

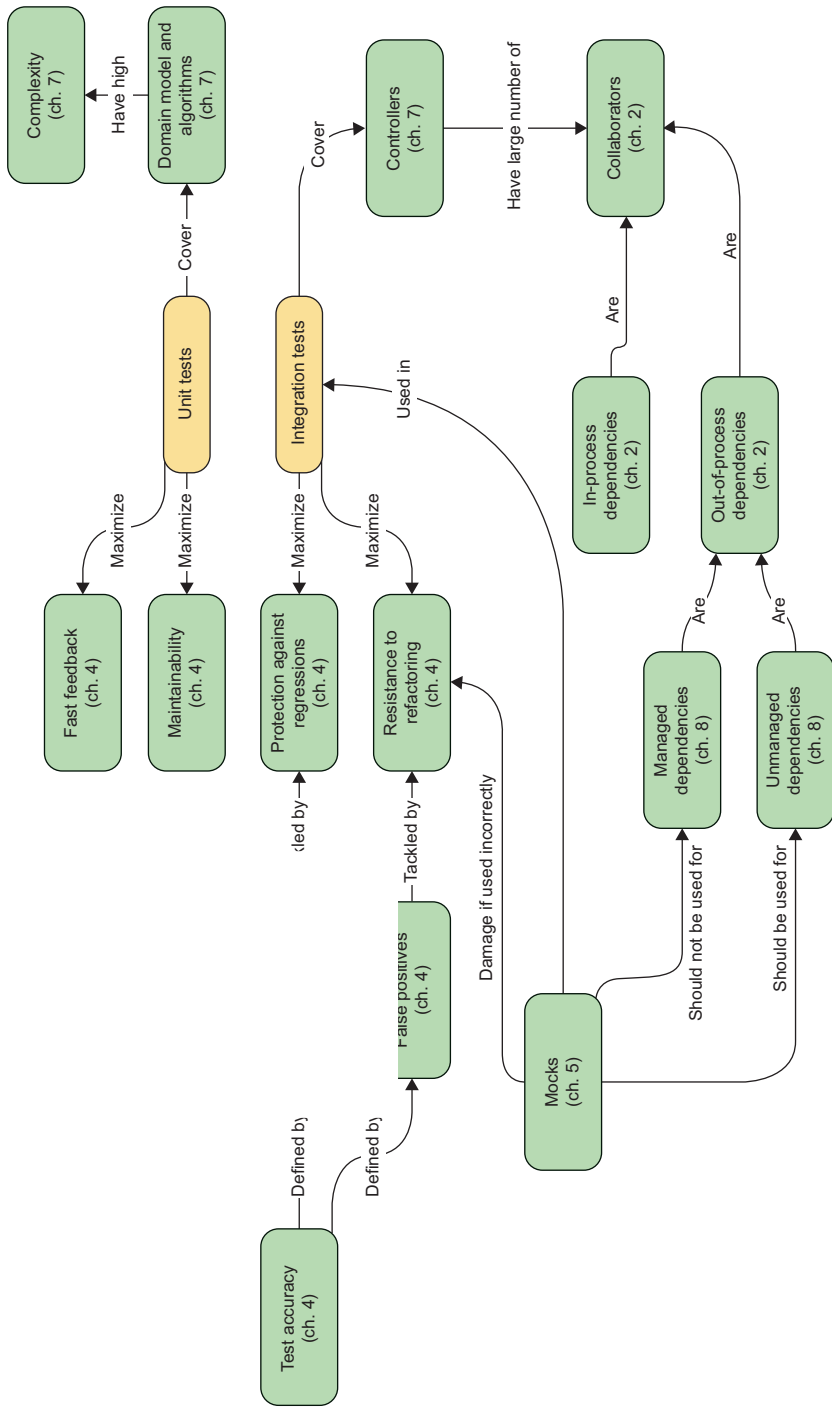


# Unit Testing

Principles, Practices, and Patterns

Vladimir Khorikov

# Chapter Map



# *Unit Testing: Principles, Practices, and Patterns*

VLADIMIR KHORIKOV



MANNING  
SHELTER ISLAND

For online information and ordering of this and other Manning books, please visit [www.manning.com](http://www.manning.com). The publisher offers discounts on this book when ordered in quantity. For more information, please contact


Special Sales Department  
Manning Publications Co.  
20 Baldwin Road  
PO Box 761  
Shelter Island, NY 11964  
Email: [orders@manning.com](mailto:orders@manning.com)

©2020 by Manning Publications Co. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by means electronic, mechanical, photocopying, or otherwise, without prior written permission of the publisher.

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in the book, and Manning Publications was aware of a trademark claim, the designations have been printed in initial caps or all caps.

☺ Recognizing the importance of preserving what has been written, it is Manning's policy to have the books we publish printed on acid-free paper, and we exert our best efforts to that end. Recognizing also our responsibility to conserve the resources of our planet, Manning books are printed on paper that is at least 15 percent recycled and processed without the use of elemental chlorine.

 Manning Publications Co.  
20 Baldwin Road  
PO Box 761  
Shelter Island, NY 11964

Acquisitions editor: Mike Stephens  
Development editor: Marina Michaels  
Technical development editor: Sam Zaydel  
Review editor: Aleksandar Dragosavljević  
Production editor: Anthony Calcara  
Copy editor: Tiffany Taylor  
ESL copyeditor: Frances Buran  
Proofreader: Keri Hales  
Technical proofreader: Alessandro Campeis  
Typesetter: Dennis Dalinnik  
Cover designer: Marija Tudor

ISBN: 9781617296277

Printed in the United States of America

*To my wife, Nina*



# *brief contents*

---

<b>PART 1</b>	<b>THE BIGGER PICTURE.....</b>	<b>1</b>
	1 ■ The goal of unit testing	3
	2 ■ What is a unit test?	20
	3 ■ The anatomy of a unit test	41
<b>PART 2</b>	<b>MAKING YOUR TESTS WORK FOR YOU.....</b>	<b>65</b>
	4 ■ The four pillars of a good unit test	67
	5 ■ Mocks and test fragility	92
	6 ■ Styles of unit testing	119
	7 ■ Refactoring toward valuable unit tests	151
<b>PART 3</b>	<b>INTEGRATION TESTING.....</b>	<b>183</b>
	8 ■ Why integration testing?	185
	9 ■ Mocking best practices	216
	10 ■ Testing the database	229
<b>PART 4</b>	<b>UNIT TESTING ANTI-PATTERNS.....</b>	<b>257</b>
	11 ■ Unit testing anti-patterns	259





# contents

---

*preface* xiv  
*acknowledgments* xv  
*about this book* xvi  
*about the author* xix  
*about the cover illustration* xx

## PART 1 THE BIGGER PICTURE .....1

### **1** *The goal of unit testing* 3

- 1.1 The current state of unit testing 4
- 1.2 The goal of unit testing 5
  - What makes a good or bad test?* 7
- 1.3 Using coverage metrics to measure test suite quality 8
  - Understanding the code coverage metric* 9 ■ *Understanding the branch coverage metric* 10 ■ *Problems with coverage metrics* 12
  - Aiming at a particular coverage number* 15
- 1.4 What makes a successful test suite? 15
  - It's integrated into the development cycle* 16 ■ *It targets only the most important parts of your code base* 16 ■ *It provides maximum value with minimum maintenance costs* 17
- 1.5 What you will learn in this book 17

## 2 What is a unit test? 20

- 2.1 The definition of “unit test” 21
  - The isolation issue: The London take* 21
  - *The isolation issue: The classical take* 27
- 2.2 The classical and London schools of unit testing 30
  - How the classical and London schools handle dependencies* 30
- 2.3 Contrasting the classical and London schools of unit testing 34
  - Unit testing one class at a time* 34
  - *Unit testing a large graph of interconnected classes* 35
  - *Revealing the precise bug location* 36
  - Other differences between the classical and London schools* 36
- 2.4 Integration tests in the two schools 37
  - End-to-end tests are a subset of integration tests* 38

## 3 The anatomy of a unit test 41

- 3.1 How to structure a unit test 42
  - Using the AAA pattern* 42
  - *Avoid multiple arrange, act, and assert sections* 43
  - *Avoid if statements in tests* 44
  - How large should each section be?* 45
  - *How many assertions should the assert section hold?* 47
  - *What about the teardown phase?* 47
  - *Differentiating the system under test* 47
  - Dropping the arrange, act, and assert comments from tests* 48
- 3.2 Exploring the xUnit testing framework 49
- 3.3 Reusing test fixtures between tests 50
  - High coupling between tests is an anti-pattern* 52
  - *The use of constructors in tests diminishes test readability* 52
  - *A better way to reuse test fixtures* 52
- 3.4 Naming a unit test 54
  - Unit test naming guidelines* 56
  - *Example: Renaming a test toward the guidelines* 56
- 3.5 Refactoring to parameterized tests 58
  - Generating data for parameterized tests* 60
- 3.6 Using an assertion library to further improve test readability 62

## PART 2 MAKING YOUR TESTS WORK FOR YOU .....65

## 4 *The four pillars of a good unit test* 67

- 4.1 Diving into the four pillars of a good unit test 68
  - The first pillar: Protection against regressions* 68
  - *The second pillar: Resistance to refactoring* 69
  - *What causes false positives?* 71
  - *Aim at the end result instead of implementation details* 74
- 4.2 The intrinsic connection between the first two attributes 76
  - Maximizing test accuracy* 76
  - *The importance of false positives and false negatives: The dynamics* 78
- 4.3 The third and fourth pillars: Fast feedback and maintainability 79
- 4.4 In search of an ideal test 80
  - Is it possible to create an ideal test?* 81
  - *Extreme case #1: End-to-end tests* 81
  - *Extreme case #2: Trivial tests* 82
  - *Extreme case #3: Brittle tests* 83
  - *In search of an ideal test: The results* 84
- 4.5 Exploring well-known test automation concepts 87
  - Breaking down the Test Pyramid* 87
  - *Choosing between black-box and white-box testing* 89

## 5 *Mocks and test fragility* 92

- 5.1 Differentiating mocks from stubs 93
  - The types of test doubles* 93
  - *Mock (the tool) vs. mock (the test double)* 94
  - *Don't assert interactions with stubs* 96
  - *Using mocks and stubs together* 97
  - *How mocks and stubs relate to commands and queries* 97
- 5.2 Observable behavior vs. implementation details 99
  - Observable behavior is not the same as a public API* 99
  - *Leaking implementation details: An example with an operation* 100
  - *Well-designed API and encapsulation* 103
  - *Leaking implementation details: An example with state* 104
- 5.3 The relationship between mocks and test fragility 106
  - Defining hexagonal architecture* 106
  - *Intra-system vs. inter-system communications* 110
  - *Intra-system vs. inter-system communications: An example* 111

- 5.4 The classical vs. London schools of unit testing, revisited 114
  - Not all out-of-process dependencies should be mocked out* 115
  - Using mocks to verify behavior* 116

## 6 *Styles of unit testing* 119

- 6.1 The three styles of unit testing 120
  - Defining the output-based style* 120
  - *Defining the state-based style* 121
  - *Defining the communication-based style* 122
- 6.2 Comparing the three styles of unit testing 123
  - Comparing the styles using the metrics of protection against regressions and feedback speed* 124
  - *Comparing the styles using the metric of resistance to refactoring* 124
  - *Comparing the styles using the metric of maintainability* 125
  - *Comparing the styles: The results* 127
- 6.3 Understanding functional architecture 128
  - What is functional programming?* 128
  - *What is functional architecture?* 132
  - *Comparing functional and hexagonal architectures* 133
- 6.4 Transitioning to functional architecture and output-based testing 135
  - Introducing an audit system* 135
  - *Using mocks to decouple tests from the filesystem* 137
  - *Refactoring toward functional architecture* 140
  - *Looking forward to further developments* 146
- 6.5 Understanding the drawbacks of functional architecture 146
  - Applicability of functional architecture* 147
  - *Performance drawbacks* 148
  - *Increase in the code base size* 149

## 7 *Refactoring toward valuable unit tests* 151

- 7.1 Identifying the code to refactor 152
  - The four types of code* 152
  - *Using the Humble Object pattern to split overcomplicated code* 155
- 7.2 Refactoring toward valuable unit tests 158
  - Introducing a customer management system* 158
  - *Take 1: Making implicit dependencies explicit* 160
  - *Take 2: Introducing an application services layer* 160
  - *Take 3: Removing complexity from the application service* 163
  - *Take 4: Introducing a new Company class* 164

- 7.3 Analysis of optimal unit test coverage 167
  - Testing the domain layer and utility code* 167
  - *Testing the code from the other three quadrants* 168
  - *Should you test preconditions?* 169
- 7.4 Handling conditional logic in controllers 169
  - Using the CanExecute/Execute pattern* 172
  - *Using domain events to track changes in the domain model* 175
- 7.5 Conclusion 178

## PART 3 INTEGRATION TESTING.....183

### 8 Why integration testing? 185

- 8.1 What is an integration test? 186
  - The role of integration tests* 186
  - *The Test Pyramid revisited* 187
  - *Integration testing vs. failing fast* 188
- 8.2 Which out-of-process dependencies to test directly 190
  - The two types of out-of-process dependencies* 190
  - *Working with both managed and unmanaged dependencies* 191
  - *What if you can't use a real database in integration tests?* 192
- 8.3 Integration testing: An example 193
  - What scenarios to test?* 194
  - *Categorizing the database and the message bus* 195
  - *What about end-to-end testing?* 195
  - Integration testing: The first try* 196
- 8.4 Using interfaces to abstract dependencies 197
  - Interfaces and loose coupling* 198
  - *Why use interfaces for out-of-process dependencies?* 199
  - *Using interfaces for in-process dependencies* 199
- 8.5 Integration testing best practices 200
  - Making domain model boundaries explicit* 200
  - *Reducing the number of layers* 200
  - *Eliminating circular dependencies* 202
  - Using multiple act sections in a test* 204
- 8.6 How to test logging functionality 205
  - Should you test logging?* 205
  - *How should you test logging?* 207
  - *How much logging is enough?* 212
  - How do you pass around logger instances?* 212
- 8.7 Conclusion 213

## 9 Mocking best practices 216

- 9.1 Maximizing mocks' value 217
  - Verifying interactions at the system edges* 219
  - Replacing mocks with spies* 222
  - What about IDomainLogger?* 224
- 9.2 Mocking best practices 225
  - Mocks are for integration tests only* 225
  - Not just one mock per test* 225
  - Verifying the number of calls* 226
  - Only mock types that you own* 227

## 10 Testing the database 229

- 10.1 Prerequisites for testing the database 230
  - Keeping the database in the source control system* 230
  - Reference data is part of the database schema* 231
  - Separate instance for every developer* 232
  - State-based vs. migration-based database delivery* 232
- 10.2 Database transaction management 234
  - Managing database transactions in production code* 235
  - Managing database transactions in integration tests* 242
- 10.3 Test data life cycle 243
  - Parallel vs. sequential test execution* 243
  - Clearing data between test runs* 244
  - Avoid in-memory databases* 246
- 10.4 Reusing code in test sections 246
  - Reusing code in arrange sections* 246
  - Reusing code in act sections* 249
  - Reusing code in assert sections* 250
  - Does the test create too many database transactions?* 251
- 10.5 Common database testing questions 252
  - Should you test reads?* 252
  - Should you test repositories?* 253
- 10.6 Conclusion 254

## PART 3 UNIT TESTING ANTI-PATTERNS.....257

### 11 Unit testing anti-patterns 259

- 11.1 Unit testing private methods 260
  - Private methods and test fragility* 260
  - Private methods and insufficient coverage* 260
  - When testing private methods is acceptable* 261
- 11.2 Exposing private state 263
- 11.3 Leaking domain knowledge to tests 264

11.4	Code pollution	266
11.5	Mocking concrete classes	268
11.6	Working with time	271
	<i>Time as an ambient context</i>	271
	<i>Time as an explicit dependency</i>	272
11.7	Conclusion	273
	<i>index</i>	275

## *preface*

---

I remember my first project where I tried out unit testing. It went relatively well; but after it was finished, I looked at the tests and thought that a lot of them were a pure waste of time. Most of my unit tests spent a great deal of time setting up expectations and wiring up a complicated web of dependencies—all that, just to check that the three lines of code in my controller were correct. I couldn't pinpoint what exactly was wrong with the tests, but my sense of proportion sent me unambiguous signals that something was off.

Luckily, I didn't abandon unit testing and continued applying it in subsequent projects. However, disagreement with common (at that time) unit testing practices has been growing in me ever since. Throughout the years, I've written a lot about unit testing. In those writings, I finally managed to crystallize what exactly was wrong with my first tests and generalized this knowledge to broader areas of unit testing. This book is a culmination of all my research, trial, and error during that period—compiled, refined, and distilled.

I come from a mathematical background and strongly believe that guidelines in programming, like theorems in math, should be derived from first principles. I've tried to structure this book in a similar way: start with a blank slate by not jumping to conclusions or throwing around unsubstantiated claims, and gradually build my case from the ground up. Interestingly enough, once you establish such first principles, guidelines and best practices often flow naturally as mere implications.

I believe that unit testing is becoming a de facto requirement for software projects, and this book will give you everything you need to create valuable, highly maintainable tests.



## *acknowledgments*

---

This book was a lot of work. Even though I was prepared mentally, it was still much more work than I could ever have imagined.

A big “thank you” to Sam Zaydel, Alessandro Campeis, Frances Buran, Tiffany Taylor, and especially Marina Michaels, whose invaluable feedback helped shape the book and made me a better writer along the way. Thanks also to everyone else at Manning who worked on this book in production and behind the scenes.

I’d also like to thank the reviewers who took the time to read my manuscript at various stages during its development and who provided valuable feedback: Aaron Barton, Alessandro Campeis, Conor Redmond, Dror Helper, Greg Wright, Hemant Koneru, Jeremy Lange, Jorge Ezequiel Bo, Jort Rodenburg, Mark Nenadov, Marko Umek, Markus Matzker, Srihari Sridharan, Stephen John Warnett, Sumant Tambe, Tim van Deurzen, and Vladimir Kuptsov.

Above all, I would like to thank my wife Nina, who supported me during the whole process.

## *about this book*

---

*Unit Testing: Principles, Practices, and Patterns* provides insights into the best practices and common anti-patterns that surround the topic of unit testing. After reading this book, armed with your newfound skills, you'll have the knowledge needed to become an expert at delivering successful projects that are easy to maintain and extend, thanks to the tests you build along the way.

### ***Who should read this book***

Most online and print resources have one drawback: they focus on the basics of unit testing but don't go much beyond that. There's a lot of value in such resources, but the learning doesn't end there. There's a next level: not just writing tests, but doing it in a way that gives you the best return on your efforts. When you reach this point on the learning curve, you're pretty much left to your own devices to figure out how to get to the next level.

This book takes you to that next level. It teaches a scientific, precise definition of the ideal unit test. That definition provides a universal frame of reference, which will help you look at many of your tests in a new light and see which of them contribute to the project and which must be refactored or removed.

If you don't have much experience with unit testing, you'll learn a lot from this book. If you're an experienced programmer, you most likely already understand some of the ideas taught in this book. The book will help you articulate why the techniques and best practices you've been using all along are so helpful. And don't underestimate this skill: the ability to clearly communicate your ideas to colleagues is priceless.

## ***How this book is organized: A roadmap***

The book's 11 chapters are divided into 4 parts. Part 1 introduces unit testing and gives a refresher on some of the more generic unit testing principles:

- Chapter 1 defines the goal of unit testing and gives an overview of how to differentiate a good test from a bad one.
- Chapter 2 explores the definition of *unit test* and discusses the two schools of unit testing.
- Chapter 3 provides a refresher on some basic topics, such as structuring of unit tests, reusing test fixtures, and test parameterization.

Part 2 gets to the heart of the subject—it shows what makes a good unit test and provides details about how to refactor your tests toward being more valuable:

- Chapter 4 defines the four pillars that form a good unit test and provide a common frame of reference that is used throughout the book.
- Chapter 5 builds a case for mocks and explores their relation to test fragility.
- Chapter 6 examines the three styles of unit testing, along with which of those styles produces tests of the best quality and why.
- Chapter 7 teaches you how to refactor away from bloated, overcomplicated tests and achieve tests that provide maximum value with minimum maintenance costs.

Part 3 explores the topic of integration testing:

- Chapter 8 looks at integration testing in general along with its benefits and trade-offs.
- Chapter 9 discusses mocks and how to use them in a way that benefits your tests the most.
- Chapter 10 explores working with relational databases in tests.

Part 4's chapter 11 covers common unit testing anti-patterns, some of which you've possibly encountered before.

## ***About the Code***

The code samples are written in C#, but the topics they illustrate are applicable to any object-oriented language, such as Java or C++. C# is just the language that I happen to work with the most.

I tried not to use any C#-specific language features, and I made the sample code as simple as possible, so you shouldn't have any trouble understanding it. You can download all of the code samples online at [www.manning.com/books/unit-testing](http://www.manning.com/books/unit-testing).

### ***liveBook discussion forum***

Purchase of *Unit Testing: Principles, Practices, and Patterns* includes free access to a private web forum run by Manning Publications where you can make comments about the book, ask technical questions, and receive help from the author and from other users. To access the forum, go to <https://livebook.manning.com/#!/book/unit-testing/discussion>. You can also learn more about Manning's forums and the rules of conduct at <https://livebook.manning.com/#!/discussion>.

Manning's commitment to our readers is to provide a venue where a meaningful dialogue between individual readers and between readers and the author can take place. It is not a commitment to any specific amount of participation on the part of the author, whose contribution to the forum remains voluntary (and unpaid). We suggest you try asking the author some challenging questions lest his interest stray! The forum and the archives of previous discussions will be accessible from the publisher's website as long as the book is in print.

### ***Other online resources***

- My blog is at [EnterpriseCraftsmanship.com](http://EnterpriseCraftsmanship.com).
- I also have an online course about unit testing (in the works, as of this writing), which you can enroll in at [UnitTestingCourse.com](http://UnitTestingCourse.com).

## *about the author*

---

VLADIMIR KHORIKOV is a software engineer, Microsoft MVP, and Pluralsight author. He has been professionally involved in software development for over 15 years, including mentoring teams on the ins and outs of unit testing. During the past several years, Vladimir has written several popular blog post series and an online training course on the topic of unit testing. The biggest advantage of his teaching style, and the one students often praise, is his tendency to have a strong theoretic background, which he then applies to practical examples.

## *about the cover illustration*

---

The figure on the cover of *Unit Testing: Principles, Practices, and Patterns* is captioned “Esthinienne.” The illustration is taken from a collection of dress costumes from various countries by Jacques Grasset de Saint-Sauveur (1757–1810), titled *Costumes Civils Actuels de Tous les Peuples Connus*, published in France in 1788. Each illustration is finely drawn and colored by hand. The rich variety of Grasset de Saint-Sauveur’s collection reminds us vividly of how culturally apart the world’s towns and regions were just 200 years ago. Isolated from each other, people spoke different dialects and languages. In the streets or in the countryside, it was easy to identify where they lived and what their trade or station in life was just by their dress.

The way we dress has changed since then and the diversity by region, so rich at the time, has faded away. It is now hard to tell apart the inhabitants of different continents, let alone different towns, regions, or countries. Perhaps we have traded cultural diversity for a more varied personal life—certainly for a more varied and fast-paced technological life.

At a time when it is hard to tell one computer book from another, Manning celebrates the inventiveness and initiative of the computer business with book covers based on the rich diversity of regional life of two centuries ago, brought back to life by Grasset de Saint-Sauveur’s pictures.

# Part 1

## *The bigger picture*

**T**his part of the book will get you up to speed with the current state of unit testing. In chapter 1, I'll define the goal of unit testing and give an overview of how to differentiate a good test from a bad one. We'll talk about coverage metrics and discuss properties of a good unit test in general.

In chapter 2, we'll look at the definition of *unit test*. A seemingly minor disagreement over this definition has led to the formation of two schools of unit testing, which we'll also dive into. Chapter 3 provides a refresher on some basic topics, such as structuring of unit tests, reusing test fixtures, and test parametrization.





# *The goal of unit testing*

---



## ***This chapter covers***

- The state of unit testing
- The goal of unit testing
- Consequences of having a bad test suite
- Using coverage metrics to measure test suite quality
- Attributes of a successful test suite

Learning unit testing doesn't stop at mastering the technical bits of it, such as your favorite test framework, mocking library, and so on. There's much more to unit testing than the act of writing tests. You always have to strive to achieve the best return on the time you invest in unit testing, minimizing the effort you put into tests and maximizing the benefits they provide. Achieving both things isn't an easy task.

It's fascinating to watch projects that have achieved this balance: they grow effortlessly, don't require much maintenance, and can quickly adapt to their customers' ever-changing needs. It's equally frustrating to see projects that failed to do so. Despite all the effort and an impressive number of unit tests, such projects drag on slowly, with lots of bugs and upkeep costs.

That's the difference between various unit testing techniques. Some yield great outcomes and help maintain software quality. Others don't: they result in tests that don't contribute much, break often, and require a lot of maintenance in general.

What you learn in this book will help you differentiate between good and bad unit testing techniques. You'll learn how to do a cost-benefit analysis of your tests and apply proper testing techniques in your particular situation. You'll also learn how to avoid common anti-patterns—patterns that may make sense at first but lead to trouble down the road.

But let's start with the basics. This chapter gives a quick overview of the state of unit testing in the software industry, describes the goal behind writing and maintaining tests, and provides you with the idea of what makes a test suite successful.

## 1.1 *The current state of unit testing*

For the past two decades, there's been a push toward adopting unit testing. The push has been so successful that unit testing is now considered mandatory in most companies. Most programmers practice unit testing and understand its importance. There's no longer any dispute as to whether you should do it. Unless you're working on a throwaway project, the answer is, yes, you do.

When it comes to enterprise application development, almost every project includes at least some unit tests. A significant percentage of such projects go far beyond that: they achieve good code coverage with lots and lots of unit and integration tests. The ratio between the production code and the test code could be anywhere between 1:1 and 1:3 (for each line of production code, there are one to three lines of test code). Sometimes, this ratio goes much higher than that, to a whopping 1:10.

But as with all new technologies, unit testing continues to evolve. The discussion has shifted from "Should we *write* unit tests?" to "What does it mean to write *good* unit tests?" This is where the main confusion still lies.

You can see the results of this confusion in software projects. Many projects have automated tests; they may even have a lot of them. But the existence of those tests often doesn't provide the results the developers hope for. It can still take programmers a lot of effort to make progress in such projects. New features take forever to implement, new bugs constantly appear in the already implemented and accepted functionality, and the unit tests that are supposed to help don't seem to mitigate this situation at all. They can even make it worse.

It's a horrible situation for anyone to be in—and it's the result of having unit tests that don't do their job properly. The difference between good and bad tests is not merely a matter of taste or personal preference, it's a matter of succeeding or failing at this critical project you're working on.

It's hard to overestimate the importance of the discussion of what makes a good unit test. Still, this discussion isn't occurring much in the software development industry

today. You'll find a few articles and conference talks online, but I've yet to see any comprehensive material on this topic.

The situation in books isn't any better; most of them focus on the basics of unit testing but don't go much beyond that. Don't get me wrong. There's a lot of value in such books, especially when you are just starting out with unit testing. However, the learning doesn't end with the basics. There's a next level: not just writing tests, but doing unit testing in a way that provides you with the best return on your efforts. When you reach this point, most books pretty much leave you to your own devices to figure out how to get to that next level.

This book takes you there. It teaches a precise, scientific definition of the ideal unit test. You'll see how this definition can be applied to practical, real-world examples. My hope is that this book will help you understand why your particular project may have gone sideways despite having a good number of tests, and how to correct its course for the better.

You'll get the most value out of this book if you work in enterprise application development, but the core ideas are applicable to any software project.

### What is an enterprise application?

An *enterprise application* is an application that aims at automating or assisting an organization's inner processes. It can take many forms, but usually the characteristics of an enterprise software are

- High business logic complexity
- Long project lifespan
- Moderate amounts of data
- Low or moderate performance requirements

## 1.2 The goal of unit testing

Before taking a deep dive into the topic of unit testing, let's step back and consider the goal that unit testing helps you to achieve. It's often said that unit testing practices lead to a better design. And it's true: the necessity to write unit tests for a code base normally leads to a better design. But that's not the main goal of unit testing; it's merely a pleasant side effect.

### The relationship between unit testing and code design

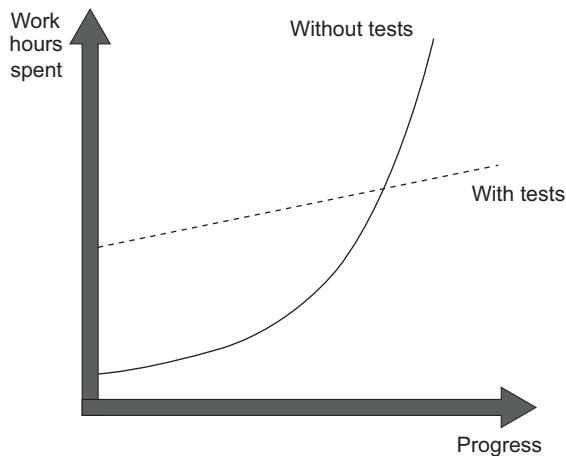
The ability to unit test a piece of code is a nice litmus test, but it only works in one direction. It's a good negative indicator—it points out poor-quality code with relatively high accuracy. If you find that code is hard to unit test, it's a strong sign that the code needs improvement. The poor quality usually manifests itself in *tight coupling*, which means different pieces of production code are not decoupled from each other enough, and it's hard to test them separately.

**(continued)**

Unfortunately, the ability to unit test a piece of code is a bad positive indicator. The fact that you can easily unit test your code base doesn't necessarily mean it's of good quality. The project can be a disaster even when it exhibits a high degree of decoupling.

What is the goal of unit testing, then? The goal is to enable sustainable growth of the software project. The term *sustainable* is key. It's quite easy to grow a project, especially when you start from scratch. It's much harder to sustain this growth over time.

Figure 1.1 shows the growth dynamic of a typical project without tests. You start off quickly because there's nothing dragging you down. No bad architectural decisions have been made yet, and there isn't any existing code to worry about. As time goes by, however, you have to put in more and more hours to make the same amount of progress you showed at the beginning. Eventually, the development speed slows down significantly, sometimes even to the point where you can't make any progress whatsoever.



**Figure 1.1** The difference in growth dynamics between projects with and without tests. A project without tests has a head start but quickly slows down to the point that it's hard to make any progress.

This phenomenon of quickly decreasing development speed is also known as *software entropy*. Entropy (the amount of disorder in a system) is a mathematical and scientific concept that can also apply to software systems. (If you're interested in the math and science of entropy, look up the second law of thermodynamics.)

In software, entropy manifests in the form of code that tends to deteriorate. Each time you change something in a code base, the amount of disorder in it, or entropy, increases. If left without proper care, such as constant cleaning and refactoring, the system becomes increasingly complex and disorganized. Fixing one bug introduces more bugs, and modifying one part of the software breaks several others—it's like a

domino effect. Eventually, the code base becomes unreliable. And worst of all, it's hard to bring it back to stability.

Tests help overturn this tendency. They act as a safety net—a tool that provides insurance against a vast majority of regressions. Tests help make sure the existing functionality works, even after you introduce new features or refactor the code to better fit new requirements.

**DEFINITION** A *regression* is when a feature stops working as intended after a certain event (usually, a code modification). The terms *regression* and *software bug* are synonyms and can be used interchangeably.

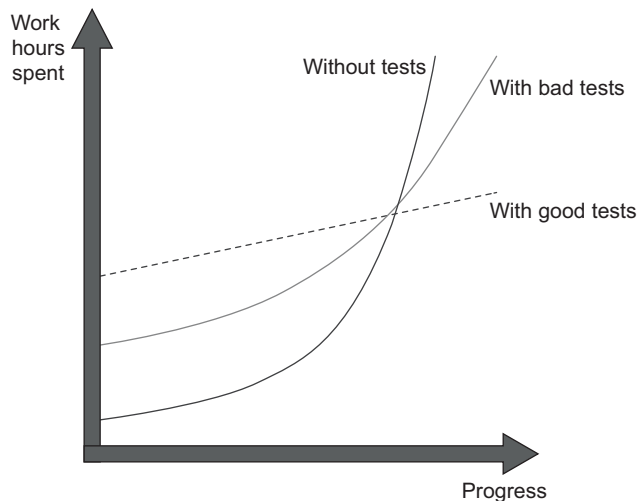
The downside here is that tests require initial—sometimes significant—effort. But they pay for themselves in the long run by helping the project to grow in the later stages. Software development without the help of tests that constantly verify the code base simply doesn't scale.

Sustainability and scalability are the keys. They allow you to maintain development speed in the long run.

### 1.2.1 What makes a good or bad test?

Although unit testing helps maintain project growth, it's not enough to just write tests. Badly written tests still result in the same picture.

As shown in figure 1.2, bad tests do help to slow down code deterioration at the beginning: the decline in development speed is less prominent compared to the situation with no tests at all. But nothing really changes in the grand scheme of things. It might take longer for such a project to enter the stagnation phase, but stagnation is still inevitable.



**Figure 1.2** The difference in growth dynamics between projects with good and bad tests. A project with badly written tests exhibits the properties of a project with good tests at the beginning, but it eventually falls into the stagnation phase.

Remember, *not all tests are created equal*. Some of them are valuable and contribute a lot to overall software quality. Others don't. They raise false alarms, don't help you catch regression errors, and are slow and difficult to maintain. It's easy to fall into the trap of writing unit tests for the sake of unit testing without a clear picture of whether it helps the project.

You can't achieve the goal of unit testing by just throwing more tests at the project. You need to consider both the test's value and its upkeep cost. The cost component is determined by the amount of time spent on various activities:

- Refactoring the test when you refactor the underlying code
- Running the test on each code change
- Dealing with false alarms raised by the test
- Spending time reading the test when you're trying to understand how the underlying code behaves

It's easy to create tests whose net value is close to zero or even is negative due to high maintenance costs. To enable sustainable project growth, you have to exclusively focus on high-quality tests—those are the only type of tests that are worth keeping in the test suite.

#### **Production code vs. test code**

People often think production code and test code are different. Tests are assumed to be an addition to production code and have no cost of ownership. By extension, people often believe that the more tests, the better. This isn't the case. *Code is a liability, not an asset*. The more code you introduce, the more you extend the surface area for potential bugs in your software, and the higher the project's upkeep cost. It's always better to solve problems with as little code as possible.

*Tests are code, too*. You should view them as the part of your code base that aims at solving a particular problem: ensuring the application's correctness. Unit tests, just like any other code, are also vulnerable to bugs and require maintenance.

It's crucial to learn how to differentiate between good and bad unit tests. I cover this topic in chapter 4.

### **1.3 Using coverage metrics to measure test suite quality**

In this section, I talk about the two most popular coverage metrics—code coverage and branch coverage—how to calculate them, how they're used, and problems with them. I'll show why it's detrimental for programmers to aim at a particular coverage number and why you can't just rely on coverage metrics to determine the quality of your test suite.

**DEFINITION** A *coverage metric* shows how much source code a test suite executes, from none to 100%.

There are different types of coverage metrics, and they're often used to assess the quality of a test suite. The common belief is that the higher the coverage number, the better.

Unfortunately, it's not that simple, and coverage metrics, while providing valuable feedback, can't be used to effectively measure the *quality* of a test suite. It's the same situation as with the ability to unit test the code: coverage metrics are a good negative indicator but a bad positive one.

If a metric shows that there's too little coverage in your code base—say, only 10%—that's a good indication that you are not testing enough. But the reverse isn't true: even 100% coverage isn't a guarantee that you have a good-quality test suite. A test suite that provides high coverage can still be of poor quality.

I already touched on why this is so—you can't just throw random tests at your project with the hope those tests will improve the situation. But let's discuss this problem in detail with respect to the code coverage metric.

### 1.3.1 Understanding the code coverage metric

The first and most-used coverage metric is *code coverage*, also known as *test coverage*; see figure 1.3. This metric shows the ratio of the number of code lines executed by at least one test and the total number of lines in the production code base.

$$\text{Code coverage (test coverage)} = \frac{\text{Lines of code executed}}{\text{Total number of lines}}$$

**Figure 1.3** The code coverage (test coverage) metric is calculated as the ratio between the number of code lines executed by the test suite and the total number of lines in the production code base.

Let's see an example to better understand how this works. Listing 1.1 shows an `IsStringLong` method and a test that covers it. The method determines whether a string provided to it as an input parameter is long (here, the definition of *long* is any string with the length greater than five characters). The test exercises the method using "abc" and checks that this string is not considered long.

**Listing 1.1** A sample method partially covered by a test

```
public static bool IsStringLong(string input)
{
    if (input.Length > 5)
        return true;
    return false;
}
```

Not covered by the test

Covered by the test

```
public void Test()
{
    bool result = IsStringLong("abc");
    Assert.Equal(false, result);
}
```

It's easy to calculate the code coverage here. The total number of lines in the method is five (curly braces count, too). The number of lines executed by the test is four—the test goes through all the code lines except for the `return true;` statement. This gives us  $4/5 = 0.8 = 80\%$  code coverage.

Now, what if I refactor the method and inline the unnecessary `if` statement, like this?

```
public static bool IsStringLong(string input)
{
    return input.Length > 5;
}

public void Test()
{
    bool result = IsStringLong("abc");
    Assert.Equal(false, result);
}
```

Does the code coverage number change? Yes, it does. Because the test now exercises all three lines of code (the `return` statement plus two curly braces), the code coverage increases to 100%.

But did I improve the test suite with this refactoring? Of course not. I just shuffled the code inside the method. The test still verifies the same number of possible outcomes.

This simple example shows how easy it is to game the coverage numbers. The more compact your code is, the better the test coverage metric becomes, because it only accounts for the raw line numbers. At the same time, squashing more code into less space doesn't (and shouldn't) change the value of the test suite or the maintainability of the underlying code base.

### 1.3.2 *Understanding the branch coverage metric*

Another coverage metric is called *branch coverage*. Branch coverage provides more precise results than code coverage because it helps cope with code coverage's shortcomings. Instead of using the raw number of code lines, this metric focuses on control structures, such as `if` and `switch` statements. It shows how many of such control structures are traversed by at least one test in the suite, as shown in figure 1.4.

$$\text{Branch coverage} = \frac{\text{Branches traversed}}{\text{Total number of branches}}$$

**Figure 1.4** The branch metric is calculated as the ratio of the number of code branches exercised by the test suite and the total number of branches in the production code base.



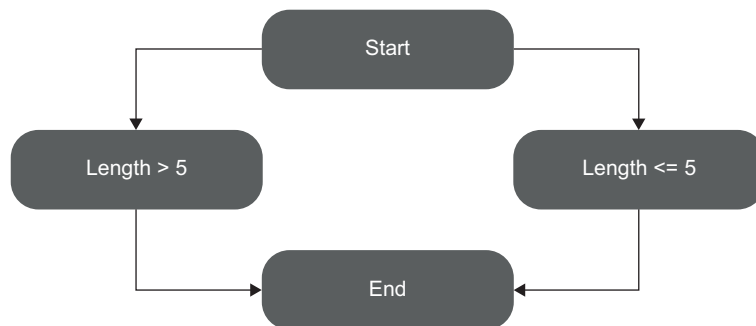
To calculate the branch coverage metric, you need to sum up all possible branches in your code base and see how many of them are visited by tests. Let's take our previous example again:

```
public static bool IsStringLong(string input)
{
    return input.Length > 5;
}

public void Test()
{
    bool result = IsStringLong("abc");
    Assert.Equal(false, result);
}
```

There are two branches in the `IsStringLong` method: one for the situation when the length of the string argument is greater than five characters, and the other one when it's not. The test covers only one of these branches, so the branch coverage metric is  $1/2 = 0.5 = 50\%$ . And it doesn't matter how we represent the code under test—whether we use an `if` statement as before or use the shorter notation. The branch coverage metric only accounts for the number of branches; it doesn't take into consideration how many lines of code it took to implement those branches.

Figure 1.5 shows a helpful way to visualize this metric. You can represent all possible paths the code under test can take as a graph and see how many of them have been traversed. `IsStringLong` has two such paths, and the test exercises only one of them.



**Figure 1.5** The method `IsStringLong` represented as a graph of possible code paths. Test covers only one of the two code paths, thus providing 50% branch coverage.

### 1.3.3 Problems with coverage metrics

Although the branch coverage metric yields better results than code coverage, you still can't rely on either of them to determine the quality of your test suite, for two reasons:

- You can't guarantee that the test verifies all the possible outcomes of the system under test.
- No coverage metric can take into account code paths in external libraries.

Let's look more closely at each of these reasons.

#### YOU CAN'T GUARANTEE THAT THE TEST VERIFIES ALL THE POSSIBLE OUTCOMES

For the code paths to be actually tested and not just exercised, your unit tests must have appropriate assertions. In other words, you need to check that the outcome the system under test produces is the exact outcome you expect it to produce. Moreover, this outcome may have several components; and for the coverage metrics to be meaningful, you need to verify all of them.

The next listing shows another version of the `IsStringLong` method. It records the last result into a public `WasLastStringLong` property.

**Listing 1.2** Version of `IsStringLong` that records the last result

```
public static bool WasLastStringLong { get; private set; }

public static bool IsStringLong(string input)
{
    bool result = input.Length > 5;
    WasLastStringLong = result;
    return result;
}

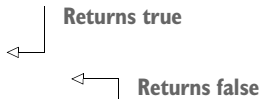
public void Test()
{
    bool result = IsStringLong("abc");
    Assert.Equal(false, result);
}
```

The `IsStringLong` method now has two outcomes: an explicit one, which is encoded by the return value; and an implicit one, which is the new value of the property. And in spite of not verifying the second, implicit outcome, the coverage metrics would still show the same results: 100% for the code coverage and 50% for the branch coverage. As you can see, the coverage metrics don't guarantee that the underlying code is tested, only that it has been executed at some point.

An extreme version of this situation with partially tested outcomes is *assertion-free testing*, which is when you write tests that don't have any assertion statements in them whatsoever. Here's an example of assertion-free testing.

**Listing 1.3 A test with no assertions always passes.**

```
public void Test()  
{  
    bool result1 = IsStringLong("abc");  
    bool result2 = IsStringLong("abcdef");  
}
```



This test has both code and branch coverage metrics showing 100%. But at the same time, it is completely useless because it doesn't verify anything.

**A story from the trenches**

The concept of assertion-free testing might look like a dumb idea, but it does happen in the wild.

Years ago, I worked on a project where management imposed a strict requirement of having 100% code coverage for every project under development. This initiative had noble intentions. It was during the time when unit testing wasn't as prevalent as it is today. Few people in the organization practiced it, and even fewer did unit testing consistently.

A group of developers had gone to a conference where many talks were devoted to unit testing. After returning, they decided to put their new knowledge into practice. Upper management supported them, and the great conversion to better programming techniques began. Internal presentations were given. New tools were installed. And, more importantly, a new company-wide rule was imposed: all development teams had to focus on writing tests exclusively until they reached the 100% code coverage mark. After they reached this goal, any code check-in that lowered the metric had to be rejected by the build systems.

As you might guess, this didn't play out well. Crushed by this severe limitation, developers started to seek ways to game the system. Naturally, many of them came to the same realization: if you wrap all tests with `try/catch` blocks and don't introduce any assertions in them, those tests are guaranteed to pass. People started to mindlessly create tests for the sake of meeting the mandatory 100% coverage requirement. Needless to say, those tests didn't add any value to the projects. Moreover, they damaged the projects because of all the effort and time they steered away from productive activities, and because of the upkeep costs required to maintain the tests moving forward.

Eventually, the requirement was lowered to 90% and then to 80%; after some period of time, it was retracted altogether (for the better!).

But let's say that you thoroughly verify each outcome of the code under test. Does this, in combination with the branch coverage metric, provide a reliable mechanism, which you can use to determine the quality of your test suite? Unfortunately, no.

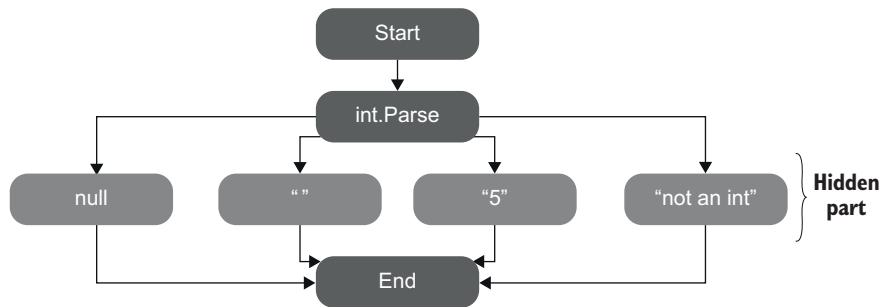
**NO COVERAGE METRIC CAN TAKE INTO ACCOUNT CODE PATHS IN EXTERNAL LIBRARIES**

The second problem with all coverage metrics is that they don't take into account code paths that external libraries go through when the system under test calls methods on them. Let's take the following example:

```
public static int Parse(string input)
{
    return int.Parse(input);
}

public void Test()
{
    int result = Parse("5");
    Assert.Equal(5, result);
}
```

The branch coverage metric shows 100%, and the test verifies all components of the method's outcome. It has a single such component anyway—the return value. At the same time, this test is nowhere near being exhaustive. It doesn't take into account the code paths the .NET Framework's `int.Parse` method may go through. And there are quite a number of code paths, even in this simple method, as you can see in figure 1.6.



**Figure 1.6** Hidden code paths of external libraries. Coverage metrics have no way to see how many of them there are and how many of them your tests exercise.

The built-in integer type has plenty of branches that are hidden from the test and that might lead to different results, should you change the method's input parameter. Here are just a few possible arguments that can't be transformed into an integer:

- Null value
- An empty string
- "Not an int"
- A string that's too large

You can fall into numerous edge cases, and there's no way to see if your tests account for all of them.

This is not to say that coverage metrics *should* take into account code paths in external libraries (they shouldn't), but rather to show you that you can't rely on those metrics to see how good or bad your unit tests are. Coverage metrics can't possibly tell whether your tests are exhaustive; nor can they say if you have enough tests.

### 1.3.4 Aiming at a particular coverage number

At this point, I hope you can see that relying on coverage metrics to determine the quality of your test suite is not enough. It can also lead to dangerous territory if you start making a specific coverage number a target, be it 100%, 90%, or even a moderate 70%. The best way to view a coverage metric is as an indicator, not a goal in and of itself.

Think of a patient in a hospital. Their high temperature might indicate a fever and is a helpful observation. But the hospital shouldn't make the proper temperature of this patient a goal to target by any means necessary. Otherwise, the hospital might end up with the quick and "efficient" solution of installing an air conditioner next to the patient and regulating their temperature by adjusting the amount of cold air flowing onto their skin. Of course, this approach doesn't make any sense.

Likewise, targeting a specific coverage number creates a perverse incentive that goes against the goal of unit testing. Instead of focusing on testing the things that matter, people start to seek ways to attain this artificial target. Proper unit testing is difficult enough already. Imposing a mandatory coverage number only distracts developers from being mindful about what they test, and makes proper unit testing even harder to achieve.

**TIP** It's good to have a high level of coverage in core parts of your system. It's bad to make this high level a requirement. The difference is subtle but critical.

Let me repeat myself: coverage metrics are a good negative indicator, but a bad positive one. Low coverage numbers—say, below 60%—are a certain sign of trouble. They mean there's a lot of untested code in your code base. But high numbers don't mean anything. Thus, measuring the code coverage should be only a first step on the way to a quality test suite.

## 1.4 What makes a successful test suite?

I've spent most of this chapter discussing improper ways to measure the quality of a test suite: using coverage metrics. What about a proper way? How should you measure your test suite's quality? The only reliable way is to evaluate each test in the suite individually, one by one. Of course, you don't have to evaluate all of them at

once; that could be quite a large undertaking and require significant upfront effort. You can perform this evaluation gradually. The point is that there's no automated way to see how good your test suite is. You have to apply your personal judgment.

Let's look at a broader picture of what makes a test suite successful as a whole. (We'll dive into the specifics of differentiating between good and bad tests in chapter 4.) A successful test suite has the following properties:

- It's integrated into the development cycle.
- It targets only the most important parts of your code base.
- It provides maximum value with minimum maintenance costs.

#### **1.4.1** *It's integrated into the development cycle*

The only point in having automated tests is if you constantly use them. All tests should be integrated into the development cycle. Ideally, you should execute them on every code change, even the smallest one.

#### **1.4.2** *It targets only the most important parts of your code base*

Just as all tests are not created equal, not all parts of your code base are worth the same attention in terms of unit testing. The value the tests provide is not only in how those tests themselves are structured, but also in the code they verify.

It's important to direct your unit testing efforts to the most critical parts of the system and verify the others only briefly or indirectly. In most applications, the most important part is the part that contains business logic—the *domain model*.<sup>1</sup> Testing business logic gives you the best return on your time investment.

All other parts can be divided into three categories:

- Infrastructure code
- External services and dependencies, such as the database and third-party systems
- Code that glues everything together

Some of these other parts may still need thorough unit testing, though. For example, the infrastructure code may contain complex and important algorithms, so it would make sense to cover them with a lot of tests, too. But in general, most of your attention should be spent on the domain model.

Some of your tests, such as integration tests, can go beyond the domain model and verify how the system works as a whole, including the noncritical parts of the code base. And that's fine. But the focus should remain on the domain model.

Note that in order to follow this guideline, you should isolate the domain model from the non-essential parts of the code base. You have to keep the domain model separated from all other application concerns so you can focus your unit testing

---

<sup>1</sup> See *Domain-Driven Design: Tackling Complexity in the Heart of Software* by Eric Evans (Addison-Wesley, 2003).

efforts on that domain model exclusively. We talk about all this in detail in part 2 of the book.

### 1.4.3 *It provides maximum value with minimum maintenance costs*

The most difficult part of unit testing is achieving maximum value with minimum maintenance costs. That's the main focus of this book.

It's not enough to incorporate tests into a build system, and it's not enough to maintain high test coverage of the domain model. It's also crucial to keep in the suite only the tests whose value exceeds their upkeep costs by a good margin.

This last attribute can be divided in two:

- Recognizing a valuable test (and, by extension, a test of low value)
- Writing a valuable test

Although these skills may seem similar, they're different by nature. To recognize a test of high value, you need a frame of reference. On the other hand, *writing* a valuable test requires you to also know code design techniques. Unit tests and the underlying code are highly intertwined, and it's impossible to create valuable tests without putting significant effort into the code base they cover.

You can view it as the difference between recognizing a good song and being able to compose one. The amount of effort required to become a composer is asymmetrically larger than the effort required to differentiate between good and bad music. The same is true for unit tests. Writing a new test requires more effort than examining an existing one, mostly because you don't write tests in a vacuum: you have to take into account the underlying code. And so although I focus on unit tests, I also devote a significant portion of this book to discussing code design.

## 1.5 *What you will learn in this book*

This book teaches a frame of reference that you can use to analyze any test in your test suite. This frame of reference is foundational. After learning it, you'll be able to look at many of your tests in a new light and see which of them contribute to the project and which must be refactored or gotten rid of altogether.

After setting this stage (chapter 4), the book analyzes the existing unit testing techniques and practices (chapters 4–6, and part of 7). It doesn't matter whether you're familiar with those techniques and practices. If you are familiar with them, you'll see them from a new angle. Most likely, you already *get* them at the intuitive level. This book can help you articulate *why* the techniques and best practices you've been using all along are so helpful.

Don't underestimate this skill. The ability to clearly communicate your ideas to colleagues is priceless. A software developer—even a great one—rarely gets full credit for a design decision if they can't explain why, exactly, that decision was made. This book can help you transform your knowledge from the realm of the unconscious to something you are able to talk about with anyone.

If you don't have much experience with unit testing techniques and best practices, you'll learn a lot. In addition to the frame of reference that you can use to analyze any test in a test suite, the book teaches

- How to refactor the test suite along with the production code it covers
- How to apply different styles of unit testing
- Using integration tests to verify the behavior of the system as a whole
- Identifying and avoiding anti-patterns in unit tests

In addition to unit tests, this book covers the entire topic of automated testing, so you'll also learn about integration and end-to-end tests.

I use C# and .NET in my code samples, but you don't have to be a C# professional to read this book; C# is just the language that I happen to work with the most. All the concepts I talk about are non-language-specific and can be applied to any other object-oriented language, such as Java or C++.

## Summary

- Code tends to deteriorate. Each time you change something in a code base, the amount of disorder in it, or entropy, increases. Without proper care, such as constant cleaning and refactoring, the system becomes increasingly complex and disorganized. Tests help overturn this tendency. They act as a safety net—a tool that provides insurance against the vast majority of regressions.
- It's important to write unit tests. It's equally important to write *good* unit tests. The end result for projects with bad tests or no tests is the same: either stagnation or a lot of regressions with every new release.
- The goal of unit testing is to enable sustainable growth of the software project. A good unit test suite helps avoid the stagnation phase and maintain the development pace over time. With such a suite, you're confident that your changes won't lead to regressions. This, in turn, makes it easier to refactor the code or add new features.
- All tests are not created equal. Each test has a cost and a benefit component, and you need to carefully weigh one against the other. Keep only tests of positive net value in the suite, and get rid of all others. Both the application code and the test code are liabilities, not assets.
- The ability to unit test code is a good litmus test, but it only works in one direction. It's a good negative indicator (if you can't unit test the code, it's of poor quality) but a bad positive one (the ability to unit test the code doesn't guarantee its quality).
- Likewise, coverage metrics are a good negative indicator but a bad positive one. Low coverage numbers are a certain sign of trouble, but a high coverage number doesn't automatically mean your test suite is of high quality.
- Branch coverage provides better insight into the completeness of the test suite but still can't indicate whether the suite is good enough. It doesn't take into



account the presence of assertions, and it can't account for code paths in third-party libraries that your code base uses.

- Imposing a particular coverage number creates a perverse incentive. It's good to have a high level of coverage in core parts of your system, but it's bad to make this high level a requirement.
- A successful test suite exhibits the following attributes:
  - It is integrated into the development cycle.
  - It targets only the most important parts of your code base.
  - It provides maximum value with minimum maintenance costs.
- The only way to achieve the goal of unit testing (that is, enabling sustainable project growth) is to
  - Learn how to differentiate between a good and a bad test.
  - Be able to refactor a test to make it more valuable.