

# Todo list

■ <b>Important:</b> Lay abstract. . . . .	iii
■ <b>Important:</b> Abstract. . . . .	iv
■ <b>Important:</b> Acknowledgements. . . . .	v
■ <b>Important:</b> Replace reading notes. . . . .	xii
■ <b>Important:</b> Declaration of Academic Achievement. . . . .	xiii
■ <i>Easy:</i> Such as this one, but check out Section 2.5 for more options. . . .	3
■ <b>Important:</b> “Important” notes. . . . .	6
■ Generic inlined notes. . . . .	6
■ <i>Later:</i> TODO notes for later! For finishing touches, etc. . . . .	6
■ <i>Easy:</i> Easier notes. . . . .	6
■ <i>Needs time:</i> Tedious notes. . . . .	6
■ <b>Q #1:</b> Questions I might have? . . . . .	7
■ investigate more: Steele 1990? . . . . .	8
■ add reference requirement and test case . . . . .	18
■ add relevant code . . . . .	18
■ cite Dr. Smith . . . . .	18
■ add refs to ‘underlying Theory’ comment and ‘not all outputs be IMs’ comment . . . . .	18
■ add constraints . . . . .	18

## THE GENERATION OF TEST CASES IN DRASIL

# THE GENERATION OF TEST CASES IN DRASIL

By SAMUEL CRAWFORD, B.Eng.

A THESIS  
SUBMITTED TO THE DEPARTMENT OF COMPUTING AND SOFTWARE  
AND THE SCHOOL OF GRADUATE STUDIES  
OF MCMASTER UNIVERSITY  
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF  
MASTER OF APPLIED SCIENCE

Master of Applied Science (2023)  
(Department of Computing and Software)

McMaster University  
Hamilton, Ontario

TITLE: The Generation of Test Cases in Drasil  
AUTHOR: Samuel Crawford, B.Eng.  
SUPERVISOR: Dr. Carette and Dr. Smith  
PAGES: **xiii, 20**

# Lay Abstract

**Important:** Lay abstract.

# Abstract

**Important:** Abstract.

# Acknowledgements

**Important:** Acknowledgements.

# Contents

<b>Todo list</b>	<b>i</b>
<b>Lay Abstract</b>	<b>iii</b>
<b>Abstract</b>	<b>iv</b>
<b>Acknowledgements</b>	<b>v</b>
<b>Contents</b>	<b>vi</b>
<b>List of Figures</b>	<b>viii</b>
<b>List of Tables</b>	<b>ix</b>
<b>List of Source Codes</b>	<b>x</b>
<b>List of Abbreviations and Symbols</b>	<b>xi</b>
<b>Reading Notes</b>	<b>xii</b>
<b>Declaration of Academic Achievement</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Template Organization . . . . .	1
1.2 Writing Tips . . . . .	2
1.3 Development Recommendations . . . . .	3
1.4 Troubleshooting . . . . .	4
<b>2 Extras</b>	<b>5</b>
2.1 Writing Directives . . . . .	5
2.2 Acronyms . . . . .	5
2.3 HREFs . . . . .	5
2.4 Code Snippets . . . . .	6
2.5 TODOs . . . . .	6
<b>3 Notes</b>	<b>8</b>
3.1 A Survey of Metaprogramming Languages . . . . .	8
3.1.1 Definitions . . . . .	8



3.1.2	Metaprogramming Models . . . . .	9
3.2	Software Metrics . . . . .	10
3.3	Software Testing . . . . .	10
3.3.1	General Testing Notes . . . . .	10
3.3.2	Types of Testing . . . . .	10
<b>4</b>	<b>Development Process</b>	<b>16</b>
4.1	Improvements to Manual Test Code . . . . .	17
4.1.1	Testing with Mocks . . . . .	17
4.2	The Use of Assertions in Code . . . . .	18
4.3	Generating Requirements . . . . .	18
	<b>Bibliography</b>	<b>19</b>
	<b>Appendix</b>	<b>20</b>

# List of Figures

# List of Tables

1.1 Template Organization . . . . .	1
-------------------------------------	---

# List of Source Codes

2.1	Pseudocode: exWD . . . . .	5
2.2	Pseudocode: exPHref . . . . .	6
A.1	“MultiDefinitions” (MultiDefn) Definition . . . . .	20
A.2	Pseudocode: Broken QuantityDict Chunk Retriever . . . . .	20

# List of Abbreviations and Symbols

<b>AST</b>	Abstract Syntax Tree
<b>HREF</b>	Hypertext REference
<b>HTML</b>	HyperText Markup Language
<b>IDE</b>	Integrated Development Environment
<b>PDF</b>	Portable Document Format
<b>SST</b>	Skeleton Syntax Tree

# Reading Notes

Before reading this thesis, I encourage you to read through these notes, keeping them in mind while reading.

- The source code of this thesis is [publicly available](#).
- This thesis template is primarily intended for usage by the computer science community<sup>1</sup>. However, anyone is free to use it.
- I’ve tried my best to make this template conform to the thesis requirements as per [those set forth in 2021 by McMaster University](#). However, you should double-check that your usage of this template is compliant with whatever the “current” rules are.

**Important:** Replace reading notes.

---

<sup>1</sup>Hence why there are some  $\text{\LaTeX}$  macros for “code” snippets.

# Declaration of Academic Achievement

**Important:** Declaration of Academic Achievement.

# Chapter 1

## Introduction

Congratulations! If you're seeing this, it means you've managed to compile the PDF, which also means you can get started on typesetting your thesis<sup>1</sup>.

This template is adapted from my [thesis](#). If you'd like to see an example of this template in practice, please feel free to use my thesis as an example.

### 1.1 Template Organization

I've broken up the template according to my preferred organization: chapters in separate files, various kinds of assets (images, tables, code snippets, macros, etc.) in separate files, etc. The split is approximately according to [Table 1.1](#).

Table 1.1: Template Organization

File/Folder	Intended Usage & Description
<code>thesis.tex</code>	Focal L <sup>A</sup> T <sub>E</sub> X file that collects everything and is used to build your thesis/report document.
<code>Makefile</code>	A basic <code>Makefile</code> configuration. See <code>make help</code> for a list of helpful commands.
<code>build/</code>	When you build your PDF, this folder is used as the working directory of LuaLaTeX. Using this allows us to quickly get rid of L <sup>A</sup> T <sub>E</sub> X build files that can cause problems when we re-build documents.
<code>manifest.tex</code>	Basic options that you should certainly configure according to your needs.
<code>chapters.tex</code>	All chapters of your thesis should be included here.

---

<sup>1</sup>Or report or ...



<code>chapters/</code>	Enumeration of the chapters of your thesis. I prefer using a two-digit indexing pattern for the prefix of file names so that I can quickly open up by chapter number using VS Codium.
<code>assets.tex</code>	Enumeration of the various kinds of “assets” in the <code>assets/</code> folder. See the file for examples on how you can write your extra utility macros.
<code>assets/</code>	Enumeration of various kinds of “assets,” with subdirectories for images and figures, tables, and code snippets.
<code>front.tex</code>	All front matter of your thesis should be included here.
<code>front/</code>	Enumeration of the front chapters of your thesis. These chapters should all be numbered using Roman numerals.
<code>back.tex</code>	All back matter of your thesis should be included here.
<code>back/</code>	Enumeration of the back matter content.
<code>acronyms.tex</code>	List of acronyms you intend to use in your thesis. This uses the “acro” $\LaTeX$ package.
<code>macros.tex</code>	Helpful macros!
<code>unicode_chars.tex</code>	At times, you might find issues with unicode characters, especially in verbatim environments, where you might need to manually define them using other font glyphs.
<code>mcmaster_colours.tex</code>	Macros for the McMaster colour palette.
<code>README.md</code>	Read it!
<code>.gitignore</code>	List of files in the working directory that should be ignored by git.
<code>latexmkrc</code>	Used for setting the timezone for latexmk, but can be used for other options.

## 1.2 Writing Tips

When drafting chapters, I:

1. wrote “writing directives” for each chapter to understand what I need to write about (see [Section 2.1](#)),
2. wrote “todo” notes for tedious things that I might want to do later (such as citations, figures, code snippets, etc., see [Section 2.5](#)), and

3. regularly built my thesis using `make debug` to make sure that whatever I wrote didn't break the  $\text{\LaTeX}$  code.

For workflow recommendations, you should speak with your supervisor as they might prefer you work in a specific way with them.

## 1.3 Development Recommendations

Other than the basic tools I used for this template, I enjoyed using the following tools while writing my thesis:

1. **VS Codium**/**VS Code**<sup>2</sup> with the following extensions:
  - (a)  **$\text{\LaTeX}$  Workshop**, for  $\text{\LaTeX}$  syntax highlighting, code formatting (this is highly recommended), and code completion,
  - (b) **LT $\text{\TeX}$  - LanguageTool grammar/spell checking**, for grammar checking using **LanguageTool**, and
  - (c) **Todo Tree**, for quickly listing all of my TODO notes in my IDE (in addition to the list at the top of the PDF).
2. **texcount** (which should come with your  $\text{\LaTeX}$  installation) to quickly check the word count of individual  $\text{\LaTeX}$  files, and
3. **Zotero** for collecting my references and quickly exporting bib entries that I could use.

In particular, when writing, I found it particularly helpful to use VS Code's "Zen Mode" (to see your keybind, press **CTRL+ALT+P** and search for "Zen"), which enters a stripped-down full-screen version of the current working file, keeping your eyes purely focused on the document in front of you. Being comfortable with the keybinds is particularly helpful for working effectively in this setup. For example, I found the following<sup>3</sup> to be helpful: **CTRL+TAB** and **CTRL+SHIFT+TAB** to scroll between open files, **CTRL+P** to quickly open up recent files, **CTRL+ALT+P** to run commands you forgot the keybind for, **CTRL+O** to open up files out of the current working directory.

While writing, I enjoyed:

1. using "TODO" notes

*Easy:* Such as this one, but check out **Section 2.5** for more options.

to collect notes that I would want to do later,

2. formatting the  $\text{\LaTeX}$  code to make it easier to read (the  $\text{\LaTeX}$  Workshop plugin has functionality for this),

---

<sup>2</sup>I prefer VS Codium simply because I prefer libre software.

<sup>3</sup>If you're not using Linux, I cannot guarantee that these will be the same for you, so you should use **CTRL+ALT+P** to look for your appropriate bound keybinds.

3. breaking the non-textual content into separate files and “include”-ing them in the  $\text{\LaTeX}$  code so that they didn’t cause large visual interruptions,
4. using git to version control copies of my thesis, chapters, etc.,
5. using [TikZ](#) and [draw.io/diagrams.net](#) to build graphics and diagrams, and
6. building the thesis often using `make debug` to quickly debug issues in the written code.

## 1.4 Troubleshooting

“StackOverflow” is a great area to look for solutions to common  $\text{\LaTeX}$  issues. Otherwise, feel free to use create a ticket or sending an email to me.

# Chapter 2

## Extras

### Writing Directives

- What macros do I want the reader to know about?

## 2.1 Writing Directives

I enjoy writing directives (mostly questions) to navigate what I should be writing about in each chapter. You can do this using:

Source Code 2.1: Pseudocode: exWD

---

```
\begin{writingdirectives}
  \item What macros do I want the reader to know about?
\end{writingdirectives}
```

---

Personally, I put them at the top of chapter files, just after chapter declarations.

## 2.2 Acronyms

I used a lot of acronyms in my thesis, and I wanted to add a glossary to the front matter. This is compulsory for McMaster theses, but you may remove it if you don't need it. I used the `acro` package, feel free to read their [documentation](#). For example, I might want to write HyperText Markup Language (HTML) (written using `\acf{html}`) or just HTML (written using `\acs{html}`). To define them, you can find examples in the `acronyms.tex` file.

## 2.3 HREFs

For PDFs, we have (at least) 2 ways of viewing them: on our computers, and printed out on paper. If you choose to view through your computer, reading links

(as they are linked in this example, inlined everywhere with “clickable” links) is fine. However, if you choose to read it on printed paper, you will find trouble clicking on those same links. To mitigate this issue, I built the “porthref” macro (see `macros.tex` for the definition) to build links that appear as clickable text when “compiling for computer-focused reading,” and adds links to footnotes when “compiling for printing-focused reading.” There is an option (`compilingforprinting`) in the `manifest.tex` file that controls whether PDF builds should be done for computers or for printers. For example, by default, **McMaster** is made with clickable functionality, but if you change the `manifest.tex` option as mentioned, then you will see the link in a footnote (try it out!).

Source Code 2.2: Pseudocode: exPHref

---

```
\porthref{McMaster}{https://www.mcmaster.ca/}
```

---

## 2.4 Code Snippets

Since I did my Master’s in computer science, I needed code snippet listings. For them, I chose to use the `minted` package (which lets you write colourized code snippets, supporting syntax highlighting for many languages). However, I also wanted source code listings to be linked to the main Drasil [1] repository. For example, to write **Source Code A.1**<sup>1</sup>, I had to write `assets/code/example.tex` and create appropriate links in `assets.tex`. Additionally, for pseudocode, you can also use the pseudocode environment, such as that used in **Source Code A.2** (similarly built).

## 2.5 TODOs

While writing, I plastered my thesis with notes for future work because, for whatever reason, I just didn’t want to, or wasn’t able to, do said work at that time. To help me sort out my notes, I used the `todonotes` package with a few extra macros (defined in `macros.tex`). For example,...

Important notes:

**Important:** “Important” notes.

Generic inlined notes:

Generic inlined notes.

Notes for later:

Some “easy” notes:

*Easy:* Easier notes.

Tedious work:

*Later:* TODO notes for later! For finishing touches, etc.

---

<sup>1</sup>Note: the header links use the portable HREFs (as per **Section 2.3**)!

*Needs time:* Tedious notes.

**Q #1:** Questions I might have?

Questions:

# Chapter 3

## Notes

### 3.1 A Survey of Metaprogramming Languages

investigate more:  
Steele 1990?

- Often done with an Abstract Syntax Tree (AST), although other bases are used:
  - Skeleton Syntax Tree (SST), used by Dylan [2, p. 113:6]
- Allows for improvements in:
  - "performance by generating efficient specialized programs based on specifications instead of using generic but inefficient programs" [2, p. 113:2]
  - reasoning about object programs through "analyzing and discovering object-program characteristics that enable applying further optimizations as well as inspecting and validating the behavior of the object program" [2, p. 113:2]
  - code reuse through capturing "code patterns that cannot be abstracted" [2, p. 113:2]

#### 3.1.1 Definitions

"*Metaprogramming* is the process of writing computer programs, called *metaprograms*, that [can] ... generate new programs or modify existing ones" [2, p. 113:1].

- Metalanguage: "the language in which the metaprogram is written" [2, p. 113:1]
- Object language: "the language in which the generated or transformed program is written" [2, p. 113:1]
- Homogeneous metaprogramming: when "the object language and the metalanguage are the same" [2, p. 113:1]
- Heterogeneous metaprogramming: when "the object language and the metalanguage are ... different" [2, p. 113:1]

### 3.1.2 Metaprogramming Models

#### Macro Systems [2, p. 113:3-7]

- Map specified input sequences in a source file to corresponding output sequences ("macro expansion") until no input sequences remain [2, p. 113:3]; this process can be:
  1. procedural (involving algorithms), or
  2. pattern-based (only using pattern matching) [2, p. 113:4]
- Must avoid variable capture (unintended name conflicts) by being "hygienic" [2, p. 113:4]; this may be overridden to allow for "intentional variable capture", such as Scheme's *syntax-case* macro [2, p. 113:5]

#### Lexical Macros

- Language agnostic [2, p. 113:3]
- Usually only sufficient for basic metaprogramming since changes to the code without considering its meaning "may cause unintended side effects or name clashes and may introduce difficult-to-solve bugs" [2, p. 113:5]
- Marco was the first safe, language-independent macro system that "enforce[s] specific rules that can be checked by special oracles" for given languages (as long as the languages "produce descriptive error messages") [2, p. 113:6]

#### Syntactic Macros

- "Aware of the language syntax and semantics" [2, p. 113:3]
- MS<sup>2</sup> "was the first programmable syntactic macro system for syntactically rich languages", including by using "a type system to ensure that all generated code fragments are syntactically correct" [2, p. 113:5]

#### Reflection Systems [2, p. 113:7-9]

- "Perform computations on [themselves] in the same way as for the target application, enabling one to adjust the system behavior based on the needs of its execution" [2, p. 113:7]
- Requires that the system can examine ("introspection") and modify ("intercession") how it is represented [2, p. 113:7]
  - The representation of a system can either be structural or behavioural (e.g., variable assignment) [2, p. 113:7]
- "Runtime code generation based on source text can be impractical, inefficient, and unsafe, so alternatives have been explored based on ASTs and quasi-quote operators, offering a structured approach that is subject to typing for expressing and combining code at runtime" [2, p. 113:8]



- "Not limited to runtime systems", as some "compile-time systems ...rely on some form of structural introspection to perform code generation" [2, p. 113:9]

## 3.2 Software Metrics

- The following branches of testing started as parts of quality testing:
  - Reliability testing [3, p. 18, ch. 10]
  - Performance testing [3, p. 18, ch. 7]
- Reliability and maintainability can start to be tested even without code by “measur[ing] structural attributes of representations of the software” [3, p. 18]
- The US Software Engineering Institute has a checklist for determining which types of lines of code are included when counting [3, pp. 30-31]
- Measurements should include an entity to be measured, a specific attribute to measure, and the actual measure (i.e., units, starting state, ending state, what to include) [3, p. 36]
  - These attributes must be defined before they can be measured [3, p. 38]

## 3.3 Software Testing

### 3.3.1 General Testing Notes

- Simple, normal test cases (test-to-pass) should always be developed and run before more complicated, unusual test cases (test-to-fail) [4, p. 66]

### 3.3.2 Types of Testing

#### Static Black-Box (Specification) Testing [4, p. 56-62]

Most of this section is irrelevant to generating test cases, as they require human involvement (e.g., Pretend to Be the Customer [4, p. 57-58], Research Existing Standards and Guidelines [4, p. 58-59]). However, it provides a "Specification Terminology Checklist" [4, p. 61] that includes some keywords that, if found, could trigger an applicable warning to the user (similar to the idea behind the correctness/consistency checks project):

- **Potentially unrealistic:** always, every, all, none, every, certainly, therefore, clearly, obviously, evidently
- **Potentially vague:** some, sometimes, often, usually, ordinarily, customarily, most, mostly, good, high-quality, fast, quickly, cheap, inexpensive, efficient, small, stable

- **Potentially incomplete:** etc., and so forth, and so on, such as, handled, processed, rejected, skipped, eliminated, if ... then ... (without "else" or "otherwise")

### Dynamic Black-Box (Behavioural) Testing [4, p. 64-65]

This is the process of "entering inputs, receiving outputs, and checking the results" [4, p. 64]. Note that while black-box testing is usually done at a higher (e.g., system) level, unit testing can also be black-box [5, p. 1].

### Requirements

- Requirements documentation (definition of what the software does) [4, p. 64]; relevant information could be:
  - Requirements: Input-Values and Output-Values
  - Input/output data constraints

### Exploratory Testing [4, p. 65]

An alternative to dynamic black-box testing when a specification is not available [4, p. 65]. The software is explored to determine its features, and these features are then tested [4, p. 65]. Finding any bugs using this method is a positive thing [4, p. 65], since despite not knowing what the software *should* do, you were able to determine that something is wrong.

This is not applicable to Drasil, because not only does it already generate a specification, making this type of testing unnecessary, there is also a lot of human-based trial and error required for this kind of testing [6].

### Equivalence Partitioning/Classing [4, p. 67-69]

The process of dividing the infinite set of test cases into a finite set that is just as effective (i.e., by revealing the same bugs) [4, p. 67].

### Requirements

- Ranges of possible values [4, p. 67]; could be obtained through:
  - Input/output data constraints
  - Case statements

### Data Testing [4, p. 70-79]

The process of "checking that information the user inputs [and] results", both final and intermediate, "are handled correctly" [4, p. 70].

**Boundary Conditions** [4, p. 70-74] "[S]ituations at the edge of the planned operational limits of the software" [4, p. 72]. Often affects types of data (e.g., numeric, speed, character, location, position, size, quantity [4, p. 72]) each with its own set of (e.g., first/last, min/max, start/finish, over/under, empty/full, shortest/longest, slowest/fastest, soonest/latest, largest/smallest, highest/lowest, next-to/farthest-from [4, p. 72-73]). Data at these boundaries should be included in an equivalence partition, but so should data in between them [4, p. 73]. Boundary conditions should be tested using "the valid data just inside the boundary, ... the last possible valid data, and ... the invalid data just outside the boundary" [4, p. 73].

### Requirements

- Ranges of possible values [4, p. 67, 73]; could be obtained through:
  - Case statements
  - Input/output data constraints (e.g., inputs that would lead to a boundary output)

**Buffer Overruns** [4, p. 201-205] *Buffer overruns* are "the number one cause of software security issues" [4, p. 75]. They occur when the size of the destination for some data is smaller than the data itself, causing existing data (including code) to be overwritten and malicious code to potentially be injected [4, p. 202, 204-205]. They often arise from bad programming practices in "languages [sic] such as C and C++, that lack safe string handling functions" [4, p. 201]. Any unsafe versions of these functions that are used should be replaced with the corresponding safe versions [4, p. 203-204].

**Sub-Boundary Conditions** [4, p. 75-77] Boundary conditions "that are internal to the software [but] aren't necessarily apparent to an end user" [4, p. 75]. These include powers of two [4, p. 75-76] and ASCII and Unicode tables [4, p. 76-77].

While this is of interest to the domain of scientific computing, this is too involved for Drasil right now, and the existing software constraints limit much of the potential errors from over/underflow [6]. Additionally, strings are not really used as inputs to Drasil and only occur in output with predefined values, so testing these values are unlikely to be fruitful.

### Requirements

- Increased knowledge of data type structures (e.g., monoids, rings, etc. [6]); this would capture these sub-boundaries, as well as other information like relevant tests cases, along with our notion of these data types (**Space**)

**Default, Empty, Blank, Null, Zero, and None** [4, p. 77-78] These should be their own equivalence class, since "the software usually handles them differently" than "the valid cases or ... invalid cases" [4, p. 78].

Since these values may not always be applicable to a given scenario (e.g., a test case for zero doesn't make sense if there is a constraint that the value in question cannot be zero), the user should likely be able to select categories of tests to generate instead of Drasil just generating all possible test cases based on the inputs [6].

### Requirements

- Knowledge of an "empty" value for each **Space** (stored alongside each type in **Space**?)
- Knowledge of how input data could be omitted from an input (e.g., a missing command line argument, an empty line in a file); could be obtained from:
  - User responsibilities
- Knowledge of how a programming language deals with **Null** values and how these can be passed as arguments

**Invalid, Wrong, Incorrect, and Garbage Data** [4, p. 78-79] This is testing-to-fail [4, p. 77].

**Requirements** This seems to be the most open-ended category of testing.

- Specification of correct inputs that can be ignored; could be obtained through:
  - Input/output data constraints (e.g., inputs that would lead to a violated output constraint)
  - Type information for each input (e.g., passing a string instead of a number)

### State Testing [4, p. 79-87]

The process of testing "a program's states and the transitions between them" [4, p. 79].

**Logic Flow Testing** [4, p. 80-84] This is done by creating a state transition diagram that includes:

- Every possible unique state
- The condition(s) that take(s) the program between states
- The condition(s) and output(s) when a state is entered or exited

to map out the logic flow from the user's perspective [4, p. 81-82]. Next, these states should be partitioned using one (or more) of the following methods:

1. Test each state once
2. Test the most common state transitions
3. Test the least common state transitions
4. Test all error states and error return transitions
5. Test random state transitions [4, p. 82-83]

For all of these tests, the values of the state variables should be verified [4, p. 83].

### Requirements

- Knowledge of the different states of the program [4, p. 82]; could be obtained through:
  - The program's modules and/or functions
  - The program's exceptions
- Knowledge about the different state transitions [4, p. 82]; could be obtained through:
  - Testing the state transitions near the beginning of a workflow more?

**Testing States to Fail [4, p. 84-87]** The goal here is to try and put the program in a fail state by doing things that are out of the ordinary. These include:

- Race Conditions and Bad Timing [4, p. 85-86] (Is this relevant to our examples?)
- Repetition Testing: "doing the same operation over and over", potentially up to "thousands of attempts" [4, p. 86]
- Stress Testing: "running the software under less-than-ideal conditions" [4, p. 86]
- Load testing: running the software with as large of a load as possible (e.g., large inputs, many peripherals) [4, p. 86]

### Requirements

- Repetition Testing: The types of operations that are likely to lead to errors when repeated (e.g., overwriting files?)
- Stress testing: can these be automated with pytest or are they outside our scope?
- Load testing: Knowledge about the types of inputs that could overload the system (e.g., upper bounds on values of certain types)

**Other Black-Box Testing** [4, p. 87-89]

- Act like an inexperienced user (likely cannot be generated by Drasil)
- Look for bugs where they've already been found (keep track of previous failed test cases?)
- Think like a hacker (is this out of scope?)
- Follow experience (this will implicitly be done just by using Drasil)

# Chapter 4

## Development Process

The following is a rough outline of the steps I have gone through this far for this project:

- start development of system tests (this was pushed too later to focus on unit tests)
- test inputting default values as `floats` and `ints`
- check constraints for valid input
- check constraints for invalid input
- test calculation of:
  - `t_flight`
  - `p_land`
  - `d_offset`
  - `s`
- test valid output of writing output
- test for projectile going long
- integrate system tests into existing unit tests
- test for assumption violation of `g`
  - code generation could be flawed, so we can't assume assumptions are respected
  - test cases shouldn't necessarily match what is done by the code; for example, `g = 0` shouldn't really give a `ZeroDivisionError`; it should be a `ValueError`
  - this inspired the potential for [The Use of Assertions in Code](#)
- test that calculations stop on a constraint violation; this is a requirement should be met by the software

- test for empty input file
- start creation of test summary (for `InputParameters` module)
  - it was difficult to judge test case coverage/quality from the code itself
  - this is not really a test plan, as it doesn't capture the testing philosophy
  - rationale for each test explains why it supports coverage and how Drasil derived (would derive) it
- start researching testing
- start the **Generating Requirements** subproject

## 4.1 Improvements to Manual Test Code

Even though this code will eventually be generated by Drasil, it is important that it is still human-readable, for the benefit of those reading the code later. This is one of the goals of Drasil (see [#3417](#) for an example of a similar issue). As such, the following improvements were discovered and implement in the manually created testing code:

- use pytest's parameterization
- reuse functions/data for consistency
- improve import structure
- use `conftest` for running code before all tests of a module

### 4.1.1 Testing with Mocks

When testing code, it is common to first test lower-level modules, then assume that these modules work when testing higher-level modules. An example would be using an input module to set up test cases for a calculation module after testing the input module. This makes sense when writing test cases manually since it reduces the amount of code that needs to be written and still provides a reasonably high assurance in the software; if there is an issue with the input module that affects the calculation module tests, the issue would be revealed when testing the input module.

However, since these test cases will be generated by Drasil, they can be consistently generated with no additional effort. This means that the testing of each module can be done completely independently, increasing the confidence in the tests.



## 4.2 The Use of Assertions in Code

While assertions are often only used when testing, they can also be used in the code itself to enforce constraints or preconditions; they act like documentation that determines behaviour! For example, they could be used to ensure that assumptions about values (like the value for gravitational acceleration) are respected by the code, which gives a higher degree of confidence in the code.

## 4.3 Generating Requirements

I structured my manually created test cases around Projectile’s functional requirements, as these are the most objective aspects of the generated code to test automatically. As such, I created a test case to ensure that if an input constraint was violated, the calculations would stop. However, this test case failed, since the actual implementation of the code did *not* stop upon an input constraint violation. This was because the code choice for what to do on a constraint violation was "disconnected" from the manually written requirement ([#3523](#)). This problem has been encountered before ([#3259](#)) and presented a good opportunity for generation to encourage reusability and consistency. However, since it makes sense to first verify outputs before actually outputting them and inserting generated requirements among manually created ones seemed challenging, it made sense to first generate an output requirement.

While working on Drasil in the summer of 2019, I implemented the generation of an input requirement across most examples ([#1844](#)). I had also attempted to generate an output requirement, but due to time constraints, this was not feasible. The main issue with this change was the desire to capture the source of each output for traceability; this source was attached to the `InstanceModel` (or rarely, `DataDefinition`) and not the underlying `Quantity` that was used for a program’s outputs. The way I had attempted to do this was to add the reference as a `Sentence` in a tuple.

Taking another look at this four years later allowed us to see that we should be storing the outputs of a program as their underlying models, allowing us to keep the source information with it. While there is some discussion about how this might change in the future, for now, all outputs of a program should be `InstanceModels`. Since this change required adding the `Referable` constraints to the output field of `SystemInformation`, the outputs of all examples needed to be updated to satisfy this constraint; this meant that generating the output requirement of each example was nearly trivial once the outputs were specified correctly. After modifying `DataDefinitions` in GlassBR that were outputs to be `InstanceModels` ([#3569](#); [#3583](#)), reorganizing the requirements of SWHS ([#3589](#); [#3607](#)), and clarifying the outputs of SWHS ([#3589](#)), `SglPend` ([#3533](#)), `DblPend` ([#3533](#)), `GamePhysics` ([#3609](#)), and `SSP` ([#3630](#)), the output requirement was ready to be generated.

add reference requirement and test case

add relevant code

cite Dr. Smith

add refs to ‘underlying Theory’ comment and ‘not all outputs be IMs’ comment

add constraints

# Bibliography

- [1] Jacques Carette, Spencer Smith, Jason Balaci, Ting-Yu Wu, Samuel Crawford, Dong Chen, Dan Szymczak, Brooks MacLachlan, Dan Scime, and Maryyam Niazi. *Drasil*. Version v0.1-alpha. 2021-02-09. DOI: [10.5281/ZENODO.4526445](https://doi.org/10.5281/ZENODO.4526445). URL: <https://github.com/JacquesCarette/Drasil/tree/v0.1-alpha> (cit. on p. 6).
- [2] Yannis Lilis and Anthony Savidis. “A Survey of Metaprogramming Languages”. In: *ACM Computing Surveys*. Vol. 52. Association for Computing Machinery, 2019-10, 113:1–40. DOI: <https://doi.org/10.1145/3354584> (cit. on pp. 8–10).
- [3] Norman E. Fenton and Shari Lawrence Pfleeger. *Software Metrics: A Rigorous & Practical Approach*. 2nd ed. Boston, MA, USA: PWS Publishing Company, 1997. ISBN: 0-534-95425-1 (cit. on p. 10).
- [4] Ron Patton. *Software Testing*. 2nd ed. Indianapolis, IN, USA: Sams Publishing, 2006. ISBN: 0-672-32798-8 (cit. on pp. 10–15).
- [5] Pramod Mathew Jacob and M. Prasanna. “A Comparative analysis on Black Box Testing Strategies”. In: *2016 International Conference on Information Science (ICIS)*. Manhattan, NY, USA: IEEE, 2016-08, pp. 1–6. DOI: [10.1109/INFOSCI.2016.7845290](https://doi.org/10.1109/INFOSCI.2016.7845290). URL: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7845290> (visited on 2023-07-05) (cit. on p. 11).
- [6] W. Spencer Smith and Jacques Carette. Private Communication. Hamilton, ON, Canada, 2023-07 (cit. on pp. 11–13).

# Appendix

Source Code A.1: “MultiDefinitions” (MultiDefn) Definition

---

```
-- | 'MultiDefn's are QDefinition factories, used for showing one
  ↳ or more ways
--   we can define a QDefinition.
data MultiDefn e = MultiDefn{
  -- | UID
  _rUid :: UID,
  -- | Underlying quantity it defines.
  _qd :: QuantityDict,
  -- | Explanation of the different ways we can define a quantity.
  _rDesc :: Sentence,
  -- | All possible ways we can define the related quantity.
  _rvs :: NE.NonEmpty (DefiningExpr e)
}
```

---

Source Code A.2: Pseudocode: Broken QuantityDict Chunk Retriever

---

```
retrieveQD :: UID -> ChunkDB -> Maybe QuantityDict
retrieveQD u cdb = do
  (Chunk expectedQd) <- lookup u cdb
  pure expectedQd
```

---