

A Design Brief on Improving Student Productivity by Addressing the Planning Fallacy

October 25, 2020

Contents

1	Background and Purpose	2
2	Stakeholders and Their Objectives	3
2.1	Students	3
2.2	Professors and TAs	3
2.3	The Facilitators of the Solution	3
3	Objectives	3
3.1	Providing an Estimate	3
3.2	Information Acquisition and Privacy	3
3.3	Accessibility and Ease of Use	4
4	Metrics	4
4.1	Measuring Value and Consistency of Estimate	4
4.2	Measuring Effective Information Acquisition and Upholding of Privacy	4
4.3	Measuring the Accessibility and User-Friendliness	4
5	Criteria	4
6	Constraints	6
7	Evaluations of Reference Designs	6
8	Source Extracts	7

1 Background and Purpose

One’s ability to estimate how long it will take them to complete a task is often directly related to the productivity of that person. Unfortunately, it is typical for people, particularly students who make these predictions often, to hold an optimistic mindset when estimating how long it will take them to complete a task, despite having made the same mistake many times [1]. This phenomenon, generally referred to as the Planning Fallacy, is illustrated in Figure 1.

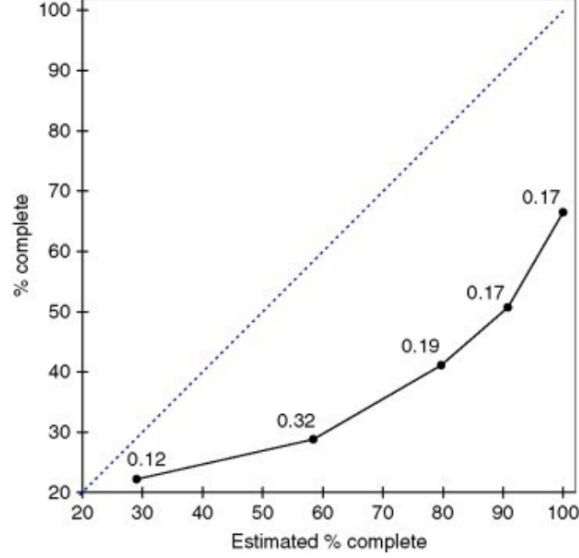


Figure 1: Calibration curve for estimated and actual probability of meeting task completion predictions. Numbers indicate the proportion of responses represented by each point on the curve. Dashed line represents perfect calibration [1].

As shown in Figure 1, a person’s estimate for how long a task will take them is often significantly less than the actual time it takes them to complete a task. Since students are going to be working remotely for the foreseeable future, they will be responsible for accounting for significantly more time than they have in the past. It is therefore important to take this time to optimize work behaviours by diminishing problems like the Planning Fallacy, thereby helping students work more productively and consistently under these conditions.

This design brief seeks to frame an opportunity that could benefit students across the University of Toronto by taking the human factor out of time estimates, thereby giving students the means to overcome the Planning Fallacy. In other words, by implementing a software solution, we can help students work more productively by helping them better understand how much time they should spend on what task to achieve the desired result.

2 Stakeholders and Their Objectives

While students are the main stakeholders in terms of use of the product, several other stakeholders must be considered for a solution to this opportunity. The order of the list of stakeholders and their objectives is not in any particular order.

2.1 Students

As users of the solution, students must be catered to in a solution to this problem. Currently, many students are using nothing better than "dead-reckoning" to judge how long and how well a project will go for them, despite the fact that students tend to be consistently overconfident in these areas. Students in general would therefore benefit from some sort of solution that provides them with guidance for making these decisions.

2.2 Professors and TAs

As the people administering the work students are working on, it is important to keep in mind how they will view and potentially interact with the solution. While the participation or cooperation of course administrators cannot be guaranteed, it is always important to ensure that solutions don't violate their standards for academic honesty and the standards of their course.

2.3 The Facilitators of the Solution

While this group isn't necessary for every conceivable solution, it is likely that there will be a group of people distinct from the greater student body that are responsible for the upkeep and/or distribution of the solution. The function and therefore needs of this group depends on the solution, but should nevertheless be considered.

3 Objectives

By adapting the stakeholder requirements, we can come up with the following objectives to outline the requirements for the solution.

3.1 Providing an Estimate

- The solution must provide an estimate and have some degree of accuracy in the estimate it provides to students - a random guess is not a sufficient solution. In general, this estimate should be better than the estimate the student provides under normal circumstances.
- To produce the estimate, the solution must consider past data from at least the user using the solution, but may include data from multiple or many students.
- The estimate provided should in some way be specific to the user. That is, since it takes different people different amounts of time to do something, the solution should be able to identify patterns in individual behaviours and use that to predict future behaviours.

3.2 Information Acquisition and Privacy

- The solution should require information from the user in order to both provide user specific results and help find results for other users.
- Any personal information about the user must be protected against third party access to prevent breaches of privacy.
- Data integrity must be maintained by periodic backup and data classification.

3.3 Accessibility and Ease of Use

- The solution must be accessible by any student working remotely.
- The solution should be easy for a student to use - the user interface should not be so complex that someone would struggle to navigate and use the solution.

4 Metrics

In order to measure how well a solution meets the objectives listed, we will consider the following metrics, each corresponding to its respective ordinal objective.

4.1 Measuring Value and Consistency of Estimate

The value of the estimate is related to both the absolute accuracy of the estimate and the related confidence with which the estimate is calculated. To measure the absolute accuracy, the formula:

$$\frac{estimate}{actual} \cdot 100\% \quad (1)$$

can be employed, where a value closer to 100% is preferable. The confidence of the measurement provided by the solution should be less than one standard deviation of the raw data. That is to say the uncertainty of the estimate provided should be less than the statistical uncertainty of all data points that have been considered. Note that many algorithms and machine learning libraries require a wide set of learning cases before they can provide a result with a low uncertainty. This metric can be applied after the 'learning period' is expired. The learning period should however not exceed a month of data under any circumstances.

4.2 Measuring Effective Information Acquisition and Upholding of Privacy

The following qualitative metric rubric are dedicated to measure security level of design, modified from control objectives obtained from [Table A.1 CSA ISO/IEC 27001:14] [2]:

4.3 Measuring the Accessibility and User-Friendliness

The following qualitative metric rubric are designed to measure the accessibility as well as the ease of use for a potential solution:

5 Criteria

Criteria for the first quantitative metric can be represented using a graph similar to Figure 1. Instead of estimating the percentage of completion of a certain task, using equation 1, it is possible to graph scattered points on a graph with the estimated time on the x-axis, and actual time on the y-axis, where a calibration line can be drawn as the parity between the two, shown in Figure 2.

Figure 2: Calibration curve for estimated and actual task completion time. Dashed line represents calibration line, indicating a perfect estimate where work progresses linearly.

As a result, when enough experimental data points are plotted, the closer that these data points are to the perfect calibration line, the better accuracy that the approximation model has.

Criteria for qualitative metrics (4.2 and 4.3) can be assessed from the above established rubric in Section 4. The three respective metric standards (Unacceptable, Good, Outstanding) are suggested by the rubric title,

Sub Metrics	Unacceptable	Good	Outstanding	Metrics Source & Reasoning
Information Acquisition	No information is acquired from users to help produce systematic estimations	A minimum of estimated time to completion, actual time to completion and task type are collected and used as parameters.	Additional parameters that help better the estimate are included.	This metrics is used to gauge the variety and utility of the parameters collected and used in the solution.
Personal Information Protection	No explicit policies for security protection are made or defined. No event logs are produced.	A short set of policies for information security are provided, but not reviewed by authentic source. Event logs are only created to record errors. No scheduled monitoring.	A set of policies for information security shall be defined, confirmed against industry standards and maintained. Event logs are created to record user activities and errors, logs are reviewed on a regular basis	This metrics is aimed to evaluate the security of the user's information in the solution, in accordance with security standards from the ISO [2].
Data Integrity Maintenance	Information is stored in accordance to legal requirements. No backup for recorded information.	Information is classified in accordance with legal requirements. The information is secured with respect to its sensitivity from unauthorized disclosure or modification. No backup copies of information are available for data loss incidence.	The information is secured with respect to its sensitivity from unauthorized disclosure or modification. Backup copies of information shall be taken and tested regularly in accordance with an agreed backup policy.	This metrics is aimed to evaluate the maintenance and incidence policy of the solution design, along with the techniques used for secure storage as outlined by the ISO [2].

Sub Metrics	Unacceptable	Good	Outstanding	Metrics Source & Reasoning
Accessibility	The solution requires the student to implement a number of plug-in tools for the solution to be usable.	The solution must be downloaded onto the user's device.	The solution is accessible on Chrome, Safari, Edge, and Firefox with all major operations running server side.	Having the solution on a browser is the most accessible, since students already use a browser to access their online material. This keeps the hardware requirements of the solution low, allowing more students to access the solution.
Organization	The UI is confusing with inconsistent graphics. Hard to learn to navigate, determined through group testing.	The UI has an easy style to follow, however elements are no optimized for ease of use, determined through group testing.	The UI is intuitive; graphics and formatting are consistent and helpful to the user.	Standards from ISO 9241 [3].

6 Constraints

The constraints for the first metric are as follows. Firstly, the actual data collection process must have first year Engineering Science student involved as they are the primary stakeholder as well as users. Secondly, sufficient trials of data must taken in order to create a observable trend for determining the accuracy of the estimation model. Last but not least, minimal amount of outliers should not be included in final analysis for the legitimacy of conclusion.

The constraints for the other two qualitative metrics are equivalent to the "Unacceptable" standard indicated in the above rubrics.

7 Evaluations of Reference Designs

There are currently no products that address the specific problem of the planning fallacy using computational analysis techniques. However, there are countless examples of products that use machine learning libraries and adaptations of recognized algorithms to solve similar problems. Much of the design challenge in this case is deciding which approach would be most efficient for the purposes of this opportunity. To help develop an understanding of this, we will explore the potential integration of Gradient Boosting as a potential way of delivering a prediction to the user.

Gradient Boosting is a machine learning technique by which decision trees of a specified size (typically 8-32 branches deep) are continuously constructed and then assigned weights based on their residuals with the given data, where the weight of the tree corresponds to how much 'say' the tree has in the final decision [4]. The process starts by averaging the results, in our case these would be the reported times to complete, and generates a set of residuals based on how the average compares to the actual data. A decision tree is then constructed using parameters provided which relate to each datum (like estimated time to complete, quality of work, etc.) to, hopefully, decrease the value of the residuals. Depending on how well it does this, the decision tree is assigned a weight before the process is repeated until a 'perfect tree' which matches the data exactly is created or a predefined limit to the number of trees is met. The resulting model can be used to predict future values depending on the parameters given to the program (estimated time to complete, quality of work, etc. in this example).

While we cannot know exactly how effective an exclusive implementation of gradient boosting will be until an experimentation is performed, we can speculate about its shortcomings. For instance, a lot of data will need to be provided before a model that can consistently predict time to completions will emerge. This would likely require that the information from many students be collected and processed before a model that approximates group behaviour better than it does individual behaviour. While this may be better addressed by applying a clustering algorithm before gradient boosting to find regularity in subsets of students, thereby increasing the proportion of a group a single student represents, the result would still not be personalized to the student; only a group the student can be grouped into that is smaller than the sample.

8 Source Extracts

[1] (P1) The planning fallacy refers to a readily observable phenomenon: the conviction that a current project will go as well as planned even though most projects from a relevant comparison set have failed to fulfill their planned outcomes. The term was first introduced to the psychological literature by Kahneman and Tversky (1979, 1982a, p. 415) to describe people’s tendency ‘to underestimate the time required to complete a project, even when they have considerable experience of past failures to live up to planned schedules.

[1] (P7) Figure 1.1 presents data from a related study (Griffin & Buehler, 1999, Study 1) where students reported probability estimates for their best guess completion times for 10 current projects across their academic and personal lives. Once again, about 45% of the projects were completed, compared to an average confidence level of 73%. The calibration plot reveals two interesting patterns: First, probability judgments are overly optimistic across the entire range; this does not match the canonical pattern of overconfidence in knowledge which is marked by underconfidence or underestimation at the low end of the probability scale and overconfidence or overestimation at the upper end, a pattern known as over-extremity. Thus, the phenomenon to explain is an optimistic bias, not an extremity bias...

[5] (P267) The act of prediction, by its very nature, elicits a focus on the future rather than on the past; a future orientation may prevent individuals from looking backward in time.

[5] (P366) This study tested 3 main hypotheses concerning people’s predictions of task completion times: (a) People underestimate their own but not others’ completion times, (b) people focus on plan-based scenarios rather than on relevant past experiences while generating their predictions, and (c) people’s attributions diminish the relevance of past experiences. Results supported each hypothesis.

[5] (P376) The findings suggest that people make more realistic completion estimates when they use their past experiences to inform their predictions

[4] (P1) Gradient boosting constructs additive regression models by sequentially fitting a simple parameterized function (base learner) to current “pseudo”-residuals by least squares at each iteration. The pseudo-residuals are the gradient of the loss functional being minimized, with respect to the model values at each training data point evaluated at the current step. It is shown that both the approximation accuracy and execution speed of gradient boosting can be substantially improved by incorporating randomization into the procedure. Specifically, at each iteration a subsample of the training data is drawn at random (without replacement) from the full training data set. This randomly selected subsample is then used in place of the full sample to *t the base learner and compute the model update for the current iteration. This randomized approach also increases robustness against overcapacity of the base learner.

[3] (P31) Table A.1 - Control Objectives and Controls A.5.1.1: Policies for information security: a set of policies for information security shall be defined... published and communicated to employees and relevant external parties. (P33) A.8.2.1: Classification of information: Information shall be classified in terms of legal requirements... to unauthorised disclosure or modification. (P37) A.12.3.1: Information Backup: Backup copies of information... shall be taken and tested regularly in accordance with an agreed backup policy. (P37) A.12.4: Event Logging: Event logs ... shall be produced, kept and regularly reviewed.

References

- [1] R. Buehler, D. Griffin, and J. Peetz, “The planning fallacy: Cognitive, motivational, and social origins,” in *Advances in experimental social psychology*, vol. 43, pp. 1–62, Elsevier, 2010.
- [2] “Information technology - Security techniques - Information security management systems - Requirements,” tech. rep., Canadian Standards Association, 2013.
- [3] “Iso 9241-110:2020 ergonomics of human-system interaction — part 110: Interaction principles,” tech. rep., International Standardization Organization, 2020.
- [4] J. H. Friedman, “Stochastic gradient boosting,” *Computational statistics & data analysis*, vol. 38, no. 4, pp. 367–378, 2002.
- [5] R. Buehler, D. Griffin, and M. Ross, “Exploring the” planning fallacy”: Why people underestimate their task completion times,” *Journal of personality and social psychology*, vol. 67, no. 3, p. 366, 1994.