Proceedings of the ASME 2013 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference
IDETC/CIE 2013
August 4-7, 2013, Portland, Oregon, USA

DETC2013-12700

THE INFLUENCE OF TIMING IN EXPLORATORY PROTOTYPING AND OTHER ACTIVITIES IN DESIGN PROJECTS

Anders Häggman
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139
E-mail: haggman@mit.edu

Tomonori Honda
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139
E-mail: tomonori@mit.edu

Maria C. Yang
Department of Mechanical Engineering
and Engineering Systems Division
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139
E-mail: mcyang@mit.edu

ABSTRACT

The importance of prototyping in the design process has been widely recognized, but less research emphasis has been placed on the appropriate timing and detail of so-called "throwaway" prototyping during the preliminary design phase. Based on a study of mid-career professional graduate students, statistically significant correlations were found between the time such prototypes were created and design outcome. Building prototypes early on in the design process, or performing additional rounds of benchmarking and user interaction later on during the project (in addition to the typical early stage efforts), correlated with better design outcome, although the total time spent on these activities did not. The correlation between project presentations and reviewer scores are also touched upon. These findings suggest that the timing of design activities is more important than the time spent on them.

INTRODUCTION

It is widely recognized that decisions made in the early stages of the design process have great bearing on the outcome of the design in terms of design success, cost of manufacturing and time to market. The later in the design process problems are identified and changes made to the design, the more costly it becomes [1-3]. The question then is, how should designers go about exploring the design space in the most efficient way possible, find a desirable design direction, and execute a successful design?

This research examines the timing of prototyping related activities in the early stages of the design process. Research has been conducted on prototyping in the early stages of the design process, with some arguing for early and frequent prototyping as a way to test ideas early [4, 5] and to help build confidence in design concepts in a team [6].

There are several different design activities that have been recognized as important to design outcome and have been widely researched. These range from prototyping and sketching to benchmarking and collecting input from end-users.

Prototyping is often thought of in terms of the particular technology or materials involved in creating the prototype, such as 3D printing or aluminum stock. This paper considers prototypes from a different perspective, that of as an artifact that design teams use to evaluate potential design concepts before further development [7, 8]. These early stage prototypes are created with the express understanding that they will be discarded after evaluation, and are thus "throwaway" prototypes [9]. Using this strategy encourages building the "cheapest" prototype that can still provide needed information, meaning that such prototypes may be made of relatively inexpensive materials and are fast to fabricate. This approach to fabrication has been expressed as "fail early to succeed sooner," popularized by the design firm IDEO and has been widely cited as a strategy for early stage design [4].

Although there has been a significant amount of research on effective ways to conduct different design activities themselves (for example the most appropriate types of sketching a design team can do [10, 11] or how brainstorming should be conducted [12]), less research emphasis has been placed on the specific timing of these activities. A traditionally accepted belief is that certain design activities such as benchmarking, gathering user input and ideation, to name a few, should be conducted in the early stages of the design process [13].

This paper examines the relationships between the amount of time spent on, and the timing of, different design activities (in other words, *when* a certain activity was conducted) with design outcome.

RESEARCH QUESTIONS

Our research questions were primarily aimed at investigating the timing of prototyping and other design activities in the early stages of the design process. The research questions were specifically:

1. Did the timing of the prototypes, in other words *when* the prototypes were built, matter with respect to design outcome? And did teams that spent more time prototyping at various points in the design cycle fare better than those teams that spent less time?
2. Did the timing of other design activities, such as sketching, user interaction, benchmarking, and presentation preparation correlate with better design performance?

RELATED WORK

Physical prototypes are tangible and visual representations of design concepts, and a means to communicate an idea to others [14, 15] and as such, act as a shared vision for all stakeholders involved. The process of building and developing physical models can bring up design issues in ways that alternative representations often cannot [16].

As mentioned earlier, there is some disagreement with regard to optimal prototyping practices, with some advocating on behalf of early and frequent prototyping [4, 17], whereas others caution against excessive prototyping due to the time and cost involved in doing so [18]. Prototyping may introduce the risk of premature design fixation, or commitment, to a design choice [19, 20].

Research with novice designers using physical models as idea generation tools found that they did not experience design fixation, and generated more functional ideas than those who only sketched [21], while other research found that engineering design faculty experienced design fixation when sketching and presented with an inferior design alternative as an example [22].

Some preliminary results regarding prototyping suggest that the making of physical models supports the designers' internal reasoning processes by rectifying flaws in their mental models, leading them to produce a larger proportion of ideas that satisfy the design requirements [19, 21].

Front-loaded problem-solving is presented as a strategy to improve development performance by shifting the identification and solving of design problems to earlier phases of the design process, which is in essence the aim of early-stage throwaway prototyping; identifying problems with current concepts and generating alternative design directions with minimal effort and invested time. Solving problems is cheaper and less time-consuming the earlier they are identified in the design process, before having committed to a certain design direction [3].

The building of prototypes is often a trade-off between the level of detail of the prototype, and invested effort, time and cost, and should therefore be built as inexpensively and quickly as possible, while still providing useful insights for the designers [23]. Although time spent on a design project is important, committing raw hours in and of itself is not a guarantee for success, and it has been found to be more useful to spend time consistently, and to put forth effort on a well scoped design [16]. Some research also suggests that spending proportionately more time on prototyping [16], sketching [24], and user feedback [25, 26] in the early stages of design process correlates with better design outcome.

METHODS

The data for this study was gathered during a semester long, graduate level design course at a US university in 2011. The class consisted of 67 mid-career professional students who were observed throughout the semester. In other words, the students were enrolled as full-time students, but had a considerable amount of previous work experience. Another study found that in previous years the students taking the class in question had an average of 10.2 years of industry work experience in a technical field and were experienced in product development activities [25], and it is reasonable to assume that the students during the year in question had a comparable amount of experience. This class was also the basis of another study by the authors [27], though this current paper delves more deeply into prototyping specifically. This current paper draws on some of the same data from that earlier study, in particular estimations of design outcome measures and ways to address missing timesheet data.

The course was a semester long graduate level design course, where students worked in teams ranging in size from 4 to 7, with 5 students per team being the median team size. The theme for 2011 was "Healthcare and Healthy Living" and the teams were tasked with creating a functioning proof-of-concept prototype at the end of the semester, and were given a budget of $800 for doing so.

Figure 1 shows three final proof-of-concept prototypes. From left to right; a seat cushion for exercising your abdominal muscles while at work, an armband holder for exercise necessities such as disinfectant wipes for the gym, and a customized iPad holder for use on treadmills and other exercise equipment.

Figure 1 Example proof-of-concept prototypes delivered by the teams at the end of the course

Data collected: The course was divided into 7 time periods, each roughly 2 weeks long. The main type of data collected was timesheets. This refers to bi-weekly reports that the students submitted electronically which detailed the number of minutes the student had spent on 13 different activity categories. The name and description of each category is presented in Table 1. below, in the form it was presented to the students. For clarity, in this paper 'building' will be referred to as 'prototyping', and 'presentation preparation' will be referred to simply as 'presentation'.

Table 1 Categories of activities documented on timesheets

Category	*Description*
1. User Interaction	Investigating the needs of customers and users, and testing concepts and prototypes.
2. Market Research	Investigating or identifying markets and customers at a large scale.
3. Benchmarking	Researching existing and competing products. Includes patent search.
4. Concept Generation	Formulating design solutions, i.e. brainstorming.
5. Concept Selection	Evaluating and choosing concepts (includes product testing).
6. Design: Sketching	Planning the details of your concept and how it will function and look, using sketching.
7. Design: CAD	Planning the details of your concept and how it will function and look, using CAD and other software tools.
8. Design: Anything Else	Planning the details of your concept and how it will function and look, where the activities don't fit in either of the previous two sections.
9. Building	Fabricating or coding a concept. Includes prototyping.
10. Business Plan	Planning future development and financial projections.
11. Presentation Preparation	Preparing for team presentations.
12. Administrative	Scheduling and managing meetings. Includes time spent in meetings that does *not* involve the activities above.
13. Other	Please explain.

Supporting data about the teams' prototypes and their design process was collected through questionnaires about the prototypes, photographs of prototypes, e-mail interviews with the students, scans of their design notebooks, through milestone presentations throughout the semester and through the teams' final documentation. Timesheets were completed every second week, and other supporting material less often; students were required to scan and submit at least 3 significant sketches from their design notebooks every two weeks and prototype questionnaires were requested to be filled in after each prototype. There were three milestone design reviews throughout the semester, in addition to the final presentations. An overview of the timeline with relevant milestones is given in Figure 2 below.

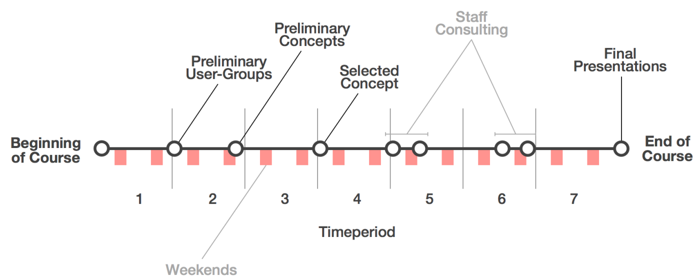

Figure 2 Timeline of course with relevant milestones

A total of 401 accepted timesheets were collected (some duplicates and empty timesheets were rejected), which corresponds to a reply rate of 85.5%.

Although the response rate was good, two sets of data were used in the subsequent analysis to determine if the missing timesheets amounted to a significant omission. In the original data set, missing timesheets counted as 0 minutes towards the project during that time period. In a modified data set, the missing data values were imputed based on the students' average effort level compared to the team, as well as on that student's average effort level in the different work categories. For example, if a student spent on average 1.25 times more time working on the project than his or her teammates, it was assumed that during the time period of the missing timesheet report, the student also spent 1.25 times more time on the project than the average team member. In other words, the estimated work effort by the student was affected by how much time his or her team members were also spending on the project.

Likewise, for each student missing a timesheet, an average effort level per category was computed based on the timesheets that the student had already submitted. These were used to estimate the time spent on various activities during the time periods with missing timesheets. For example, if a student spent a quarter of their time (on average) sketching during the other time periods during which they had submitted a report, it was assumed that in the time period with the missing timesheet report the student had also spent a quarter of their time sketching. This was repeated for every student, until a "complete" data set was generated. However, the original data set was also retained, and the two were compared later on to check if the missing timesheets led to any significant changes in the results.

Design outcome measures. One of the enduring problems with analyzing design team performance is the difficulty of evaluating the end results. In this study, an expert panel grading the teams at the end of the course was used to evaluate the design success of each team.

At the end of the course, the final projects were presented in front of the class, as well as a panel of 8 industry professionals and academics. The panel consisted of two professors, two senior lecturers, and four industry experts working in product design or product development. The professors and senior lecturers also had a background in industry. Four of the eight panel members were not familiar with the previous work they had done during the semester, whereas the remaining four academic panel members had witnessed the teams' progress throughout the term. Interestingly, however, analysis showed that there was no statistically significant correlation between reviewers who had followed the teams throughout the year and the scores that the teams received in their evaluation. In other words, there seemed to be no grading bias based on familiarity when the teams were graded on their final presentations and prototypes during the panel review.

Four different types of evaluations were conducted. Each jury member gave points on a 7-point Likert scale to the teams in several distinct categories detailed below in Table 2..

Table 2 Final presentation review questionnaire given to panel members

Nr.	*Topic*	*Question*
1. a)	User and market need	Do you think this user need is compelling, clearly defined and unmet by existing products?
b)		Does the team have an understanding of where the product fits in with its competitors?
c)		Do you think there is a viable market for this product?
d)		Does the concept the team has developed meet the user need?
2. a)	Prototype	Does the prototype work as it is intended?
b)		Does the prototype effectively execute the design intention?
c)		Does the prototype communicate the product concept convincingly?
d)		Did the team thoughtfully consider users throughout their design process?
3.	Business assessment	Is the business case for the product plausible?
4.	Presentation	Was the presentation well structured and delivered?

The jury members also gave the teams an overall ranking (i.e. “Team C was the fourth best team”), after which the jury convened, and discussed the rankings further to achieve consensus, and finally, each team member received a final grade for the class.

Based on the problems with the other types of grading (discussed later), the individual team rankings of the jury members (before convening and discussing the scoring) were used. It was seen to be the most accurate and representative opinion of the expert panel members for a team’s performance. This resulted in each team receiving a rank number (1st, 2nd, 3rd, and so on) from each jury member.

The main data used in the analysis was gathered through bi-weekly timesheets, and consisted of minutes of activity, divided into 13 different activity groups (as can be seen in Table 1). The times were then correlated with team performance in five different ways, for both the original data (with missing timesheet reports) and the modified data (with estimated work times inserted). The five different ways the times were used were: 1. absolute, 2. cumulative, 3. percentage per task, 4. percentage per time period, and 5. percentage of total time.

For 1., the absolute time, the time spent on any particular activity during one time period was compared to the success of the design team (success being measured in this study as getting favorable ratings from the reviewers). For example, did teams that spent more time prototyping during time period three fare better than those teams that spent less time?

For 2., the cumulative time, the time spent from the beginning of the course up until that time period was compared to overall success. For example, did teams that spent more time sketching up until the fourth time period do better than teams that spent less time sketching?

For 3., the percentage per task, the percentage of the total time used on that task, that was used in a certain time period, is compared to overall design success. In other words, regardless of how much actual time in minutes was used on a certain task (for example, market research), the percentage of it that was used in a certain time period was correlated with design success. For example, if a team spent a larger percentage of their total market research time during a certain time period (time period two, for example), did those teams fare better than teams that spent a smaller percentage of their total market research time during that time period?

For 4., percentage per time period, the percentage of the total time that the team spent on the project during a certain time period was correlated with overall success. To give an example, if a team spent a larger percentage of the total time that they spent during time period three on user interaction, did they fare better or worse than a team who spent a smaller percentage of their time on it? In other words, the time spent on each of the 13 activities was summed up per time period, and then percentages calculated for each activity. For example, team B might spend 15% of the total time they spent during time period three on prototyping.

For 5., percentage of total time, the percentage of time that a team spent on a specific activity during a specific time period out of the total time that the team spent on the project throughout the semester was compared to the team's overall success.

As there were five different categories or ways at looking at the time spent on the project, and two data sets (original and modified), there were ten different categories altogether, and seven time periods, that were correlated with design success. Hence, there were 70 different values per team that were correlated with their overall success. A Spearman rank correlation was used to correlate time spent on various design activities with design success. Qualitative data from interviews and prototype questionnaires was also used as supplemental and supporting data for findings based on the timesheets.

RESULTS AND DISCUSSION

Four different types of grading data were available for each team: 1. ratings (scores) for the ten different rating categories detailed in Table 2. on the previous page, 2. ranking scores (1st, 2nd, and so on), 3. modified ranking scores obtained after reviewers met to discuss grading, and 4. final grades received in the class.

In aggregating ratings, calculations had to take into consideration that the reviewers do not specify a weighting for each of the categories, thereby making each category equally weighted. However, the reviewers internally create a weighting for the different criteria, meaning that simply summing up the scores would give an inaccurate picture of a team's overall performance.

Hence, the overall ranking order that the reviewers created for the teams was seen as the most comprehensive and accurate measure of a team's design performance. The reason that the ranking scores prior to the group discussion were used instead of the ones generated after the group discussion was that analysis showed significant shifts in consensus ranking, perhaps due to social dynamics between the reviewers. Since all eight reviewers were assumed to have roughly the same level of knowledge and expertise, the initial rankings of each reviewer were used instead, giving equal weight to the opinions of each of the eight reviewers.

After compiling all the rankings from the panel members, it was obvious that there was limited agreement between the eight reviewers. The ranking scores of each reviewer were further analyzed, to see if there were any correlations between background of the reviewer (industrial design, business, engineering), the type of reviewer (academic or industry expert), how familiar they were with the students and their progress throughout the course, or their gender and the ranking scores that they gave. No such correlations were found. In other words, no obvious reason was found that would clearly describe why the rankings were so different, other than personal preference. With a panel of eight experts, it had to be accepted that there were going to be differing opinions with regards to the success of each of the designs.

In order to be able to evaluate the efficacy of differing design processes the teams used, agreement on the results had to first be achieved. Four different criteria were used to summarize the ranking scores of the eight reviewers: mean, median, and two different Borda count methods [28]. The Borda count methods used were a) $\sum(14-R_i)$ and b) $\sum(1/R_i)$, where R_n is the rank given by reviewer n. These will later be referred to simply as Borda count 1 and Borda count 2.

When combining the ranking scores given by the reviewers, depending on the criteria used (mean, median, Borda count 1 or Borda count 2), the rank order of the teams changed. However, regardless of the sorting criteria used, the top 5 and bottom 8 teams remained the same. That is to say, that the teams could be robustly divided into two groups of teams: top tier teams, and bottom tier teams, as can be seen from Table 3. Teams C, I, A, L, B were classified as top tier teams, and teams K, H, D, E, M, J, G, F were classified as bottom tier teams.

Table 3 Ranking of project teams depending on sorting criteria

Rank	*Mean*	*Median*	*Borda 1*	*Borda 2*
1	C	C	C	C
2	I	B	I	L
3	A	L	A	A
4	L	I	L	I
5	B	A	B	B
6	K	H	K	D
7	H	M	H	E
8	D	D	D	K
9	E	K	E	H
10	M	E	M	J
11	J	G	J	M
12	G	J	G	G
13	F	F	F	F

Using eight reviewers instead of only two or three meant that it was likely that there would be some level of disagreement, but ultimately using a larger number of reviewers means that the results are less sensitive to any single reviewers personal preferences.

After the teams were divided into top tier and bottom tier teams, the next step was to look at the timesheet data, and see if effort in terms of time spent would correlate with better design performance (higher reviewer rankings). Analysis showed that the total time spent on the project during the course did not correlate with overall design success, even when controlling for the differing number of students per team. This also applies if inspecting the total time spent on any specific activity other than "presentation preparation" in which total time spent correlated with better rankings. For all other design activities, the total time a team spent on that activity did not correlate with design success. However, the timing of that effort was highly important, and when the teams conducted different design activities was statistically significant (Spearman, $p<0.05$).

Calculating Spearman's correlation between the ten different categories (five different ways of looking at the time, for both the modified and original data sets) and whether a team was top tier or bottom tier yielded some interesting results. Even though whether a correlation was statistically significant depended somewhat on which data set was used, and in which way the time was viewed (absolute, cumulative, and so on), four design categories emerged for which statistically significant correlations were present for almost all of the conditions. There were also other design categories that did exhibit statistically significant correlations during certain conditions, but since the correlations were only visible during a few combinations of the criteria, they were disregarded in the main analysis.

The four different categories that showed statistically significant correlations in many of the different cases were: 1. user interaction, 3. benchmarking, 9. prototyping, and 11. presentation. Again, it is important to note, that the total time spent on 'user interaction', 'benchmarking' and 'prototyping' activities throughout the course was not significant per se, it was *when* time was invested in these activities that mattered. Even for 'presentation' in which the total time was important, timing also clearly played and important role. Table 4 shows the time periods during which these four activities were statistically significant with regards to the final outcome, and the number of different configurations of the ten possible configurations (two sets of data, with five different categories each) that were statistically significant ($p<0.05$). As it is hard to justify why a certain way of looking at the time information would be more correct than another, all ten different ways were considered, and the assumption was made that if significant correlations were found regardless of the ways in which the data was viewed, important correlations existed.

Table 4 Spearman correlations between key activities and final outcome. The numbers in the table indicate the number of statistically significant ($p<0.05$) correlations that were found, out of a maximum of 10 possible different correlations. Highlighted time periods were chosen for further discussion due to the large number of statistically significant correlations with final outcome.

	Time period						
Design activity	1	2	3	4	5	6	7
User interaction	0	0	0	6	0	0	0
Benchmarking	0	0	0	0	8	0	0
Prototyping	9	10	2	0	0	0	0
Presentation	0	5	1	1	0	2	1

As can be seen, user interaction, benchmarking and prototyping had certain time periods that were clearly important, and had numerous statistically significant correlations with overall design success, whereas the presentation category correlated with design success during a larger number of different time periods. However, it is important to note that the time spent on presentation correlated in a statistically significant way with the end result in most cases only when looking at the cumulative time spent from the beginning of the course. The only time period with more than two different ways of correlating in a statistically significant way was time period two. The discussion will therefore focus only on the time teams spent working on their presentations in time period two.

Apart from prototyping in time period two, all other design activities were dependent on in which way the times were viewed (in other words, was the absolute time spent during a certain time period correlated with design outcome, or was the cumulative time from the beginning of the course spent on a certain design activity correlated with design outcome, to name a few examples). Building prototypes in the second time period always correlated with better design outcome regardless of the way in which the time was viewed.

Tables 5 through 8 detail the correlations for the four chosen design activities during the five time periods chosen as being the most important with regard to design outcome (shaded in Table 4).

As can be seen from Table 5, the time spent on 'user interaction' in the fourth time period ranged from being highly statistically significant, to not at all significant, depending on how the time spent was viewed (absolute time spent on user interaction during time period four, cumulative time spent on user interaction from the beginning of the course up until time period four, percentage of time spent on user interaction out of the total time spent during time period four, and so on). The correlation rho- and p-values for 'user interaction' during the fourth time period are given in Table 5.

Table 5 Spearman correlations for 'user interaction' in time period 4

*Spearman correlations for **User interaction** in time period 4*	*p-value*	*rho-value*
Unmodified data-set		
Absolute	**0.010**	**0.682**
Cumulative	0.629	0.148
Percentage per task	0.061	0.532
Percentage per time period	**0.001**	**0.813**
Percentage of total time	**0.002**	**0.765**
Modified data-set		
Absolute	**0.010**	**0.682**
Cumulative	0.891	0.042
Percentage per task	0.074	0.511
Percentage per time period	**0.001**	**0.813**
Percentage of total time	**0.019**	**0.637**

Interestingly, time spent on user interaction was not significant during the early stages of the design process, although that is when one would generally expect design teams or designers to engage the end user, to get insight into the design challenge. However, what happened in the design course studied was that there was very little difference between the teams in terms of how much of their time they spent on user interaction in the early stages of the course (perhaps due to the fact that user interaction was a topic in lectures, and the students were reminded about the importance of talking with the proposed end users of their products). Therefore, the fact that user interaction was only statistically significant in the fourth time period does not mean that user interaction should not be conducted at the beginning of a design process. It is simply a result of the fact that there were very small differences between the teams in terms of how much time they spent on user interaction in the early stages of the class. However, top tier teams went back to talk with the user later in time period four, *after* they had already chosen their final concept, to get feedback on their designs and iterate further.

Presumably the bottom tier teams also continued iterating on their design until the end of the class, but they did so with much less input from the users. Figure 3 shows the percentage of their time that teams spent on "user interaction" during the different time periods, with the error bars indicating ± 1 standard deviation. In other words, bottom tier teams spent on average 33% of the total time they worked on the project in time period one, on user interaction. Top tier teams, to give another example, spent 10% of the total time they spent on the project during time period four on user interaction.

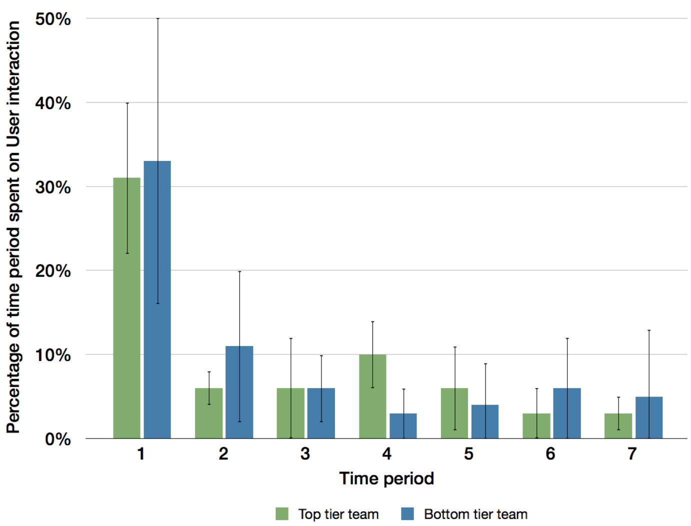

Figure 3 Percentage of time period spent on 'user interaction'

As can be seen, the percentage of time spent on 'user interaction' in the first time period (first two weeks of the class) is similar. However, after selecting their final concepts at the end of time period three, the top tier teams went back to talk to the user (and spent 10% of their time doing so), whereas the bottom tier teams only spent 3% of their time on average engaging the end user in time period four. The difference between top tier and bottom tier teams is highly statistically significant (p-value 0.001, rho-value 0.813). This is consistent with the findings of Lai, et al. [25] that suggest the role of user feedback in the later stages of the design process is critical, possibly even more so than initial user research.

Similarly, and perhaps somewhat surprisingly, the time spent on 'benchmarking' was statistically significant only in the fifth time period, which again, at first glance seems very late in the design process (the course being only seven time periods long). The correlation rho- and p-values for "benchmarking" during the fifth time period are given below in Table 6..

Table 6 Spearman correlations for 'benchmarking' in time period 5

*Spearman correlations for **Benchmarking** in time period 5*	*p-value*	*rho-value*
Unmodified data-set		
Absolute	**0.030**	**0.600**
Cumulative	0.487	0.212
Percentage per task	**0.017**	**0.645**
Percentage per time period	**0.023**	**0.624**
Percentage of total time	**0.023**	**0.623**
Modified data-set		
Absolute	**0.030**	**0.600**
Cumulative	1.000	0.000
Percentage per task	**0.017**	**0.645**
Percentage per time period	**0.023**	**0.624**
Percentage of total time	**0.017**	**0.645**

As with 'user interaction', there is a distinct initial hump in how much time was spent on benchmarking, again, with little difference between top tier and bottom tier teams. However, similarly to 'user interaction' top tier teams went back and spent time looking at their competition in more detail in the fifth time period (as can be seen in Figure 4), *after* they had chosen their final concept for their design at the end of the third time period, and *after* they had talked with their users in time period four.

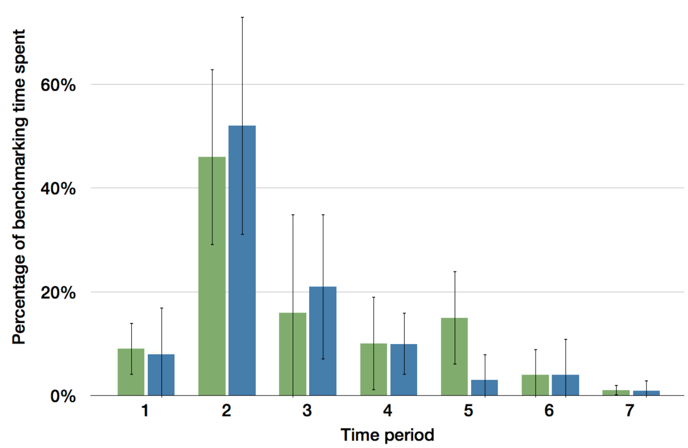

Figure 4 Percentage of time period spent on 'benchmarking'

Figure 4 above shows how large a percentage of the time teams spent on 'benchmarking' was used in the different time periods. 'Benchmarking' in the early stages of the design process did not correlate with design success as all the teams were doing roughly similar amounts of benchmarking and differences between teams were small, but as mentioned, benchmarking in the fifth time period was statistically significant (p-value 0.017, rho-value 0.645).

Contrary to 'user interaction' and 'benchmarking', which were only statistically significant in one specific time period, building prototypes was highly statistically significant during the first two time periods, as well as the third time period when looking at the cumulative time spent from the beginning of the project until the end of the third time period (both for the modified and un-modified data set) with p-values of 0.017 and 0.018 for the unmodified and modified data-sets respectively. The p- and rho-values for prototyping in the first and second time period are give in Tables 7 and 8 below.

Table 7 Spearman correlations for 'prototyping' in time period 1

Spearman correlations for ***Prototyping*** *in time period 1*	*p-value*	*rho-value*
Unmodified data-set		
Absolute	**0.010**	**0.686**
Cumulative	**0.010**	**0.686**
Percentage per task	**0.009**	**0.688**
Percentage per time period	**0.010**	**0.686**
Percentage of total time	**0.010**	**0.686**
Modified data-set		
Absolute	**0.010**	**0.686**
Cumulative	**0.010**	**0.686**
Percentage per task	0.058	0.537
Percentage per time period	**0.010**	**0.686**
Percentage of total time	**0.010**	**0.686**

Table 8 Spearman correlations for 'prototyping' in time period 2

Spearman correlations for ***Prototyping*** *in time period 2*	*p-value*	*rho-value*
Unmodified data-set		
Absolute	**0.010**	**0.686**
Cumulative	**0.001**	**0.826**
Percentage per task	**0.010**	**0.686**
Percentage per time period	**0.010**	**0.686**
Percentage of total time	**0.010**	**0.686**
Modified data-set		
Absolute	**0.010**	**0.686**
Cumulative	**0.001**	**0.826**
Percentage per task	**0.009**	**0.688**
Percentage per time period	**0.010**	**0.686**
Percentage of total time	**0.010**	**0.686**

As mentioned, 'prototyping' was highly statistically significant during the first two time periods, but less so in later time periods. In other words, from a perspective of design success during a constrained semester long design class, it did not matter how much time was spent on building prototypes, but rather, that time was spent in the *early* stages of the class.

Many of the students that were interviewed through e-mail talked about the importance of building exploratory throw-away prototypes, and how they had helped form a common vision for the team, how building them had helped them in creating design concepts as well as weeding out ideas that at first seemed promising but on closer inspection were problematic. As two students in the class put it:

"Talking about ideas was not the same as seeing them… …[the prototype] brought to light some of the more subtle aspects of the concepts that we weren't aware of…"

"I think it was good for our confidence. We all liked once we could see [the prototype] and felt more engaged to the project. Besides it was fun to build it so we got more enthusiastic about the class."

Although, based on the timesheet data, it is not possible to claim causation, based on qualitative analysis of the e-mail interviews of thirteen students from top tier teams, it seems that many of the students themselves felt that building the rough "throwaway" prototypes in the early stages of the design process had been beneficial.

Another possible explanation could be that teams that were more motivated and engaged began building prototypes immediately, whereas teams with less drive waited until explicitly being told by course staff to start prototyping, in which case building early in the design process would be an effect of a motivated team, instead of a cause for better design success.

Figure 5 below compares how the top tier and bottom tier teams divided up the total time that they used on building prototypes. The error bars on the graph represent ± 1 standard deviation.

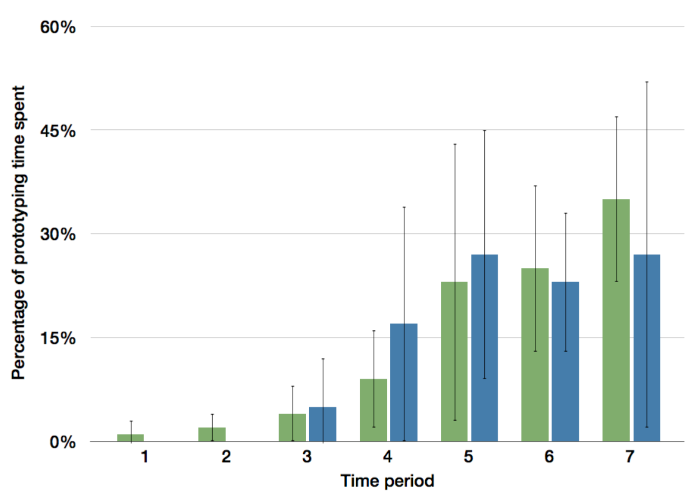

Figure 5 Percentage of time spent on 'prototyping'

The time spent working on building prototypes increased dramatically towards the end of the course, but there were no statistically significant differences in the amount of time spent between the top tier and bottom tier teams, apart from the first two time periods (and when looking at the cumulative time, the third time period). The relatively small amount of time spent building prototypes during the first two time periods was the difference between top tier and bottom tier teams.

It is also interesting to bear in mind, that top tier teams built prototypes *before* having to choose three possible user-groups, *before* having to choose three preliminary concepts, and *before* choosing their final concept, whereas none of the bottom tier teams built anything during the first two time periods (first four weeks of the roughly fourteen week long class) and four out of the eight bottom tier teams reported not building anything until *after* having chosen their final design concept.

Based on interviews and observations, the prototypes built during these early stages of the design process were by nature rough throw-away prototypes built out of paper, foamboard, wooden sticks and Legos, to name a few of the building materials used.

And finally, time spent working on the presentations (either for the final presentation, or for milestone review presentations throughout the course) correlated with success in the class. Although there were statistically significant correlations in several time periods, because most of them were only between one specific time form, the focus was placed on the second time period which had several statistically significant correlations. The correlation rho- and p-values for ‘presentation’ during the second time period are given below in Table 9.

Table 9 Spearman correlations for ‘presentation’ in time period 2

*Spearman correlations for **Presentation** in time period 2*	*p-value*	*rho-value*
Unmodified data-set		
Absolute	**0.019**	**0.638**
Cumulative	0.051	0.551
Percentage per task	0.089	0.490
Percentage per time period	**0.049**	**0.555**
Percentage of total time	0.075	0.510
Modified data-set		
Absolute	**0.019**	**0.638**
Cumulative	**0.033**	**0.592**
Percentage per task	0.223	0.363
Percentage per time period	**0.049**	**0.555**
Percentage of total time	0.126	0.447

Perhaps somewhat surprisingly, time spent on creating presentations for the class correlated with design success. Especially time spent during time period two (see Figure 6), as well as the total cumulative time spent throughout the course (Figure 7).

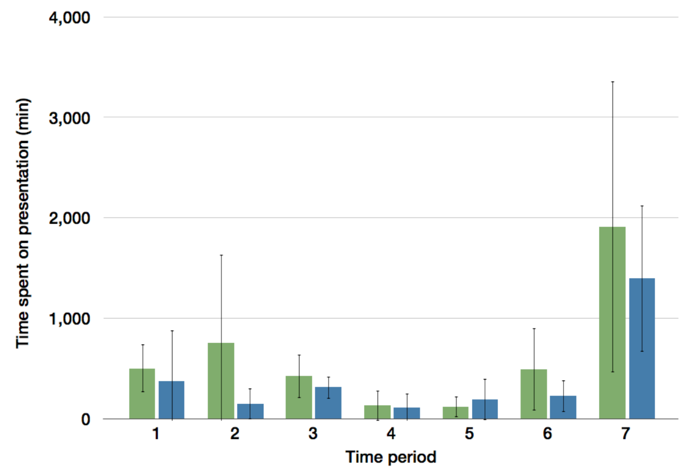

Figure 6 Time spent on ‘presentation’ during different time periods

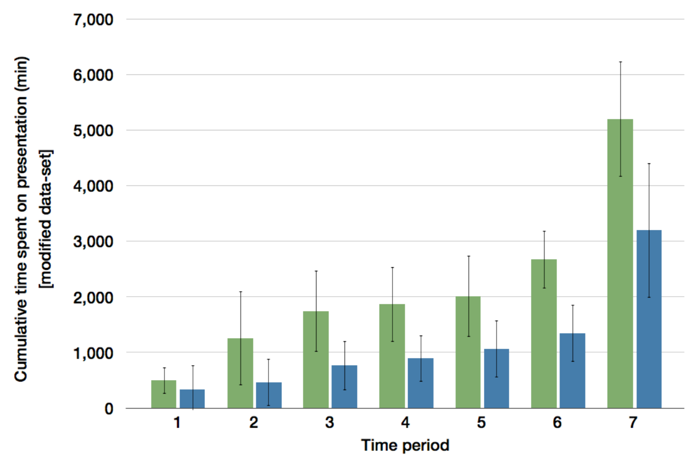

Figure 7 Cumulative time spent on ‘presentation’ during different time periods

If one accepts the assumption that more time spent working on a presentation correlates with a better presentation, then, intuitively, one can imagine that time spent during the last few time periods could correlate with a higher ranking, since a nicer looking and well thought out presentation may impress the reviewers, and therefore make them more prone to giving higher scores.

However, it is more difficult to explain the correlation between time spent working on the presentation in the early stages of the design process and the final rankings received during the final presentations over ten weeks later, since the presentations given at the final presentations were not the same ones that the teams had been working on during time period two.

One possible explanation could be that by investing time in the presentations early on before choosing their final concepts, teams spent more time mentally going through their different design directions and working through perceived issues and challenges with their preliminary concepts in more detail than teams that spent less time preparing their presentations, and consequently had a better understanding of the weaknesses and strengths of their concepts and therefore were perhaps better suited to choosing a successful design.

Another possible explanation could be that time spent on the presentations simply reflects the motivation level of the teams, and that the time spent working on the presentations is nothing more than an indication of the teams motivation, which may have been a factor in their ultimately better success. However, statistical analysis showed there was no statistically significant difference between the ratios of team members that helped create the presentations in top tier versus bottom tier teams. So, if using the number of team members that worked together on the presentation as a proxy for team motivation, there was no correlation between working on the presentation in the early stages of the process and team motivation.

Finally, as many of the reviewers did follow the teams throughout the course, giving a good presentation early on in the course could have already made some reviewers view certain teams in a more favorable light early on in the course, resulting in them giving the teams higher rankings during the final presentations. However, analysis on the reviewers also did not find any statistically significant differences between reviewers who had followed the teams throughout the course and those who were only present for the final presentations.

In addition to the importance of benchmarking, user interaction and prototyping, sketching has been widely recognized as an important part of the design process [24, 29, 30].

However, in this study, no statistically significant correlations were found between sketching and design outcome. One explanation for this could be that simply measuring time spent on sketching is not an appropriate metric to determine sketching effort in a design process. Sketching may also be too quick to be captured in high enough fidelity with the self-reported timesheets, where oftentimes students round to the closest 10 or 15min interval. It is not known if quicker sketching sessions occurred and they have gone unreported, or if they have been rounded up to 10min. A more appropriate metric to capture sketching may be to record the *number* and *fidelity* of the sketches, and at what point in the design process they occurred. These are questions that cannot be answered with the data gathered in this study, but present opportunities for future work..

Limitations. There are several limitations with the study. One of the main concerns deals with the fact that the time spent on different activities is self reported, and error may be introduced by students rounding off their times. It is of course also challenging to estimate exact times for several different design activity categories for two weeks at a time. The students were encouraged to keep personal logbooks and notes regarding their time usage to mitigate this problem. Also, the assumption is that the students will make similar estimation errors, and that with a sample of 67 students, the answers will be representative of each team and comparable between teams, even if the absolute numbers may not be completely precise.

Another limitation of the study is that the data was gathered during a university course with students. However, these students were mid-career professionals with substantial work experience. So, although these findings are based on a study of students, the students in question were not novice designers.

Furthermore, the time scale of the course may introduce error into the results, as the optimal design process in such a time constrained setting may be different from other real-world situations. In the context of a semester long design course, it may be advantageous to spend less time on ideation and exploring the design space, and place a higher than normal emphasis on the execution (fabrication quality) of the final prototype. However, statistical analysis on the data did not find any correlations between total fabrication (building) time and better design success.

Regardless, there may be important differences between design in the context of the design course in question and real-world situations, and further research is required to validate the findings of this paper.

CONCLUSIONS AND FUTURE WORK

Our original research questions spoke to the timing of prototyping as well as other design activities in early stage process. Below, we summarize our findings as responses to these initial questions, along with implications for design education and practice:

1. Did the timing of the prototypes, in other words when the prototypes were built, matter with respect to design outcome? And did teams that spent more time prototyping at various points in the design cycle fare better than those teams that spent less time?

In this study, the correlation between prototyping early (building rough exploratory prototypes) and design success was highly statistically significant. There were no statistically significant correlations between the total time a team put into building prototypes and mock-ups, but there were highly statistically significant correlations between prototyping *early* and design outcome. For both students and practitioners, the message is to *prototype cheaply and early in a project.*

2. Did the timing of other design activities, such as sketching, user interaction, benchmarking, and presentation preparation correlate with better design performance?

Several correlations were found with other activities and design outcome, with an emphasis on closed-loop design of artifacts that allow teams to compare their early designs in the context of the end user and other existing products.

It was found that going back and talking with the end-users *after* having chosen a final concept correlated with better design outcome. Successful teams sought input while developing their final concept, whereas less successful teams seemed to solicit input to help *choose* their final concept, but did not engage the end-user *after* the choice had been made to a similar extent as the more successful teams did. The take home message is: *go back and talk to the end-user*.

Similarly to user interaction, *after* choosing their final design concept, top tier teams went back for another round of benchmarking to support their design iteration of their concept, whereas bottom tier teams spent very little effort on benchmarking after having chosen their final design concept. For design teams, then, it is important to *compare your final concept with existing products*.

Perhaps somewhat surprisingly, there was a correlation between spending time working on presentations and the final evaluation of the teams' design. This suggests the importance of formulating and communicating a coherent message about a design to others. In other words, teams should *have a message about their design*.

Future work. The results of this paper suggest the importance of timing of several activities in early stage design, and further point out the importance of closing the loop between early stage design activities and the elicitation and incorporation of feedback in the process. More and more, design education and practice now emphasize generative design activities such as brainstorming and creativity, but this work suggests the value of user feedback and product contextualization in shaping how a design should move forward. Future work should consider how such evaluative behavior can most effectively be incorporated in both design curriculum as well as design practice.

ACKNOWLEDGMENTS

The work described in this paper was supported in part by the National Science Foundation under Award CMMI-1130791 and the Finnish Foundation for Technology Promotion and the Finnish Cultural Foundation. The opinions, findings, conclusions and recommendations expressed are those of the authors and do not necessarily reflect the views of the sponsors.

REFERENCES

[1] Cooper, R. G., and Kleinschmidt, E. J., 2010, Wiley International Encyclopedia of Marketing, John Wiley & Sons, Ltd, Success Factors for New-Product Development.
[2] Simpson, T. W., Rosen, D., Allen, J. K., and Mistree, F., 1998, "Metrics for Assessing Design Freedom and Information Certainty in the Early Stages of Design," ASME Journal of Mechanical Design, 120(4), pp. 628–635.
[3] Thomke, S., and Fujimoto, T., 2000, "The Effect of "Front-Loading" Problem-Solving on Product Development Performance," Journal of Product Innovation Management, 17(2), pp. 128-142.
[4] Kelley, T., and Littman, J., 2001, *The Art of Innovation: Lessons in Creativity from Ideo, America's Leading Design Firm*, Doubleday, New York, NY.
[5] Ward, A., Liker, J.K., Cristiano, J.J., and Sobek, D.K., 1995, "The Second Toyota Paradox: How Delaying Decisions Can Make Better Cars Faster," MIT Sloan Management Review, 36(3), pp. 43-61.
[6] Gerber, E., 2009, "Prototyping: Facing Uncertainty through Small Wins," International Conference on Engineering Design, Stanford, CA, USA
[7] Houde, S., and Hill, C., 1997, Handbook of Human-Computer Interaction, Elsevier Science, Amsterdam, What Do Prototypes Prototype?
[8] Buxton, B., 2006, "What Sketches (and Prototypes) Are and Are Not.," CHI 2006 One-Day Workshop on "Sketching" Nurturing Creativity: Commonalities in Art, Design, Engineering and Research, Montreal, Canada
[9] Sommerville, I., 1995, *Software Engineering*, Addison-Wesley, Wokingham, England.
[10] Goel, V., 1995, *Sketches of Thought*, MIT Press, Cambridge, Mass.
[11] Ferguson, E. S., 1992, *Engineering and the Mind's Eye*, The MIT Press, Cambridge, MA.
[12] Osborn, A. F., 1963, *Applied Imagination*, Charles Scribner and Sons, New York, NY.
[13] Ulrich, K. T., and Eppinger, S. D., 1995, *Product Design and Development*, McGraw-Hill, Inc., New York.
[14] Kolodner, J. L., and Wills, L. M., 1996, "Powers of Observation in Creative Design," Design Studies, 17(4), pp. 385-416.
[15] Schrage, M., and Peters, T., 1999, *Serious Play: How the World's Best Companies Simulate to Innovate*, Harvard Business School Press, Boston, MA.
[16] Yang, M. C., 2005, "A Study of Prototypes, Design Activity, and Design Outcome," Design Studies, 26(6), pp. 649-669.
[17] Kelley, T., 2001, "Prototyping Is the Shorthand of Innovation," Design Management Journal (Former Series), 12(3), pp. 35-42.
[18] Baxter, M., 1995, *Product Design: Practical Methods for Systematic Development of New Products*, CRC Press, USA.
[19] Viswanathan, V., and Linsey, J., 2009, "Enhancing Student Innovation: Physical Models in the Idea Generation Process," IEEE Frontiers in Education Conference, San Antonio, TX
[20] Jansson, D. G., and Smith, S. M., 1991, "Design Fixation," Design Studies, 12(1), pp. 3-11.
[21] Viswanathan, V. K., and Linsey, J. S., 2012, "Physical Models and Design Thinking: A Study of Functionality, Novelty and Variety of Ideas," Journal of Mechanical Design, 134(9), pp. 091004-13.
[22] Linsey, J., Tseng, I., Fu, K., Cagan, J., Wood, K., and Schunn, C., "A Study of Design Fixation, Its Mitigation and Perception in Engineering Design Faculty," Journal of Mechanical Design, 132(pp. 041003.
[23] Dijk, L., Vergeest, J. S. M., and Horváth, I., 1998, "Testing Shape Manipulation Tools Using Abstract Prototypes," Design Studies, 19(2), pp. 187-201.
[24] Yang, M. C., 2009, "Observations of Concept Generation and Sketching in Engineering Design Projects," Research in Engineering Design, 20(1), pp. 1-11.

[25] Lai, J., Honda, T., and Yang, M. C., 2009, "A Study of the Role of User-Centered Design Methods in Design Team Projects," AI EDAM: Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 24(3), pp. 303-316.
[26] Hannah, R., Joshi, S., and Summers, J. D., 2012, "A User Study of Interpretability of Engineering Design Representations," Journal of Engineering Design, 23(6), pp. 443-468.
[27] Elsen, C., Häggman, A., Honda, T., and Yang, M. C., 2012, "Representation in Early Stage Design: An Analysis of the Influence of Sketching and Prototyping in Design Projects," ASME International Design Engineering Technical Conferences, Chicago, IL
[28] Van Newenhizen, J., 1992, "The Borda method is most likely to respect the Condorcet principle," Econ. Theory 2, pp. 69-83.
[29] Purcell, A. T., and Gero, J. S., 1998, "Drawings and the Design Process," Design Studies, 19(4), pp. 389-430.
[30] Suwa, M., and Tversky, B., 1997, "What Do Architects and Students Perceive in Their Design Sketches? A Protocol Analysis.," Design Studies, 18(4), pp. 385-403.