Student Project Administration at a Research Laboratory

Sunil Kumar Kopparapu*, Akhilesh Srivastava, P. V. S. Rao
*TCS Innovation Lab - Mumbai, Tata Consultancy Services,
Yantra Park, Thane (West) 400601, Maharastra, India.
sunilkumar.kopparapu@tcs.com

Abstract— **This paper discusses the administration and management of software based student projects at a research laboratory. Based on completed past projects, a set of criteria which determine the success of a project have been evolved. Further a prediction model is developed to guess estimate the success of a project in progress. The paper discusses issues associated with these student projects in terms of investments and gains for the laboratory. Some case studies are presented in support of the developed prediction model.**

I. The Student Project

Student projects are an essential part of a four year graduate Engineering course in most universities world wide. In most Indian universities the project in an industrial setup is spread over a period of two semesters (part time) for a graduate (Bachelor of Engineering or Bachelor of Technology) course in any Engineering discipline; for a post graduate (Master of Engineering or Master of Technology) course it is spread over a period of one semester full time. In effect, the students spend about 8 - 10 hours a week when working part time and about 40 hours when working full time. At the end of the project duration the student is assessed by a panel of examiners and awarded a suitable grade and consequently their graduation degree.

The basic aim of such a project is to give the necessary *industrial exposure* to the graduating students. Such exposure is deemed necessary in order to bridge the transition gap between university education and a professional job. This benefits both the student and the industry and is a welcome aspect of the education process.

The student adviser is a full time member of the research staff at the Laboratory and the students spend about 8-10 hours a week discussing the project plan and related issues with him. About 15 % of his time is spent in planning and projecting the work to be done from several angles to invoke continued interest in the project, as much as 35 % of the time is spent in day to day discussions, planning the project and evaluating the results; 20 % in evaluating the project reports especially considering the intellectual property rights (IPR) angle; about 5 % in planning the infrastructure and making available the necessary computing and logistics facilities and the rest of the time goes in keeping oneself updated about the progress elsewhere in the research community and finding possible extensions to the project plan to make the project output more useful and current.

A. *Successful Project*

Success of a student project is difficult to define especially when it is setup in an R&D environment, because the probable outcome of the work is difficult to predict. We would like to mention that this is different from [1] where the authors discuss about software projects courses to integrate education and research.

Definition 1: A project is deemed successful if the project is brought to a logical conclusion and gives the person involved in the capacity of an adviser a sense of satisfaction and makes him/her more motivated to continue similar studies in the future.

Based on the data available from previous projects done at the Laboratory, a set of factors have been identified as being instrumental for determining the success of a given project [2]. The indicators have been derived from data pertaining to projects undertaken by 50 students from as many as 6 different colleges over a period of 25 months. These measures can be categorized into *essential*[1] and *non-essential*[2] in terms of driving a project towards success. They are:

1) Clarity of problem definition at the time of initiating the project
2) Team Availability: Either full time (for most of the post graduate courses) or part time (for graduate courses) and their overall dedication to the project work
3) Team Size: defined as small if the team had less than 3 students; else large.
4) Project Background: whether all available details have been made available to the students at the start of the project
5) Team cooperation: determined based on how the assignments were divided among the team members and how they interacted among themselves.
6) Motivation: the excitement within the team about the problem being addressed.
7) Adviser Time: extent of guidance from a Laboratory research staff on an ongoing basis.
8) Team composition: Boys or Girls or mixed.
9) Prerequisites: Whether students possessed the prerequisites required for the project

[1]items 1 to 11
[2]items 12 and 13

10) Interaction: with other members of the laboratory and in particular with the head of the laboratory.
11) Targets being set: short goals so that focused results are obtained in short intervals of time
12) Skill set: of the team depending on the project. For example programming skills for an implementational project.
13) Data availability if the project needs it.

B. *Unsuccessful Project*

A set of indicators (based on previous projects done at our Laboratory) that more often than were present in projects which were unsuccessful in the sense of satisfaction derived at the end of the project.

1) Part time: involvement because of their on and off pattern of work on the project, this had a significant dent on the success of a project.
2) Research like problems often failed to invoke motivation and necessary enthusiasm from the students. It seems that most of the students wanted hands-on experience in 'implementing' something rather than *'understanding'* something.
3) Large: teams (more than 3 members) were failures because of *too many cooks.*
4) Shuffling activities of people within a group was not useful, it created confusion.
5) Undefined end goals of a project infused confusion and did not provide necessary motivation for the project completion.
6) Large training time left very little to work on the actual project.
7) Lack of planning short term goals drove the projects haywire and did not result in a successful project.

Note 1: Clearly a number of these measures are very subjective and difficult to measure. But these can be handled by defining what each of the measures mean. We do not discuss this aspect in this paper.

II. Management and Related Issues

Resource allocation, time management, issues related to protecting copyrights, and intellectual property rights (IPR) are crucial from the laboratory stand point. Issues related to patenting of the work carried out at the laboratory, very often results in the students not being able to report the complete work in their thesis. This is a cause for worry for the students because it could be misinterpreted as *not much work done* by the assessing examiner. These are some of the minor pitfalls for the students working with a research laboratory dwelling on some research issues in frontier areas.

1) *Time Management and Investment:* A significant amount of the research advisers time goes into planning the project and keeping tabs of the progress and reviewing the documentation and the source code. The allocated student project titles are in line with the overall goals of the laboratory. This benefits the laboratory in in two ways, (i) the time invested by the laboratory member is in tune with the objectives of the laboratory and (ii) opens up a good platform to try out novel and drastically different ideas without any dents to the time determined goals of the laboratory.
2) *Resource Management:* Resource allocation in a small setup for a two or three student groups (averaging 3 per group) becomes a challenge. It requires extensive scheduling so that none of the groups are starved of computational resources; but at the same time three student groups coming to the laboratory on three different days of the week means that the involved Laboratory member has little time for anything else. A compromise is usually worked with all the student groups visiting the laboratory on one or two days of the week and allocating work amongst the groups to balance between use of computers and digesting literature to aid understanding concepts related to the project.
3) *Academic Management:* A student project while giving the flexibility of trying out novel ideas has the disadvantage of being very time specific, determined by the academic calendar of the college. The positive aspect of this is that ideas can be slept over and brought to a level of maturity before experimenting with the ideas as student projects. On the down-side this could result in *an unbeaten iron rod when it is hot.* Usually, the research staff plans the student project to fit into the academic schedule of the students. The project plans also provides time margins to enable students to prepare for competitive exams; which is a common practice for a final year student in India.
4) *IPR Issues:* Intellectual property rights is an important issues with any organization. A non disclosure agreement (NDA) is signed by every student who starts his project with the laboratory. At the very beginning of the project, the students are specifically cautioned that the research staff guiding them will have a final say on what goes into their final dissertation; this is primarily to protect patentable ideas generated within the laboratory. So far this stance has worked out well without undue discomfort to the project students.
5) *Students will be Students:* More often *last minute* report writing and submission seems to be the trend with the project students. This behavioral phenomenon is taxing on the member of the research staff, who not only has to adhere to his personal deadline which are quite often tight, but also has to verify the contents of the thesis at a very short notice. Reviewing the contents of the thesis by itself is not an issue, but keeping the IPR issues in mind the reviewing process is pretty stringent and has to pass through careful eyes.

III. Project related Specific Data

Table I shows at a glance the data collected over a period of 25 months[3]. These are some of the indicators that one can

[3]Note the the actual names of the colleges have been masked to protect the identity of the colleges and also because it probably does not matter.

S.No	College	Size	Course	F(P)T	Performance
1	College A	5	Grad	PT	14.2
2	College B	6	Grad	FT	18.8
3	College C	3	Grad	PT	25
		5	PGrad	FT	29
4	College D	4	Grad	PT	11
5	College E	3	Grad	PT	23
6	College F	1	Grad	PT	11
7	College G	1	Grad	PT	11
Overall	6	29			18.58

TABLE I
DATA AT A GLANCE. THE PERFORMANCE WAS CALCULATED BASED ON INTERNAL ASSESSMENT ON A SCALE OF 30, USING THE INDICATORS AS SUGGESTED IN SECTION I.

	Feature	Weights
c_1	Problem Definition	0.13
c_2	Full time and dedicated team	0.11
c_3	Small team (size 2/3)	0.08
c_4	Background of the project clear	0.08
c_5	Team cooperation	0.08
c_6	Team excitement	0.05
c_7	Good and continued guidance	0.11
c_8	Girl team	0.03
c_9	Students have prerequisites required	0.05
c_a	Lab Interaction	0.09
c_b	Short goals	0.12
c_c	Programming skills	0.07
	Total	1.00

TABLE II
CRITERIA AND THEIR RELATIVE IMPORTANCE.

derive from the data-set purely based on the performance of the students and the success of the project.

- College C and College B students seems to perform very well.
- College D, College F and College G students seem to perform below average.
- Full time (FT) students seem to perform better than their part time (PT) counterparts with the exception of College C and College E.

A. Useful Selection Criteria

The students selected to pursue a project at the Laboratory is based on certain criteria which have been listed below. The selection criteria are derived from the data in Table I.

- Full time post graduate students are preferred
- Preference to students in this order (a) College C (Full time), (b) College C (Part time), (c) College E (Part time), (d) College B (Full time), (e) College Aand (f) College D, College F, College G

While full time (FT) students seem to outperform the part time (PT) project students due to their continuous presence albeit over a short period of time; this criteria does not *always* reflect into a *successful* project. It is observed that a dedicated and more capable team of part time students make up for the inadequacy of full time presence. But from the Laboratory standpoint, full time students are preferred because of logistics and management issues.

Based on the earlier successful projects (prior to July 2004) a set of weights have been assigned to all the criteria that we believe resulted in making the project a success. The list of criteria and the associated weights are given in Table II. These weights were normalized to capture the relative importance of one over the other. For example, setting up short time goals (0.12) for the student project came next only to a well defined problem (0.13) being suggested as a student project. Here c_i's can take a value on a scale of 1 to 10.

$$\begin{aligned} f(c) &= 0.13c_1 + 0.11c_2 + 0.08c_3 + 0.08c_4 \\ &+ 0.08c_5 + 0.05c_6 + 0.11c_7 + 0.03c_8 \\ &+ 0.05c_9 + 0.09c_a + 0.12c_b + 0.07c_c \end{aligned} \quad (1)$$

Note 2: Clearly the model is a linear weighted sum of the inputs. The weights themselves have been derived by doing a regression analysis on the set of data that was available.
Observe that Table II in *some sense* models a successful student project (2) and can be used to predict the outcome (success or failure) of a yet to begin student project. The following criteria is used to predict the effectiveness of a project

$$\begin{array}{rc} 0 \leq f(c) \leq 5.0 & \text{Failure} \\ 5.1 \leq f(c) \leq 6.0 & \text{Moderate Success} \\ 6.1 \leq f(c) \leq 7.0 & \text{Success} \\ 7.1 \leq f(c) \leq 9.0 & \text{Good Success} \\ 9.1 \leq f(c) \leq 10.0 & \text{Super Success} \end{array} \quad (2)$$

Note 3: The guidelines for interpreting this weighted sum as a prediction for the likelihood of success are based on the actual data. However a large value of $f(c)$ indicates a successful outcome.
Note 4: In [3] authors talk about the development of a framework to validate software measurements. We believe this is not applicable for the R&D type of projects that we are addressing in this paper.

IV. STUDENT PROJECT PREDICTION

The overheads associated with a student project on the laboratory is quite large in terms of resource usage and research staff time. For this reason it becomes crucial from the laboratory point of view to be aware of the outcome of a project even before it has started. With this in mind, we access the outcome of a project based on the developed student project prediction model (2). A student project falling in the category of *Success, Good Success, Super Success* are the projects that are initiated and the rest are shelved until the right combination of the project and the students is found.

Three student projects were evaluated. All the student projects started in Jul '05 and completed in May '06. The initial assessments were done in Jul '05 to get an idea of how the projects would progress.

1) Offline Script Recognition – involves the recognition of script with the final aim of reliably recognizing and verifying signatures. The parameter set used in this recognition process and the procedure adopted are the key to the development of a script recognition software.
2) On-line Script Recognition – involves the development of tools and procedures to extract parameters from script data (output of a CrossPad) so as to enable recognition of cursive characters.
3) Staff Management System – involves the development of an employee attendance monitoring system to log and time the working hours of an employee. It has various features which are unique.

The predictive case studies of these projects are described in the following Sections.

A. Case Study: Offline Script Recognition

The Offline Script Recognition project was assessed as a *success*, having ranked at ≈ 7 on a scale of 10 using Table II. The weights in column three are based on our assessment of the students who chose to do complete this project at the laboratory. The weights also reflect our preparedness in terms of being able to provide background material to the students and also being able to allocate time to guide the students.

Feature	Wt	Assess		Assess	
c_1	0.13	9	1.17	10	1.30
c_2	0.11	4	0.44	5	0.55
c_3	0.08	8	0.64	10	0.80
c_4	0.08	7	0.56	10	0.80
c_5	0.08	7	0.56	10	0.80
c_6	0.05	7	0.35	10	0.50
c_7	0.11	10	1.10	10	1.10
c_8	0.03	0	0.00	0	0.00
c_9	0.05	10	0.50	10	0.50
c_a	0.09	6	0.54	10	0.90
c_b	0.12	10	1.20	10	1.20
c_c	0.07	10	0.70	10	0.70
Total	1.00	Jul '05	6.99/10	Oct '05	9.15/10

TABLE III
PREDICTION OF SUCCESS: OFFLINE SCRIPT RECOGNITION.

The student project scored 6.99 on a scale of 10 at the beginning of the evaluation process (Jul '05) and at a midway assessment (Oct '05) it stood at a proud 9.15 moving from a predicted *success* to a *strong success*.

B. Case Study: On-line Script Recognition

Based on the model (Table II) the On-line Script Recognition project scored a 6.86 on a scale of 10, making it a *success* based on our model (2) and scale (3).

The mid term (Oct '05) assessment moved from the predicted *success* to a *good success* by the midterm; moving on our scale from 6.86 to 7.89. A careful look at the assessment reveals that the shift from *success* to *good success* is mainly due to the *improvement* in our own understanding about the project and the goal settings. For some odd reasons the students seem to be less interested (compared to our initial assessment!), which is surely sad.

Feature	Wt	Assess		Assess	
c_1	0.13	9	1.17	10	1.30
c_2	0.11	4	0.44	2	0.22
c_3	0.08	8	0.64	10	0.80
c_4	0.08	8	0.64	10	0.80
c_5	0.08	8	0.64	4	0.32
c_6	0.05	4	0.20	5	0.25
c_7	0.11	7	0.77	10	1.10
c_8	0.03	0	0.00	0	0.00
c_9	0.05	7	0.35	6	0.30
c_a	0.09	7	0.63	10	0.90
c_b	0.12	8	0.96	10	1.20
c_c	0.07	6	0.42	10	0.70
Total	1.00	Jul '05	6.86/10	Oct '05	7.89/10

TABLE IV
PREDICTION OF SUCCESS: ON-LINE SCRIPT RECOGNITION.

C. Case Study: Staff Management System

The Staff Management System project was a *large* project and required a collective effort from two groups of students. One set consisted of 3 students while the other consisted of 2 students. Each of these two teams has been evaluated separately because of distinct difference between the teams and the nature of work. Based on the model (Table II) the two teams scored 5.12 and 6.41 respectively.

Feature	Wt	Assess		Assess	
c_1	0.13	9 (9)	1.17(1.17)	10(10)	1.30(1.30)
c_2	0.11	3 (7)	0.33(0.77)	10(5)	1.10(0.55)
c_3	0.08	3 (3)	0.24(0.24)	10(10)	0.80(0.80)
c_4	0.08	8 (8)	0.64(0.64)	10(10)	0.80(0.80)
c_5	0.08	7 (7)	0.56(0.56)	10(10)	0.80(0.80)
c_6	0.05	2 (7)	0.10(0.35)	10(6)	0.50(0.30)
c_7	0.11	8 (8)	0.88(0.88)	10(10)	1.10(1.10)
c_8	0.03	0 (0)	0.00(0.00)	0(0)	0.00(0.00)
c_9	0.05	2 (7)	0.10(0.35)	10(10)	0.50(0.50)
c_a	0.09	0 (0)	0.00(0.00)	0(0)	0.00(0.00)
c_b	0.12	8 (8)	0.96(0.96)	10(10)	1.20(1.20)
c_c	0.07	2 (7)	0.14(0.49)	10(10)	0.70(0.70)
Total	1.00	Jul '05	5.12 (6.41)	Oct '05	8.64 (7.56)

TABLE V
PREDICTION OF SUCCESS: STAFF MANAGEMENT SYSTEM.

The student projects were given a good ahead sign though they were according to (3) were barely into the *success* zone. There were two reasons for this (i) we wanted to see if there was a possibility of the predicted *barely success* project would nose dive or shoot up or remain a *bare success*. Midterm assessment (Oct '05) is significantly different from our initial

Student Project	Predict (Jul '05)	Observed (Oct '05)	Actual (May '06)
On-line to Offline convertion	Success	Super Success	Super Success
On-line script recognition	Success	Good Success	Success
Staff Management (Team 1)	Mod Success	Good Success	Good Success
Staff Management (Team 2)	Success	Good Success	Good Success

TABLE VI

ESTIMATES OF PROJECTS AT A GLANCE; OBSERVED, PREDICTED AND ACTUAL.

assessment. The large shift is primarily because a cooperative effort within the team significantly boosted the score and nullified the effect of a large team size. While this is good for the success of the student project it reflects that we need to refine the model (2) by giving less relative weightage to the team size.

V. CONCLUSIONS

A set of *motivated* students working cooperatively seem to be a good recipe for a successful student project. We have predicted the outcome of *in progress* student projects based on simple model developed using data from successful previous projects. The success story of the projects is tabulated in Table VI.

The initial project outcome predictive estimate have been lower than what they stood in Oct '05, which suggests that the model (2) is a pessimistic model. Necessary fine tuning of the model parameters are required to capture a more real model. The development of a weighted sum model, though simple, has been very useful in selecting students who apply to do a project with the laboratory.

Note 5: Barbara et. al [4] discuss about evaluation and assessment specifically to software engineering, however in our case though it was an software implementation it had an R&D component that required iterative experimentation and analysis.

REFERENCES

[1] Supannika Koolmanojwong and Barry W. Boehm, "Using software project courses to integrate education and research: An experience report," in *CSEE&T*. 2009, pp. 26–33, IEEE Computer Society.

[2] Norbert Seyff, Neil A. M. Maiden, Inger Kristine Karlsen, James Lockerbie, Paul Grünbacher, Florian Graf, and Cornelius Ncube, "Exploring how to use scenarios to discover requirements," *Requir. Eng.*, vol. 14, no. 2, pp. 91–111, 2009.

[3] Barbara Kitchenham, Shari Lawrence Pfleeger, and Norman E. Fenton, "Reply to: Comments on "toward a framework for software measurement validation"," *IEEE Trans. Software Eng.*, vol. 23, no. 3, pp. 189, 1997.

[4] Barbara Kitchenham, Pearl Brereton, David Budgen, Stephen G. Linkman, Vicki L. Almstrum, and Shari Lawrence Pfleeger, "Evaluation and assessment in software engineering," *Information & Software Technology*, vol. 39, no. 11, pp. 731–734, 1997.