Solberg. "Sentimos que a qualidade deve ser relatada de forma sistemática e lógica, para que seja significativa. Ele também deve ter multidisciplinar, e não se concentrar apenas em um aspecto da qualidade. Medimos três aspectos: 'experiência relatada pelo paciente' (ERP), o que o paciente pensa sobre a experiência total de receber o tratamento, 'resultado relatado pelo paciente' (RRP), como o paciente vê a eficácia do tratamento recebido e, o mais importante, 'resultado clínico relatado' (RCR), como os clínicos veem a eficácia do tratamento. É claro que essas três medidas estão interligadas. Então, o RRP, eventualmente, depende do resultado médico (RCR) e quanto desconforto e dor o tratamento desencadeou. Mas também é influenciado pela experiência do paciente (ERP), por exemplo, o quanto mantemos o paciente informado, o quanto nosso pessoal é sensível e assim por diante."*

### **"Medir a qualidade levou naturalmente à melhoria contínua"**

Os processos de medição da qualidade do hospital logo se desenvolveram em uma abordagem mais ampla para a melhoria em geral. Em particular, a ideia de melhoria contínua começou a ser discutida. *"Medir a qualidade levou naturalmente à melhoria contínua", explicou Pär Solberg. "Uma vez tendo indicadores de qualidade mensuráveis, poderíamos estabelecer metas e, o mais importante, poderíamos começar a pensar sobre o que nos impedia de melhorar a qualidade. Isso, por sua vez, levou a uma compreensão de todos os processos que afetavam os indicadores de qualidade. Foi uma grande mudança ver o hospital como um conjunto de processos que controlavam um conjunto de fluxos – fluxos de pacientes através de seus estágios de tratamento, fluxos de pessoal clínico, fluxos de informações, fluxos de produtos farmacêuticos, fluxos de equipamentos, e assim por diante. Foi uma revolução no nosso pensamento. Começamos a examinar esses fluxos e a analisar a forma como impactavam o nosso desempenho, e como poderíamos melhorar os métodos de trabalho que considerávamos significativos para os indicadores de qualidade que queríamos influenciar. Foi quando descobrimos o conceito da abordagem enxuta."*

### **"A melhoria contínua nos apresentou a abordagem enxuta"**

Foi em uma conferência de "melhoria da saúde europeia" que contou com a participação de Pär e outro colega que o Saint Bridget foi apresentado pela primeira vez à abordagem "enxuta". *"Estávamos falando com alguns representantes do National Health Service Institute do Reino Unido, que estava envolvido na introdução de princípios enxutos nos hospitais do Reino Unido. Eles explicaram que a abordagem enxuta era uma técnica de melhoria que aprimorava o fluxo e eliminava o desperdício, e que havia sido utilizada com sucesso em alguns hospitais para desenvolver a melhoria contínua. A abordagem enxuta, eles disseram, tal como desenvolvida pela Toyota, significava fazer as coisas certas no lugar certo, no momento certo, nas quantidades certas, minimizando o desperdício e sendo flexível e aberto a mudanças. Parecia valer a pena acompanhar essa tendência. No entanto, eles admitiram que nem todas as tentativas de introduzir os princípios enxutos obtiveram sucesso."*

### **Tudo poderia facilmente se tornar político**

Intrigado pela conversa, Pär entrou em contato com um dos hospitais do Reino Unido que haviam sido mencionados e falou com Marie Watson, que tinha sido a "Chefe da Abordagem Enxuta" e havia iniciado vários projetos enxutos. Ela disse que um dos problemas que precisou enfrentar foi a insistência de seu chefe executivo em levar várias empresas de consultoria para implementar as ideias da abordagem enxuta. Para tornar as coisas mais confusas, quando um novo chefe executivo foi nomeado, ele trouxe seus próprios consultores preferidos, além dos que já operavam no hospital. Marie não estava feliz com a mudança. *"Antes da mudança dos executivos, tínhamos uma maneira muito clara de como iríamos avançar e espalhar a abordagem enxuta por toda a organização, mas então as coisas ficaram muito menos claras. A ênfase mudou para obter alguns*

resultados rápidos. Mas não foi por isso que ficamos aborrecidos. Originalmente, tratava-se de ter um impacto positivo, envolver as pessoas nas práticas enxutas, engajá-las e dar-lhes capacitação para uma melhoria contínua; havia coisas que eram mensuráveis, mas depois isso mudou para ‘mostre-nos alguns resultados rápidos’. As pessoas estavam se esquecendo do lado cultural da coisa. Além disso, tudo poderia facilmente se tornar ‘político’. Todas as diferentes equipes de consultoria e as iniciativas enxutas internas tinham seus próprios territórios. Por exemplo, nós [a equipe interna de Marie] estávamos prestes a iniciar um estudo de atividades de A&E, quando lhes foi dito que se afastassem da A&E para não atrapalhar a firma de consultoria que atuava lá.”

### **“Não estamos fabricando carros; pessoas são diferentes”**

Pär estava determinado a não cometer os mesmos erros que o hospital de Marie tinha cometido, e consultou bastante seus colegas antes de experimentar quaisquer melhorias enxutas. Alguns eram céticos: *“Não estamos fabricando carros; pessoas são diferentes, e os processos que usamos com as pessoas repetidamente são mais complicados do que os processos usados na fabricação de um carro.”* Além disso, alguns funcionários mais antigos estavam duvidosos sobre mudanças que eles percebiam como ameaçando seu *status* profissional. Em vez de médicos e enfermeiros mantendo papéis separados e definidos, que se concentravam exclusivamente em seu campo de perícia médica, eles eram encorajados a trabalhar (e sentar) juntos em equipes. As equipes também eram responsáveis por sugerir melhorias no processo. Mas a maioria foi convertida. Um clínico sênior, inicialmente, afirmou que *“isso é tudo um monte de lixo. Não faz sentido mapear esse processo, todos sabemos o que acontece: o paciente vai de lá para cá e esta é a solução, e é isso que precisamos fazer.”* No entanto, poucos dias depois ele estava dizendo: *“Nunca havia percebido o que realmente acontece; aquilo que não funcionava, agora vai funcionar, e isso realmente é ótimo, pois eu nunca compreendi. Eu só via a minha parte, mas agora eu entendo o processo inteiro.”*

### **“Isso funciona e torna as coisas melhores para os pacientes”**

Com o passar do tempo, a maior parte (embora não todo) do ceticismo foi superada, principalmente porque, nas palavras de um médico, *“Isso funciona e torna as coisas melhores para os pacientes”*. À medida que outras partes do hospital se tornaram convencidas da eficácia da abordagem enxuta, as melhorias no fluxo e na qualidade dos pacientes começaram a se acumular. Algumas das primeiras melhorias foram relativamente simples, como uma mudança de sinalização (para evitar que os pacientes se perdessem). Outras simplesmente envolviam um rolo de fita amarela. Em vez de o pessoal desperdiçar um tempo precioso procurando equipamentos como desfibriladores, a fita amarela foi usada para marcar um ponto no chão, onde as máquinas eram sempre mantidas. Outra envolveu usar pontos magnéticos em um gráfico de progresso para acompanhar o progresso de cada paciente e indicar quais leitos estavam livres. Algumas eram ainda mais simples, por exemplo, dar alta aos pacientes ao longo do dia, em vez de todos ao mesmo tempo; assim, eles poderiam encontrar um táxi com mais facilidade. Outras melhorias envolveram mais análises, como a redução dos níveis de estoque mantidos (por exemplo, 25.000 pares de luvas cirúrgicas de 500 fornecedores diferentes). Algumas envolveram uma mudança completa nos pressupostos, como a eficácia do departamento de registros médicos. *“Foi incrível. Acabamos de desmentir o mito de que, quando não se conseguia anotações médicas em uma área clínica, isso era culpa dos registros médicos. Mas nunca era. Os médicos tinham anotações em seus carros, em casa, tínhamos milhares de anotações nos escritórios das secretárias, havia anotações em salas, gavetas e armários — elas estavam por toda a parte. E nos perguntamos por que não conseguíamos receber*

as anotações! Duas pessoas caminhavam mais de dez quilômetros por dia para encontrar anotações de casos!" (Pär Solberg)

### **"Precisamos seguir para o próximo nível"**

Denize Ahlgren ficou compreensivelmente impressionada com as melhorias que Pär lhe descreveu; no entanto, Pär estava surpreendentemente aborrecido sobre o futuro. *"OK, eu admito que tivemos alguns ganhos impressionantes com a melhoria contínua e, posteriormente, pela adoção de princípios enxutos. Estou especialmente impressionado com o conceito da Toyota dos sete tipos de desperdício [ver detalhes neste capítulo]. É uma ideia conceitualmente poderosa e muito prática para identificar onde podemos melhorar. Também os funcionários gostam. Mas tudo está acontecendo como um exercício de ticar itens numa lista. A busca de desperdício não é exatamente uma ideia excitante ou radical. Quanto mais eu estudo como a abordagem enxuta ocorreu na Toyota e outras montadoras, mais eu vejo que realmente não abraçamos a filosofia inteira. No entanto, ao mesmo tempo, não estou totalmente convencido de que podemos. Talvez alguns dos que duvidaram estivessem certos; um hospital não é uma fábrica de automóveis, e podemos aplicar apenas algumas ideias da abordagem enxuta."*

Ironicamente, enquanto Pär estava tendo dúvidas, alguns de seus colegas estavam esforçando-se para fazer mais. Um clínico em particular, Fredrik Olsen, médico-chefe da Clínica de Dor Lombar do Saint Bridget, pensou que sua clínica poderia se beneficiar de uma abordagem mais radical. *"Precisamos seguir para o próximo nível. Toda a filosofia da Toyota está preocupada com o fluxo síncrono e regular, mas não entramos de cabeça nisso. Eu sei que somos relutantes em falar sobre 'estoques' de pacientes, mas isso é exatamente o que são as salas de espera. Eles são 'estoques' de pessoas, e nós as usamos exatamente da mesma forma como os fabricantes faziam antes da abordagem enxuta – como um estoque de proteção contra os desajustes de curto prazo entre suprimento e demanda. O que devemos fazer é tratar das causas raiz desse desajuste. As salas de espera estão nos impedindo de mover para um fluxo regular, de valor agregado para os nossos pacientes."*

Fredrik continuou a fazer o que Denize pensou ser uma proposta interessante, mas radical. Ele propôs a desativação da sala de espera atual para a Clínica de Dor Lombar e a substituição por dois consultórios extras que seriam adicionados aos dois consultórios existentes. Os pacientes receberiam agendamentos para horários específicos em vez de serem solicitados a chegar "na hora" (efetivamente em lotes), como era feito na época. Uma enfermeira anotaria os detalhes dos pacientes e realizaria alguns testes preliminares, após o que elas chamariam o médico especialista. Os níveis de pessoal durante os horários da clínica seriam controlados por enfermeiras que também monitorariam a chegada dos pacientes, os encaminhariam para os consultórios e organizariam qualquer tratamento posterior (para exames de ressonância magnética, por exemplo).

Denize não tinha certeza sobre a proposta de Fredrik. *"Parece que pode ser um passo longe demais. Os pacientes esperam aguardar até que um médico possa vê-los, então não tenho certeza de quais benefícios resultariam da proposta. E qual é o sentido de equipar dois novos consultórios se eles não forem totalmente utilizados?"*

### **QUESTÕES**

1. Que benefícios foram obtidos pelo Saint Bridget ao adotar primeiro uma abordagem de melhoria contínua, depois enxuta?
2. Você acredita que Pär Solberg está certo ao pensar que existe um limite para o ponto ao qual um hospital pode chegar adotando as ideias da abordagem enxuta?
3. No *website* do Saint Bridget existem várias referências ao seu programa de "Cuidado de qualidade", mas não às suas

iniciativas enxutas, embora a abordagem enxuta seja considerada importante pela maioria dos médicos e administradores no hospital. Qual o motivo disso, na sua opinião?

4. Denize não consegue enxergar os benefícios da proposta de Fredrik. Qual o motivo disso, na sua opinião?
5. Algum benefício da remoção da sala de espera na clínica compensa a subutilização dos quatro consultórios que Fredrik prevê?



## PROBLEMAS E APLICAÇÕES

1

Pense na última vez que você viajou de avião. Analise a viagem em termos do tempo com valor agregado (realmente, ir a algum lugar) e do tempo sem valor agregado (tempo perdido em filas etc.), desde o momento que saiu de casa até o momento exato que chegou a seu destino final. Calcule o tempo de valor agregado da viagem.

2

Um processo simples tem quatro estágios – A, B, C e D. A carga média de trabalho necessário para processar os itens passa por esses estágios, do seguinte modo: estágio A = 68 minutos, estágio B = 55 minutos, estágio C = 72 minutos e estágio D = 60 minutos. Uma amostragem do trabalho em processo entre cada estágio revela o seguinte: entre os estágios A e B há 82 itens, entre os estágios B e C, 190 itens, e entre os estágios C e D, 89 itens.

- (a) Usando a lei de Little (veja no [Capítulo 6](#)), calcule o tempo de atravessamento do processo.
- (b) Qual é a eficiência de atravessamento do processo?

3

No exemplo anterior, o gerente de produção encarregado do processo realoca o trabalho em cada estágio para melhorar o “balanceamento” do processo. Agora, cada estágio tem uma carga média de trabalho de 64 minutos. O trabalho em processo na frente dos estágios B, C e D é de 75, 80 e 82 unidades, respectivamente. Como isso mudou a eficiência de atravessamento do processo?

4

Um processo de produção precisa produzir 1.400 unidades do produto X, 840 do produto Y e 420 do produto Z em um período de quatro semanas. Se o processo funciona sete horas por dia e cinco dias por semana, elabore uma programação com mistura de modelos em termos do número de cada um dos produtos que precisam ser produzidos por hora para satisfazer à demanda.

5

Reveja a seção “Operações na Prática” no início deste capítulo (e qualquer outra fonte de informação sobre o sistema de produção da Toyota) e (a) liste todas as diferentes técnicas e práticas adotadas pela Toyota e (b) descreva como os objetivos da produção (qualidade, velocidade, confiabilidade, flexibilidade e custo) são influenciados pelas práticas adotadas pela Toyota.

Considere como os princípios de redução do *set-up* podem ser usados nas seguintes situações:

- (a) trocar um pneu no acostamento de uma rodovia (após um furo);
- (b) limpar um avião e prepará-lo para o próximo voo, no intervalo de tempo entre a aterrissagem e o desembarque dos passageiros até a mesma aeronave estar pronta para decolar em um próximo voo;

- (c) o tempo entre o término de um procedimento cirúrgico em uma sala de cirurgia de um hospital e o início do próximo procedimento na mesma sala;
- (d) as atividades de *pitstop* durante uma corrida de Fórmula 1 (como se compara ao item (a), acima?).

## LEITURA COMPLEMENTAR SELECIONADA

BICHENO, J.; HOLWEG, M. *The lean toolbox: the essential guide to lean transformation*. 4. ed. Buckingham: PICSIE Books, 2010.

Um guia prático preparado por duas autoridades europeias em questão de práticas enxutas.

HOLWEG, M. The genealogy of lean production. *Journal of Operations Management*, v. 25, p. 420-437, 2007.

Uma excelente visão de como as ideias enxutas se desenvolveram.

HUMBLE, J.; MOLESKY, J.; O'REILLY, B. *Lean enterprise: how high performance organizations innovate at scale*. Sebastopol: O'Reilly Media, 2015.

Exemplo interessante de como (algumas) ideias enxutas se espalharam para fora da área convencional de administração da produção.

MANN, D. *Creating a lean culture*. 2. ed. New York: Productivity Press, 2010.

Trata do lado suave da filosofia enxuta.

MODIG, N.; AHLSTROM, P. *This is lean: resolving the efficiency paradox*. Estocolmo: Rheologica Publishing, 2012.

Um tratamento prático e de fácil leitura, embora intelectualmente coerente.

WOMACK, J.P.; JONES, D.T. *Lean thinking: banish waste and create wealth in your corporation*. New York: Simon & Schuster, 1996.

Algumas das lições de *The Machine that Changed the World* (abaixo), mas aplicadas em um contexto mais amplo.

\_\_\_\_\_; ROOS, D. *The machine that changed the world*. New York: Rawson Associates, 1990.

Comprovadamente o livro mais influente sobre a prática de gestão de operações dos últimos 50 anos. Embora fortemente baseado no setor automobilístico, contribuiu muito para estabelecer a filosofia enxuta.