Final Year Project
Dissertation

An analysis of the language used in early new product design sessions that contributes towards turning ideas into detailed design concepts used in solutions

Isaac Anderson
Mechanical Engineering with Advanced Design and Innovation
University of Bath

Supervisor	Assessor	Personal Tutor
Dr. Elies Dekoninck	*Mr. Martin Ould*	*Dr. Marcelle McManus*

SUMMARY

How are concepts *optimally developed* from initial ideas?

By following three undergraduate engineering teams for five weeks during the early design stages of their group business and design projects, a greater understanding of the language used in design sessions and its role in developing ideas into concepts was gathered. The results gathered were split into several sections.

The first found an increase of 7% in “block frequency” between normal and critical episodes (critical episodes discussing ideas included in the final design), with low overall rates of “support for blocks”. The second stage found while that criticism was rarely given in the early stages of NPD, standard criticism was given 164% more frequently than constructive, with increases of 8% and 20% in critical episodes respectively. Questioning rates were much higher, with the most popular type being those that helped understanding. Adding to this, questions that required a higher level of cognitive thought or critical thinking to answer were more common in groups with leaders who had industrial experience in NPD (six out of ten against five out of ten), with a dramatic increase in analytical questions (+55%) being observed, as well as a fall in understanding questions asked (-13%). Part three saw that advancement of an idea was preferred instead of no advancement over three times as much for criticism, and over five times for questioning, but with no significant difference in normal and critical episodes. Saying this, chi-squared tests found the majority of raw percentage increases in this investigation to be minor, with only the relationship between industrial experience and the level of questioning significant.

Qualitative data findings were also presented in this study. Interaction dynamics notation found that while it was rare for a block to be supported, support for the answer was common in all groups. It was also discovered that most advancement did not arise due to questioning and criticism, but through statements and external feedback, and that one team member did the majority of criticism in all the teams. An incidental finding also discovered that, like the majority of research in this area shows, individual brainstorming led to a higher amount of ideas than group brainstorming.

However, as the majority of percentage changes gathered were proven to be insignificant by chi-squared tests, no rules for concept development can be recommended at this time.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

The research presented in this dissertation would not have been possible without the help of several people. I would like to thank;

- **Elies Dekoninck**, for initiating the project, as well as her support during its duration,

- **Martin Ould**, for his support, as well as giving a different perspective on the project having worked closely with third year group business design project students in the past,

- **Gillian Ellsworth**, for providing a physical location for the teams to work in, often at short notice,

- And most importantly, to **all the members of the three teams** that I have followed, for allowing me to sit in and document their sessions, as well as discuss them in this paper. Without them, this research would not have been possible.

1. INTRODUCTION

"It's the disease of thinking that getting a really great idea is 90% of the work....The problem with that is there's a tremendous amount of craftsmanship between a great idea and a great product." – ***Steve Jobs***

"Ideas are worth nothing unless executed. They are just a multiplier. Execution is worth millions." – ***Derek Sivers***

In new product design (NPD), one of the first stages undertaken when coming up with a solution is the concept development stage (also known as the conceptual design phase, **Fig.1**). Started after the task has been clarified through a product specification, this typically takes the ideas generated in a preliminary brainstorm and grows them into complete concepts. Logical reasoning suggests that a larger amount of initial ideas would lead to a larger amount of concepts; as a result of this, most research investigates brainstorming and idea generation, as well as techniques that increase both the quality and quantity of these initial ideas.

But how are ideas crafted into detailed design concepts? Or rather, how are concepts *optimally developed* from initial ideas?

Answering this question could open up a fresh stream of research in NPD. By moving focus from idea generation to concept development, a deeper understanding of the behaviour(s) that turn an idea into a fully-fledged concept can be gained. If certain behaviours or activities are found to be more effective in developing a quality idea than others, as well as being repeatable and relatively easy to implement, it is possible that new "ground rules" for concept development can be developed in order to make it as efficient as possible. When used in conjunction with some of the concept development techniques currently recommended, these rules could optimise this stage of the design process by increasing concept quality or reducing development time.

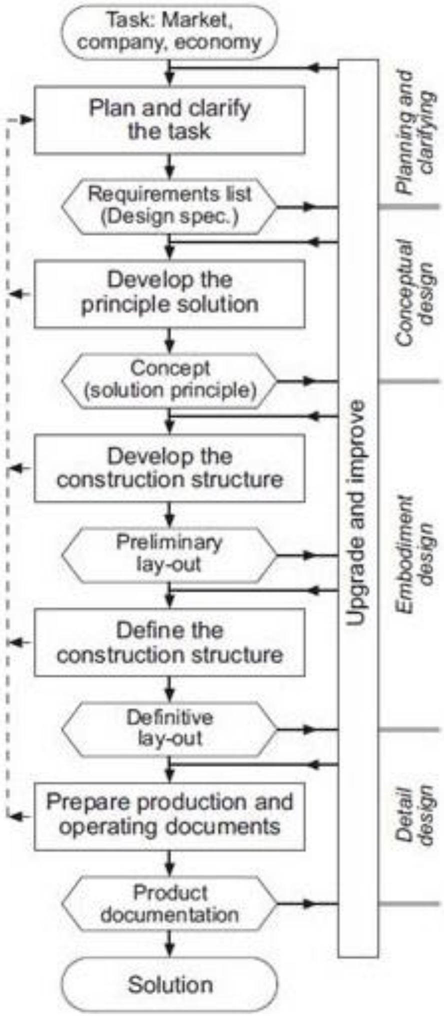

Figure 1: *The Pahl and Beitz design model, adopted widely internationally and general enough to be applied to most strands of engineering.*

The aim of this research is to analyse the language used in early NPD sessions that contributes towards turning ideas into detailed design concepts used in solutions. By focusing on the criticism and questioning that arises during these sessions specifically, as well as which types are effective in developing a quality idea, any relationship between these variables can be found. As mentioned earlier, if the findings of this research are conclusive enough, it is possible that rules for concept development can be created in order to make it as efficient as possible.

To gather data, three undergraduate engineering teams were followed for five weeks during the early design stages (mainly planning and conceptual, **Fig.1**) of their group business and design projects. At the end of this period, the teams had also completed mandatory technical proposals on their initial concepts, with the strongest being chosen to develop further. By observing and reviewing these meetings, the final design proposed by each team at the end of their projects was "reverse engineered" – allowing the key exchanges that take place while developing these ideas to be identified and evaluated.

By the end of the project;

- The interactions of each group will be **analysed** using a conversational analysis coding technique which can reveal behavioural trends of a group,

- Critical exchanges which result in features ending up in the final design will be **assessed** and **evaluated**,

- A relationship between the types of criticism/questioning and their effectivity in developing a quality idea will be **identified** if existent**,**

- Rules for concept development, to be used alongside concept development techniques, will be **generated** if analysis shows certain behaviours to be consistently effective.

Initially, this project will look at two particular behaviours common in concept development sessions; criticism and questioning. Appropriate literature on these, as well as on other relevant topics are covered in **section 2**.

2. LITERATURE REVIEW

The idea of creating a standard set of "rules" to aid at a certain stage of the design process is not a new idea. Brainstorming, an idea generation process usually undertaken in groups, has four standard rules that should be used to maximise the productivity of a session;

- No criticism of ideas is allowed.
- Freewheeling and free association is encouraged.
- Quantity is required more than quality.
- Building on ideas is encouraged (Osborn, 1953).

Osborn proved that by following these rules, the quantity of ideas generated during a brainstorming session increased when compared to a non-constrained team (+44%). However, the *quality* of the ideas generated was not measured; as an advertising executive, Osborn rationalised that a greater amount of ideas would increase the likelihood of creating a suitable solution as "it is easier to tone down a wild idea than to think up a new one". The effectiveness of his rules was confirmed in 1959 in a study that considered both the quantity *and* quality of the ideas generated. It also found;

- When used by a group of subjects trained in creative problem solving, a team that used Osborn's rules produced a "significant increment" of good quality ideas when compared to a team that did not.
- With a group of trained subjects, Osborn's rules produced a "significantly greater absolute number" of good ideas compared to a team that were not.
- Using Osborn's rules instead of a "produce good ideas, penalty for bad ideas" method "improved team performance" (Parnes & Meadow, 1959).

Brainstorming is still widely used today in a variety of different industries; in NPD it is still typically the first action taken after a task has been set. The ideas generated are then usually evaluated, with the best being further developed into complete concepts. Osborn's rules are still also commonly practiced to improve the productivity of a session. Indeed, his initial list of four rules has been expanded by some over the years; to eight (Thompson, 2013) and even ten (Carter & Wynn, 2014).

But is there a *particular* rule that is more important than the others? Is this rule *applicable* to concept development? Are its positive effects *replicable*?

2.1. CRITICISM AND ITS EFFECTS

The "no criticism" rule is by far the most tested and discussed out of Osborn's original four. Osborn himself determined this to be the most important, going as far as to state that "[only if] the deferment of judgement principle is strictly followed" was a brainstorming session valid (Osborn, 1953). However, further investigation has found this rule to be both beneficial *and* detrimental to the idea generation process.

The main body of research supporting the "no criticism" rule comes from researchers investigating whether brainstorming individually was more productive than in a group. Osborn claimed that groups could create twice as many ideas as individuals due to psychologists proving that "competition will increase accomplishment in mental work by adults by 50% or more" (Osborn, 1953). However, the psychological studies he cited were relatively outdated (from 1897); and when his claims were tested by various researchers testing different industries it was found that the opposite was true – larger quantities and higher qualities of ideas were produced individually (**Fig.2**). In regards to NPD, Lewis discovered that groups of design engineering professionals were less effective at solving realistic engineering problems when compared to the collective efforts of individuals (Lewis, et al., 1975). Like most researchers in this field, he attributed the lack of criticism during individual brainstorming sessions as a major contributor to this result.

However, these studies did not *directly* test whether criticism was the reason for nominal groups being more productive than real groups. When Diehl and Strobe investigated, they began by analysing the results of twenty-one previous studies comparing teams that brainstormed collectively with teams that brainstormed individually (**Fig.2**), before focusing on identifying *what* caused the productivity loss in groups. Three factors were hypothesised; production blocking (caused by individuals being unable to express their ideas), free riding (individuals not contributing as they believe other members will instead) and evaluation apprehension (individuals being reluctant to share their suggestions due to fear of criticism). While production blocking was found to be the main cause, evaluation apprehension was also found to "significantly reduce" the productivity of an individual. Furthermore, the effects of evaluation apprehension were found to increase as the size of a group got larger, due to bigger numbers of people increasing the probability of criticism (Diehl & Stroebe, 1987; Gallupe, et al., 1992). However, there were limitations to this study. The major one is that while four internal experiments were carried out to test Diehl and Strobe's hypotheses, the composition of each experiments groups was different (48 male high school students for one experiment, 36 male psychology undergraduates for another, and 60 mixed gender undergraduate psychology students for the others). Adding to this, their results directly contradicted with another paper that did not find correlation between productivity and evaluation apprehension (Maginn & Harris, 1980).

Results of Studies That Compare Brainstorming Productivity of Real Groups (R) With Productivity of Nominal Groups (N)

Study	Group size	Productivity: Quantity	Productivity: Quality
Taylor, Berry, & Block (1958)	4	$R < N$	TQ:$R < N$ AQ:equivocal[a] NO:$R < N$
Cohen, Whitmyre, & Funk (1960)	2	$R = N$	NO:equivocal
Dunette, Campbell, & Jaastad (1963)	4	$R < N$	TQ:$R < N$ AQ:equivocal
Milton (1965)	4	$R < N$	TQ:$R < N$
Gurman (1968)	3	$R < N$	TQ:$R < N$
Bouchard (1969) Experiment 2	4	$R < N$	TQ:$R < N$ AQ:equivocal NG:$R < N$
Rotter & Portugal (1969)	4	$R < N$	—
Vroom, Grant, & Cotton (1969)	4	$R < N$	TQ:$R < N$ AQ:equivocal NG:$R < N$
Bouchard & Hare (1970)	5, 7, 9	$R < N$	—
Torrance (1970)			
Experiment 1	2	$R = N$	NO:$R < N$
Experiment 2	2	$R = N$	NO:$R < N$
Dillon, Graham, & Aidells (1972)	4	$R < N$	—
Bouchard (1972)			
Experiment 2	4	$R < N$	NG:equivocal
Bouchard, Drauden, & Barsaloux (1974)			
Conditions E and F	4	$R < N$	—
Street (1974)	3	$R < N$	—
Harari & Graham (1975)	4	$R < N$	—
Chatterjea & Mitra (1976)	3	$R < N$	—
Madsen & Finger (1978)	4	$R < N$	—
Maginn & Harris (1980)	4	$R < N$	—
Jablin (1981)	4	$R < N$	—
Barkowski, Lamm, & Schwinger (1982)	2	$R < N$	—
Pape & Bölle (1984)	2	$R = N$	NO:$R = N$

Note. The following quality measures were used: total quality (TQ), average quality (AQ), number of original or unique ideas (NO), and number of good ideas (NG). Dashes indicate that quality was not assessed.
[a] Findings vary across different topics, subjects groups, or experimental conditions.

Figure 2: *A summary of the results found in twenty one studies comparing teams that brainstormed collectively, known as real groups, with teams that brainstormed individually, known as nominal groups (Diehl & Stroebe, 1987).*

When researchers have *directly* tested the "no criticism" rule, it is often found to be detrimental to idea generation. In 2004, Nemeth compared the productivity of a session in which the group participants were not allowed to criticise against a group in which the participants were actively encouraged to "debate", and a control group given no specific instructions (Nemeth, et al., 2004). She found that the "debate" group generated more ideas (+16% compared to traditional brainstorming). This research was broad in scope; not only did it investigate cross-cultural differences by testing French and American students, but it tested large numbers of groups (52 groups of five females were used in the French tests, 39 groups of five same-sex students for the American). It also agreed with studies that found that critical feedback improves performance on a wide range of tasks (Landy, et al., 1982; Pritchard, et al., 1988). However, like the majority of researchers in this area, she used psychology students; and in regards to NPD, it was unclear whether these results were transferrable across disciplines.

Linley replicated Nemeth's results using engineering students (Linley, 2013). However, he also investigated whether the quality of ideas increased using Osborn's rules, finding that they both did (quantity increase of 15%, quality increase of 20%). One of the key parts of this research was its use of interaction dynamics notation to identify blocks (see Sonalkar's work, **section 2.3**). By doing this, the *types* of criticism present during a session were revealed, which were;

- Negativity (highlighting an issue in an inconsiderate manner, with the aim of "killing" an idea altogether). This research will refer to negativity as destructive criticism.

- Standard criticism (simply highlighting an issue). Linley found this was the most common type of criticism.

- Constructive criticism (highlighting an issue in a considerate manner that aimed to be as useful as possible, which often included offering potential solutions to issues with ideas). Linley found this was the most effective in increasing idea quantity.

However, several limitations existed in Linley's work. While he used engineering students, they were volunteers and therefore would be unlikely to have the same drive as employees or students who are being judged on the quality of their work. Adding to this, the same group of participants was used to brainstorm various briefs, with Linley noting that the teams "brainstormed more efficiently" as the study progressed. Saying this, the results were considered to be reliable enough, as they were used in Dekoninck's 2016 paper on the same topic (Dekoninck, et al., 2016).

Several key findings have been identified from reviewing the research into the extremities of criticism (constructive and destructive) and their effects. The first is that recipients prefer constructive criticism not just because it is more considerate, but because it is usually specific (Liden & Mitchell, 1985). Other findings came from Baron; his 1988 study into the negative effects of destructive criticism found that, when compared to individuals who received either constructive criticism or no feedback at all, individuals who received this were more likely to have "lower goals, lower levels of self-efficacy and were less likely to handle future disagreements through collaboration or compromise" (Baron, 1988). In a follow-up study, he also discovered that the critic perceived their feedback to have a smaller impact on motivation and working relationships compared to the recipient (Baron, 1990). This indicated that the critic does not realise the effect of their judgement on others, and when combined with other research, led Baron to conclude that destructive criticism "may adversely affect performance". The main weakness of Baron's research is that he used undergraduates. As this group is relatively inexperienced in the workplace, they might not deal with criticism as well as others. Furthermore, a lack of soft skills developed in industry within this demographic (Schulz, 2008) could go some way to explaining why critics did not realise the effects of their words. However, in the context of this dissertation, this is a strength as the natural expectation will be for the groups to act in a similar way to what Baron observed.

Finally, Kluger and DeNisi aimed to go into greater detail about the types of feedback (including criticism), as well as the best circumstances in which to give it. Using nearly 13,000 participants, their most important findings were that feedback was most useful when;

- Task feedback about changes made from previous tasks was given,

- Task complexity was high,

- Individual goals were difficult to achieve (Kluger & DeNisi, 1996).

What makes this study so important is that it used Cohen's effect size to numerically illustrate how significant an effect was (**Fig.3**). Using Cohen's rules (which state that the influence of a variable increases as the effect size grows larger), the importance of each specific type of feedback was able to be analysed (Cohen, 1992). Furthermore, it was discovered that a third of feedback given actually decreased performance; heavily indicating that the style and situation criticism is given in is important to consider.

Moderator	Number of effects	Effect size
Correct feedback		
'Tis correct	114	0.43
'Tis incorrect	197	0.25
Task feedback about changes from previous trials		
Yes	50	0.55
No	380	0.28
Task feedback designed to discourage the student		
Yes	49	–0.14
No	388	0.33
Praise feedback about the task		
Yes	80	0.09
No	358	0.34
Feedback provided from a computer		
Yes	87	0.41
No	337	0.23
Number of times feedback was provided		
Lots	97	0.32
Little	171	0.39
Task complexity		
Very complex	107	0.03
Not complex	114	0.55
Goal setting		
Difficult goals	37	0.51
Easy, do your best goals	373	0.30
Threat to self-esteem		
Much threat	102	0.08
Little threat	170	0.47

Figure 3: *A summary of effect sizes relating to feedback type (Kluger & DeNisi, 1996). Larger effect size numbers indicate that the variable tested has a greater influence.*

2.2.THE TYPES OF QUESTIONING

Reviewing the group transcripts in Linley's research also exposed the role of questioning in NPD. Participants often gave feedback in the form of questions, which in turn developed the idea being discussed at the time. Individuals tended to ask two types of question; open (which must be answered in detail) and closed (which can be answered with a simple "yes" or "no"). In reality however, questioning is much more complex than this.

Perhaps the most complete model of questioning is Bloom's taxonomy (Bloom, 1956; Krathwohl, 2002). Originally developed as a sequence of educational objectives for lesson planning in the classroom, this model states that questions have six objectives. **Table 2-1** gives an updated version of Bloom's taxonomy (Krathwohl, 2002).

Table 2-1: *The six types of questions in Bloom's updated taxonomy, with definitions and examples.*

Question Classification	Definition	Examples
Remembering (Level 1)	*Questions that require the person answering to recall information they have learned.*	*What do you remember our client saying about the table?*
		Which machine could do the stockinette stitch?
Understanding (Level 2)	*Questions that require the person answering to summarise their idea in their own words.*	*How would the animal be secured to the table?*
		Where is the bamboo culm cut?
Applying (Level 3)	*Questions that require the person answering to define and solve problems related to their idea.*	*How can the epoxy tubes be inlaid into the design automatically?*
		What would happen if the bamboo size changed?

Analysing (Level 4)	*Questions that require the person answering to explain why they have chosen a certain idea and/or why their solution works.*	*Why do you think six degrees of freedom is the best solution?*
		What is the problem with making the cast hardening process irreversible?
Evaluating (Level 5)	*Questions that require the person answering to judge and critique their own work against external standards/criteria.*	*What is the most important feature of this table that our client will like?*
		Does this bamboo size concept match the specification requirements?
Creating (Level 6)	*Questions that require the person answering to combine ideas to build a new idea for either a new or existing situation.*	*What changes can you make to the layout of the cast?*
		How would you improve the processing time?

Bloom's taxonomy is typically used with the six W's (**w**ho, **w**hat, **w**hen, **w**here, **w**hy and ho**w**) to form probing questions. As some require more thought to answer than others, the taxonomy ranges from lower levels of cognitive thinking (i.e. remembering or understanding) to higher levels (i.e. evaluating and creating). In relation to NPD, questions that require higher levels of cognitive thinking are preferred and even encouraged, as the differing levels are designed to help develop critical thinking skills (Bloom, 1956; Christenbury & Kelly, 1983). This is the main disadvantage of this model, as Bloom designed the taxonomy to be progressively used (starting from level 1 and moving up) – but if an idea is sufficiently explained in a real-life NPD setting, the lower levels will often be skipped as they are not needed. Furthermore, while it is often believed that the question level determines the response level, Dillon discovered this this was true only about 50% of the time (Dillon, 1982). However, it must also be noted that the taxonomy was originally developed for children learning in the classroom. Generally, youngsters do not have the critical thinking skills of an adult.

Unlike standard questions, rhetorical and leading questions are often asked when an individual is trying to form an argument. Rhetorical questions (asked in order to emphasise a point rather than to get an answer) were less frequently asked than leading according to Linley's transcripts, with Blankenship and Craig finding that they increase an individual's resistance to a particular point being made (Blankenship & Craig, 2006). Therefore, during concept development when group consensus needs to be reached, these should be avoided. Leading questions (which encourage the recipient to give an answer that the questioner wants) have the ability to change an individual's opinion, and can even distort facts (Loftus & Palmer, 1974; Loftus & Zanni, 1975); with attitudes leaning towards "no" based leading questions being more effective. Edinger argues that asking "yes" based leading questions is counterproductive as it raises defences and scepticism. Voss agrees, arguing questions that produce a "no" response, while shaping the direction of the conversation, allow people to relax as they feel a sense of control (Voss & Raz, 2016; Edinger, 2017). While leading questions are less likely to antagonise individuals when compared to rhetorical, it can also be advised that these types of question are detrimental during concept development, as they have the ability to skew thoughts in one direction.

2.3. OTHER RELEVANT LITERATURE

Research conducted by Sonalkar was different to others in this field, as it studied *concept development* instead of *idea generation*, by looking at the *process* of getting to a final result. Unlike the majority of previous studies (which evaluated the effect of changing a variable), his research focused on communication between team members with the aim of finding replicable behaviour. To do this, Sonalkar created *Interaction Dynamics Notation* (IDN), a conversational analysis tool which visually represented dialogue using a series of standard responses (**Fig.4**). Then, using two teams consisting of three engineering graduates, he used IDN to analyse how design concepts were generated.

He found that;

- In concept generation interactions, not all expressions were idea expressions. These other expressions were usually either short stories, statements of general facts or statements of future uncertainty.

- Blocking (an obstruction to the content of the previous move, also see **Appendix 9.1**) usually came in a sequence, and was sustained over a period of time. However, this was not always detrimental. Instead, these blocks were resolved by the team, which often led to further development. Also, the blocks were usually resolved by the participation of the person whose original response was blocked (Sonalkar, et al., 2012).

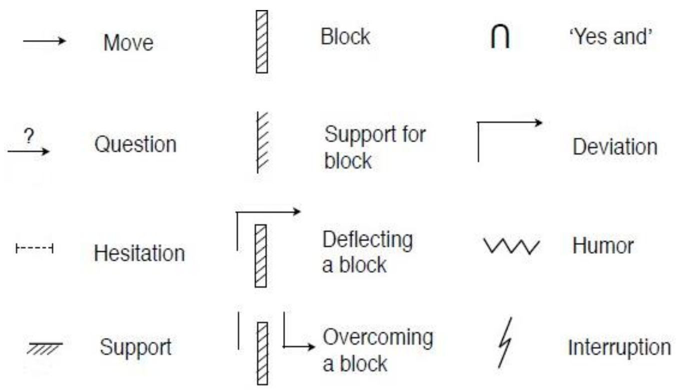

Figure 4: *The twelve symbols that make up Sonalkar's Interaction Dynamics Notation. A detailed version of this (including descriptions and examples) is given in Appendix 9.1.*

Both criticism and questioning act as blocks, as they "obstruct the flow", and overcoming them is crucial for further developing ideas and concepts. However, continual blocking can also have other effects; one of which is significant design innovation. In 1997, Nigel Cross proposed the "creative leap"; a moment of sudden illumination in a team's design process that leads to a quick rate of concept development in a short space of time, having observed this behaviour in a small team of three designers (Cross, 1997). In the minute following this leap, the team accepted, modified, developed *and* justified an idea (eventually culminating with it being developed into a fully-fledged design).

The creative leap episode[1] in Cross's research is coded using IDN in **Table 2-2**.

Table 2-2: *The minute long "creative leap" conversation in Cross's research, coded using IDN. Blocks are highlighted in red.*

Group Member	Comment	Symbol
I	Bag; put it in a bag; we're gonna need some sort of thing to do something with those straps	
K	To get this out of the way	
J	So it's either a bag, or maybe it's like a little vacuum formed tray kinda for it to sit in?	
I	Yeah, a tray, that's right, OK	

[1] This research will refer to an "episode" as an interaction that starts with a block and end with the block either being overcome or the team moving onto a new topic. Any further blocks raised while the initial episode is still unresolved will not start a new episode. Adding to this, while IDN has questioning as a symbol, questions that are asked during episodes will be referred to as blocks, as they will have the same effect (hindering the content of the previous move).

J	It would be nice, I mean just from a positioning standpoint, if we've got this (backpack) frame outline and we know that they're gonna stick with that, you can vacuum-form a tray	
I	Right, or even just a small part of the tray...	
K	Something to dress this (straps) in?	
J	Maybe the tray could have plastic snap features in it, so you just like snap your backpack down into it?	
K	Snap in these (backpack) rails	
J	It's a multi-function part	
K	You just snap in these rails	
J	Yeah, snap the rails into the tray there	
I	OK	
J	It takes care of the rooster-tail problem ...	

Aiming to prove that critical thinking is considered as one of the most important skills that an engineer can possess, Nguyen created a study of academics, students and industrial personnel to determine the essential skills of an engineer. She found that logical thinking and problem solving skills were needed by all engineers, something critical thinking is said to increase. These findings were supported by Reed who, by surveying engineering professionals, found that over 98% of her respondents thought that engineers need critical thinking skills (Nguyen, 1998; Reed, 2010). It is therefore surprising to consider that the majority of questions asked by instructors and team members are "low-level" (level 3 and under), with 68.9% of team members and 91.2% of instructors questions asked being low-level (Sellappah, et al., 1998; Saeed, et al., 2012). Phillips and Duke also observed that low-level questions were asked more frequently, but found that when comparing full-time faculty members and volunteer teachers, the faculty members asked a higher level of questions (Phillips & Duke, 2001).

The use of language could also depend on the dynamics of the actual teams being followed. Tuckman's stages of group development model proposes that there are four phases of group development (although he later added a fifth). These were;

- Forming – the stage where the teams have met for the first time, with group members being positive and restrained in critique, as well as relying on their leader for guidance and direction,

- Storming – the stage where decisions do not come easily, disagreements start to emerge and cliques possibly form,

- Norming – the stage where agreement and consensus largely forms among the team, and every individual knows their role,

- Performing – the stage where the team has a shared vision and are working efficiently and effectively, with each member knowing what they are doing and why (Tuckman, 1965).

Tuckman's model is the direct inspiration for the hypothesis made in this research, which will be revealed in **section 3**.

The final piece of literature reviewed relates to anchoring, a subconscious cognitive bias that occurs when the first piece of information given (the anchor) heavily influences the range and diversity of an individual's thoughts and ideas going forwards (Tversky & Kahneman, 1974). One of the three biases of cognition (alongside framing and decision fatigue), it is considered a key aspect of negotiation. However, Thompson argues that this effect is applicable to brainstorming also, and that only by using a different idea generation technique can this bias be overcome (Thompson, 2013). Her hunch had already been indirectly confirmed in 2010 by Howard, who at the time was investigating when vital creative ideas were generated during a brainstorming session. Having observed the sessions of five teams, he found that over half of all appropriate ideas were generated in the first ten minutes of a session, and that one team produced all their suitable ideas in seven (Howard, et al., 2010). In the context of NPD, the effects of anchoring would see the earliest ideas having a larger probability of being incorporated into the final design, as these ideas would act as the anchor that all future ideas wold be compared to. It is also a contributing factor to why only early sessions are being observed.

3. METHODOLOGY

Using the findings of the literature review on the differing types of criticism and questioning and their effects, the initial research questions that this study will explore have been formed;

- **Is there a relationship between the type of criticism and their effectivity in developing ideas in early NPD sessions?**

- **Is there a relationship between the type of questioning and their effectivity in developing ideas in early NPD sessions?**

The types of criticism tested will be the three main categories studied in the literature review – destructive, constructive and standard. Closed questions (including rhetorical and leading) will be treated as standard criticism, as these are usually not asked to develop an idea, but rather to help form or back up an argument being made. Effective criticism will be defined as criticism that helps *advance* the idea (through the evolution of a previous idea or a new suggestion that addresses the issue(s) raised), whereas non-effective criticism will result in no advancement. Likewise, the types of questioning tested will be the six categories that make Bloom's taxonomy (remembering questions, understanding questions, application questions, analytical questions, evaluative questions and creation questions). Effective questioning will be defined as questioning that helps *advance* the idea (through effectively using the six W's and Bloom's taxonomy to help improve the overall idea), whereas non-effective criticism will result in no advancement of an idea. No advancement is generally not wanted as a response, due to the fact that it does not *solve* the issue raised but instead defers or ignores it.

In this research, both criticism and questioning are defined as blocks, as they "obstruct the flow". However, Tuckman's stages of group development model suggests that the frequency of these could vary depending on the stage each group is at. As this research is following the early stages of group development (forming, with storming possibly observed if the teams "gel" quickly), it is possible that the levels of criticism could be lower due to the teams being unfamiliar with each other, and wanting to remain as positive as they can in these early stages. This has led to the following hypothesis;

- **Low levels of criticism will be observed during this project, due to the teams being in the "forming" stage of group development.**

To conduct this investigation, three undergraduate engineering teams will be followed for five weeks during the early design stages (mainly conceptual and embodiment) of their group business and design projects. Each team consists of six individuals who encompass a range of different engineering disciplines, although they had all been made aware of Osborn's brainstorming rules (including the "no criticism" rule) before beginning their projects.

Preliminary analysis on the problem solving style of the year group that the teams are in has found;

- Innovation may be difficult as individuals who do things differently, challenge the problem definition and emphasize originality are outnumbered nine to one. It is therefore possible that ideas with a higher level of inventiveness will be placed under a higher level of scrutiny.

- 80% of individuals are task-focused, meaning that put they put the quality of the outcome over people's feelings. Therefore, criticism giving by these individuals might be harsher as it is given without considering the effect it has on people (agreeing with the findings of Baron in 1990 and Schulz in 2008).

The tasks that the three teams followed in this project were working on, as well as the composition of the teams, are shown in **Table 3-1**.

Table 3-1: *A summary of the three teams observed during for this research.*

Group	A	B	C
Brief	*To design an orthopaedic wrist cast manufactured using knitting technology.*	*To design a mobile machine that allows farmers to easily process bamboo onsite.*	*To design a semi-autonomous veterinary operating theatre (focusing on the operating table).*
Team composition	***Six*** *males studying mechanical engineering;* ***two*** *with management and* ***one*** *with design.*	***Five*** *males and* ***one*** *female studying mechanical engineering;* ***one*** *with management and* ***one*** *with design.*	***Five*** *males and* ***one*** *female studying mechanical engineering;* ***two*** *with management,* ***one*** *with design and* ***one*** *with electrical engineering.*
Team industrial experience	***Four****, but* ***none in NPD.***	***Five****, with* ***one in NPD****.*	***Four****, with* ***one in NPD****.*
Design phase leader	*Mechanical engineering student. Industrial experience, but* ***none in NPD****.*	*Design engineering student.* ***Industrial experience in NPD****.*	*Design engineering student.* ***Industrial experience in NPD****.*
Project manager	*The same mechanical engineering student.*	*Mechanical engineering student. Industrial experience, but* ***none in NPD****.*	*The same design engineering student.*
Total meeting time	*15 hours*	*12 hours*	*13 hours*

This project has four aims.

- **To analyse the interactions of each group using a conversational analysis coding technique which can reveal behavioural trends of a group.**

  During sessions where the teams develop their product, audio and/or video recordings will be taken. These recordings will then be transcribed, before being coded using IDN. Exchanges called "episodes" will then be analysed; these are interactions that start with a block and end with the block either being overcome or the team moving onto a new topic. As mentioned earlier, criticism and questioning that occur during an episode will be treated as blocks. From here, the behavioural trends of a group (such as blocking frequency and support for blocks) can be identified and analysed.

- **To identify and assess critical exchanges within episodes which result in features ending up in the final design.**

  Critical episodes will be firstly identified using the final designs that the teams propose, before these designs are reverse engineered (to establish which initial ideas were carried through to the final design). While these episodes will be assessed as normal (effective criticism/questioning will either move an idea or project forwards or stop it altogether), they will be compared to a selection of standard episodes to see if there is a change in behaviour within the group during these (e.g. the frequency or type of criticism/questioning).

- **To identify a relationship between the types of criticism/questioning and their effectivity in developing a quality idea.**

  The responses to effective criticism/questioning (advancement or non-advancement of an idea) will be compared in standard and critical episodes. As all ideas that have made it *through to the final design will be considered "high quality", any increase in a particular behaviour during critical episodes that leads to further development is desirable.*

- **To generate rules for concept development.**

  If the change identified between the types of criticism/questioning and their effectivity is large enough in critical episodes when compared to standard, then rules for concept development will be generated in accordance with the behaviour that has led to these results.

3.1. STEP-BY-STEP PROCESS

The step-by-step process used to obtain the results seen in **section 4.1** was as follows.

- A preliminary viewing of the recordings was undertaken to generate a full list of the ideas the groups generated during these early sessions (see **Appendix 9.2**). During this, all episodes relevant to a particular idea were identified and documented.

- The final designs of the groups were then observed, before being reviewed against this list to identify the ideas that were included. Episodes that discussed ideas included in the final design were now called "critical episodes", as they contained interactions that were key to developing the overall solution.

- The recordings were reviewed again, with several objectives in mind.

  - Firstly, the behavioural trends of the groups were identified and compared for normal and critical episodes. These trends were the block frequency (the amount of blocks that occur within an episode, exclusive of the primary block) and support for the block (how many times other group members agreed with the reasoning behind the block).

  - The types of block were then identified (criticism or questioning), before these sub-categories of block were broken down again. These were the types of criticism and the types of questioning.

  - From here, the responses to each type of criticism or question were identified. As mentioned earlier, these were categorised as advancement or no advancement.

  - Next, the research questions were tested by taking the results gathered from normal and critical episodes and looking at changes in the raw data. These changes were then statistically tested by conducting a series of chi-squared tests.

  - Finally, any observations noted during the team meetings or when reviewing the recordings were expanded upon. The data in these was qualitative, and was split into two parts in the results; findings directly related to the project (**section 4.2**), and findings indirectly proving research given in the literature review (**section 4.3**).

4. RESULTS AND ANALYSIS

To generate ideas, all three teams conducted one main brainstorming session[2]. In these sessions, Group A came up with twenty initial ideas, Group B with twenty-six and Group C with thirty-nine, with one for Group A, six for Group B and seven for Group C integrated respectively into the final design (a list of these ideas is given in **Appendix 9.2**). Next, the recordings were reviewed to identify the number of critical episodes for each team; these were eleven for Group A, twenty-three for Group B and twenty for Group C. For accurate comparison, a random sample of normal episodes equal to the amount of critical episodes for each group was assessed.

4.1. RESULTS GAINED USING QUANTITATIVE DATA

Having reviewed the recordings to identify the critical episodes, interaction dynamics notation was used to reveal the behavioural trends of each group during these. The criteria that results were gathered for were block frequency and support for block. These findings are shown in **Table 4-1**.

Table 4-1: *Frequency of the behavioural trends (block frequency and support for the block) of the three groups for episodes of normal and critical importance.*

Group	Total episodes	Episode importance: Normal (50% of total episodes)		Episode importance: Critical (50% of total episodes)		Percentage change	
		Block frequency	Support for block	Block frequency	Support for block	Block frequency	Support for block
A	22	*16 (1.45 per episode)*	*1 (0.09 per episode)*	*17 (1.55 per episode)*	*3 (0.27 per episode)*	*+6.9%*	*+150%*
B	46	*57 (2.48 per episode)*	*9 (0.39 per episode)*	*61 (2.65 per episode)*	*6 (0.26 per episode)*	*+6.9%*	*-33%*
C	40	*69 (3.45 per episode)*	*8 (0.40 per episode)*	*74 (3.70 per episode)*	*6 (0.30 per episode)*	*+7.2%*	*-25%*
TOTAL	54	142	18	152	15		

The results show Group A with the lowest amount of episodes out of the three (**with 22**), as well as a lower block frequency per episode for both normal (**1.45 per episode**) and critical (**1.55 per episode**). The total amount of blocks during their normal and critical episodes combined (**33**) was over three times smaller than the total for Group B (**108**) and over four times smaller than Group C (**143**). Saying this, their percentage increase in block frequency between normal and critical episodes was the same as Group B (**+6.9%**) and approximately the same as Group C (**+7.2%**), with a **range of 0.3%** between all three groups calculated. Group C had the highest rate of blocking during episodes (**3.45 per normal** and **3.70 per critical**), meaning that they ended up a higher frequency of blocks than Group B, even though there were fewer episodes in which they discussed their ideas. However, the support rate for blocks among all three groups were relatively low for both normal and critical episodes. Adding to this, the rates of support for all three groups were mostly constant.

The raw data given in these results shows a small increase in the amount of blocking in critical episodes when compared to normal (average of **+7%**), indicating that there is a relationship

[2] *Group C actually held two brainstorming sessions. The first took place before a meeting with their external client; following this meeting, the team decided to focus on the operating table element of their brief. In their first brainstorming session, they came up with nineteen ideas. Of these, ten were based on the table, with two integrated into the final design.*

between the frequency of blocks and the episode importance. There are also large differences in the support for blocks, but the percentages move in both directions. To test the validity of the results acquired, chi-squared tests were carried out on both sets of data to confirm if the percentage changes were significant. The results of these tests indicated that there was no significant increase in the block frequency between normal and critical episodes, or in the support for a block, as the null hypothesis (H_0) for both was not rejected. These results will be explored further in **section 5.1**. The calculations done to get to these results are given in ***Appendix 9.3***.

Next, these blocks were broken down into criticism and questioning and their respective sub-categories, in order to determine which types of criticism and questioning were used the most in group interactions. The results of this are given in **Tables 4-2** and **4-3, Fig.5** and **Fig.6**.

Table 4-2: *Frequency of the types of criticism present in the three groups meetings for episodes of normal and critical importance.*

Type of criticism	Group						TOTAL
	A		B		C		
	Normal episode	Critical episode	Normal episode	Critical episode	Normal episode	Critical episode	
Destructive	*0*	*0*	*0*	*0*	*0*	*0*	**0**
Standard	*1*	*1*	*7*	*7*	*3*	*4*	**23**
Constructive	*0*	*1*	*3*	*4*	*1*	*0*	**9**
TOTAL	**1**	**2**	**10**	**11**	**4**	**4**	**32**

Table 4-3: *Frequency of the types of questioning present in the three groups meetings for episodes of normal and critical importance.*

Type of question	Group						TOTAL
	A		B		C		
	Normal episode	Critical episode	Normal episode	Critical episode	Normal episode	Critical episode	
Remembering	*1*	*0*	*2*	*1*	*1*	*3*	**8**
Understanding	*5*	*8*	*15*	*15*	*34*	*24*	**101**
Applying	*7*	*6*	*13*	*14*	*6*	*8*	**54**
Analysing	*2*	*1*	*12*	*14*	*8*	*19*	**56**
Evaluating	*0*	*0*	*5*	*6*	*15*	*16*	**42**
Creating	*0*	*0*	*0*	*0*	*1*	*0*	**1**
TOTAL	**15**	**15**	**47**	**50**	**65**	**70**	**262**

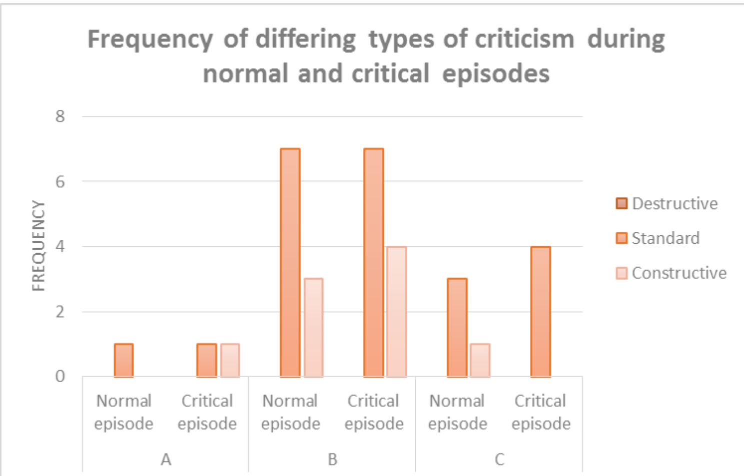

Figure 5: *A bar chart showing the frequency of differing types of criticism during normal and critical episodes.*

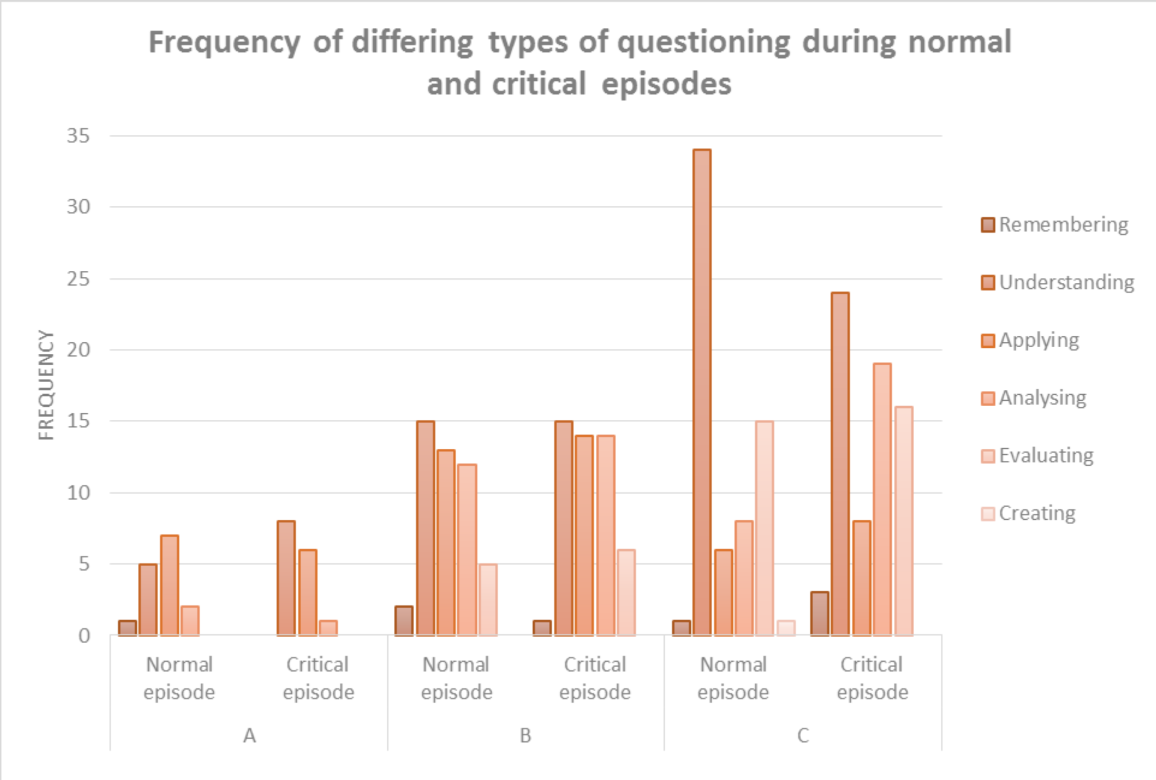

Figure 6: *A bar chart showing the frequency of differing types of criticism during normal and critical episodes.*

Overall, levels of criticism among the three groups were low, with only 32 instances being logged, **proving the hypothesis made in section 3**. Of these, Group B had the highest levels of criticism, with almost twice the amount of Groups A and C combined. While there was not a single instance of destructive criticism in all the teams, standard criticism was used **164%** more than constructive, with an increase in critical episodes of **8%**. There was also a real terms increase of **20%** for constructive criticism; however, chi-squared tests for these results again showed that the increases were not significant enough to be attributed to external circumstances.

The questioning rates within sessions were much higher than criticism, although these were spread depending on their cognitive complexity. The two extremes of questioning (remembering and creation questions, levels 1 and 6) were only asked nine times by all three groups between them, and as a result will not be evaluated due to their rarity. The other types of questions were asked more regularly, with understanding questions (L2), application questions (L3), analytical questions (L4) and evaluative questions (L5) making up **39%, 21%, 21%** and **16%** of total questioning by the groups respectively. When considering all groups, higher level questioning increased by **7%** in critical episodes compared to normal (**Fig.7**).

Again, Group A had the lowest levels of questioning with only 30 questions in total. Adding to this, the cognitive complexity of their questioning was lower; with **nine out of ten** at level 3 or under. This is a lower rate than groups B and C (**six out of 10** at level 3 or under). Group C asked the largest amount of evaluation questions; at **31**, they asked nearly three times as many as Groups B and C combined. Also observed again was a percentage change between normal and critical episodes for the differing types of questioning. Understanding questions (L2) were not asked as much (**-13%**), while there were moderate increases for application and evaluative questions (L3 and L5, **+8%** and **+10%** respectively). The biggest increase was in the amount of analytical questions, with an increase of **55%**. Saying this, chi-squared tests again proved most of these results to be non-significant. A relationship was found between industrial experience and the level of questioning asked, with chi-squared tests proving that the relationship between groups with industrial experience and higher levels of questioning was significant.

A chart showing the percentage difference for each type of questioning for normal and critical episodes is given in **Fig.8**.

Finally, the group response to criticism and questioning were determined, with section 3 of this study defining them as either advancement or non-advancement of an idea. The results of this analysis are shown in ***Tables 4-4*** and ***4-5***.

Table 4-4: *Frequency of the responses to the types of criticism present in the three groups meetings for episodes of normal and critical importance.*

Group	Type of criticism	Response type				TOTAL
		No advancement		Advancement		
		Normal	Critical	Normal	Critical	
A	Destructive	*0*	*0*	*0*	*0*	**0**
	Standard	*1*	*1*	*0*	*0*	**2**
	Constructive	*0*	*1*	*0*	*0*	**1**
B	Destructive	*0*	*0*	*0*	*0*	**0**
	Standard	*2*	*2*	*5*	*5*	**14**
	Constructive	*0*	*0*	*3*	*4*	**7**
C	Destructive	*0*	*0*	*0*	*0*	**0**
	Standard	*0*	*0*	*3*	*4*	**7**
	Constructive	*1*	*0*	*0*	*0*	**1**
TOTAL		**4**	**4**	**11**	**13**	**32**

Table 4-5: *Frequency of the responses to the types of criticism present in the three groups meetings for episodes of normal and critical importance.*

Group	Type of question	Response type				TOTAL
		No advancement		Advancement		
		Normal	Critical	Normal	Critical	
A	Remembering	*1*	*0*	*0*	*0*	**1**
	Understanding	*1*	*1*	*4*	*7*	**13**
	Applying	*2*	*2*	*5*	*4*	**13**
	Analysing	*0*	*0*	*2*	*1*	**3**
	Evaluating	*0*	*0*	*0*	*0*	**0**
	Creating	*0*	*0*	*0*	*0*	**0**
B	Remembering	*0*	*0*	*2*	*1*	**3**
	Understanding	*2*	*1*	*13*	*14*	**30**
	Applying	*1*	*3*	*12*	*11*	**27**
	Analysing	*3*	*4*	*9*	*10*	**26**
	Evaluating	*0*	*0*	*5*	*6*	**11**
	Creating	*0*	*0*	*0*	*0*	**0**
C	Remembering	*0*	*0*	*1*	*3*	**4**
	Understanding	*4*	*7*	*30*	*17*	**58**
	Applying	*1*	*1*	*5*	*7*	**14**
	Analysing	*1*	*2*	*7*	*17*	**27**
	Evaluating	*3*	*0*	*12*	*16*	**31**
	Creating	*0*	*0*	*1*	*0*	**1**
TOTAL		**19**	**21**	**108**	**114**	**262**

Overall, the teams aimed to resolve any issues arising from criticism or questioning, with advancement of an idea being their preferred response instead of no advancement over **three times as much** for criticism, and over **five times** for questioning. The main takeaway looking at the criticism results was that there was **no difference** in the response type for each type of criticism in normal and critical episodes. The questioning results did show small increases, with advancement responses **increasing by 5%** and no advancement **increasing by 10%**, but again these rises were proven to be insignificant using chi-squared tests. A further set of tests also proved that even though there was an increase in advancement compared to non-advancement, there was no significant increase for any type of question.

There were some changes in group behaviour, but most of these related to the frequency of understanding questions asked. The amount of these asked that led to advancement in the project nearly doubled for Group A, but this is the opposite of what happened to Group C, who saw a reduction of **43%**. The number of analytical questions asked by Group C resulting in advancement also increased from **seven to seventeen**.

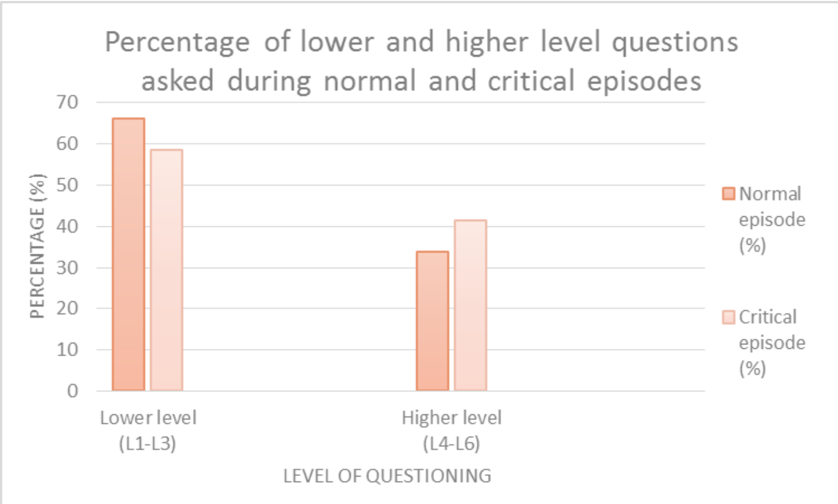

***Figure 7**: A bar chart showing the percentage difference for the levels of questioning in normal and critical episodes.*

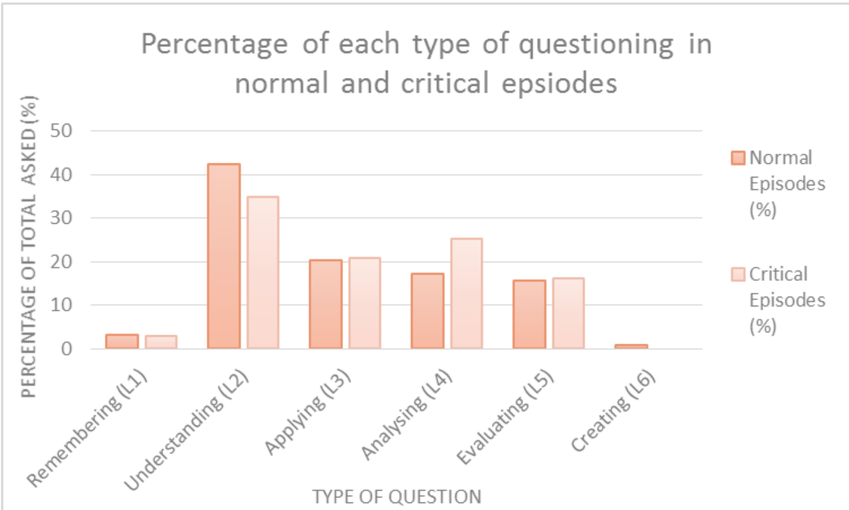

***Figure 8**: A bar chart showing the percentage difference for each type of questioning for normal and critical episodes.*

4.2. RESULTS GAINED USING QUALITATIVE DATA

As an analysis on the language used in early NPD sessions, not all of the findings are expressible using numbers. This section gives the findings gained using qualitative data from following the teams for five weeks.

- **While it was rare for a block to be supported, support for the answer was common in all the groups.**

  The results given earlier show that that support for a block was fairly uncommon, with an average support rate of 0.28 per normal episode and 0.29 per critical episode. However, support (defined by Sonalkar as an indication that "the speaker understands and agrees with the previous move", see **Appendix 9.1**) was a lot more common as team members aimed to build on a point that an individual had previously raised.

  An example of this is given in **Table 4-6**. Support for an answer is highlighted in red.

Table 4-6: An example, taken from Group C that shows team members supporting the answers in response to blocks. This is coded using IDN.

Group Member	Comment	Symbol
K	Why do we have a table that is that shape?	
N	I just hate the shape of a rectangle… I feel that it's so un-ergonomic… So I love circles…	
E	We could have rounded corners.	
N	Yeah… Plus if you're in the corner…	
K	Maybe it depends on the types of operation?	$\xrightarrow{?}$
N	Write that down as a question to ask our client…	→
K	It might be useful to switch these shapes here, so if everything is here and you move it to here then it's easier. Then when you don't want it take it away.	→
N	Basically yeah, so whatever position you put it in, it still looks like that, that sort of workspace. He can probably stay on his chair while he's working.	∩
E	Exactly!	
N	So maybe this is pre operation and this is post?	$\xrightarrow{?}$
L	It'll be universal stationary, and when necessary you can add what is needed	→
N	Yeah.	
K	But then one problem will be while you're doing the operation, let's say it takes 30 minutes but you take 15 the animal has to wait. You could do two at the same time.	
H	But then you could have multiple animals on the table at the same time.	

- **Most advancement did not arise due to questioning and criticism, but rather through statements made by team members.**

  As shown by the figures in **Table 4-2**, there were low levels of criticism overall as team members seemed reluctant to act in a negative way towards ideas. Indeed, most ideas were encouraged by individuals in all the teams, who aimed to build on them as much as they could. While they did this through questioning and to a much smaller extent criticism, most advancement actually came through general discussion, with suggestions to improve an idea normally accepted by team members.

  An example of this is given in **Table 4-7**. Statements made by team members are highlighted in red.

Table 4-7: An example, taken from Group A that shows advancement coming through statements made by team members. This is coded using IDN.

Group Member	Comment	Symbol
W	You know you were saying last week how you had a grip on the cast to keep the bone in place? If you re-watch some of the videos…	→

J	I don't think you'd need the tubey grip.	
T	No.	
W	You don't need a cast as well. The cast literally…	∩
J	It just holds it.	
T	If you read the standards as well, it says that it should be the plaster of Paris. It says that you can build it upright and around the base if you want to.	
W	It's not the tubey grip keeping everything in place, it's the cast.	
J	And the thing is, if a knitted structure forms, it can do this on its own.	∩
T	So let's not put it in then!	→

- **One team member did the majority of the criticism in each team, although this criticism often in the form of questions.**

While criticism was limited in its scope, it was interesting to note that one team member did the majority of the criticism for each team. Adding to this, the background of this individual changed in each group; in one it was an individual with industrial experience in NPD, in another it was an individual with industrial experience (but not in NPD) while the third had no industrial experience at all. This individual, according to the closing interview conducted with teams, was also more likely to give destructive criticism later in the project. Saying this, criticism of an idea or person was often not given directly; while it was sometimes given with a change in tone, often it was given as a question.

An example of this is given in **Table 4-7**. Direct or implied criticisms made by team members are highlighted in red.

Table 4-7: *An example, taken from Group A that shows criticism. This is coded using IDN.*

Group Member	Comment	Symbol
P	So you can say it's basically… You pick a fibre that you are looking to knit, and it has a re-programmed length or shape that it should adhere to… I mean, when you heat it, it becomes that shape.	→
T	How does it go hard though?	
P	Heating.	
T	It wouldn't go hard though…	
P	Well, it wouldn't *go hard*, but because of now it goes back to its original shape, if you arrange that with the knitted structure it constricts it and then…	
J	OK	
T	That sounds really hard, so I think we should go with something else.	

4.3. INCIDENTAL FINDINGS

As well as the findings related to the research questions and hypothesis, several observations were made while following the groups. While most were not directly related to the project, the following discovery confirms findings included in the literature review.

- **Individual brainstorming led to a higher amount of ideas than group brainstorming.**

  While Group A initially brainstormed as a group[3], Groups B and C split the project into sub-systems before commencing with a brainstorm. Like Group A, these sub-systems were each given to an individual, but ideas from each sub-system were presented in the original brainstorming session, with team members often adding to each concept. This led to Group A creating **20 initial ideas**, Group B creating **26** and Group C creating **39**. Breaking these results down further, Group A actually generated the most idea per sub-system (**2.4 ideas more** than Group B and **0.2 ideas more** than Group C), even though they created less ideas per individual (**1 idea less** than Group B and **3.2 ideas less** than Group C).

  A breakdown of these results is given in **Table 4-9**.

Table 4-9: *A breakdown of the ideas, sub-systems and individuals in all three groups during the idea generation phase of the project.*

	Group		
	A	**B**	**C**
Group or individual brainstorming?	*Group*	*Individual*	*Individual*
Number of ideas	*20*	*26*	*39*
Project sub-systems	*3*	*6*	*6*
Individuals per sub-system	*2*	*1*	*1*
Ideas per sub-system (ave.)	*6.7*	*4.3*	*6.5*
Ideas per individual (ave.)	*3.3*	*4.3*	*6.5*

[3] *While Group A initially brainstormed as a group (coming up with twenty ideas), they then split the project into three sub-systems and assigned two man sub-teams to come up with three complete concepts for each sub-system. Importantly, only one concept was selected from each sub-system – meaning three ideas were taken forwards overall. Saying this, their final design actually consisted of a knitting machine mountable onto a table top, with their orthopaedic wrist cast designed during these sessions used in their business proposals as a possible product the machine could make, and so one idea from their group sessions during the observation period was used in the final product design.*

5. DISCUSSION

Like **section 4**, this part of the study will be split into different sections.

5.1. RESULTS GAINED FROM QUANTITATIVE DATA

The first set of results gathered using quantitative data focused on the frequency of behavioural trends between the three groups for episodes of normal and critical importance, with the raw data showing a small average increase in block frequency of 7% in critical episodes compared to normal between the three groups. Importantly, this rise stayed constant among all three groups, independent of total episode frequency – indicating that this was a common behavioural trait. The same could be said for the support for block frequency if Group A's data in normal episodes is treated as an outlier. There are several external circumstances (see **section 5.4**), that give reasons as to why this result should be treated as such.

Saying this, a chi-squared test found there to be no significant difference in the blocking frequency in normal and critical episodes, as the values obtained were exactly the same as the values expected. This particular test also had a large sample size, with 294 observations used in the assessment. This eliminated one of the biggest limitations of the chi-squared test, which is that it is very sensitive to small sample sizes, and the effect of this can be immediately observed when comparing the results gained in the "blocking frequency" test to the results gained in the "support for block" frequency results. With a sample size of 33, percentage increases in this data-set tended to be more pronounced as there is less data to dampen the effect of one particular result. Because of this, changes ranging between -33% and +150% were observed, with chi-squared tests showing that this result (while still not significant enough to reject H_0) was more noteworthy than expected. Indeed, if the support for block sample size was similar to the block frequency sample size, the result would have been seen as significant. However, taking out the data collected from Group A would have supported the original finding (results not significant enough to suggest a relationship), further supporting the notion that results from this group were outliers.

While the raw data showed increases of 8% and 20% in standard and constructive criticism respectively during critical episodes, chi-squared tests again showed these increases to be not significant. However, what was significant were the low overall levels of criticism (proving the hypothesis made in **section 3**), and the much higher frequency of standard criticism compared to constructive and destructive. The low levels of criticism overall could be attributed to several factors. The first is that the early stages of the project consisted of mainly brainstorming and the early stages of conceptual design, and in these stages the teams actively aimed to keep criticism to a minimum, in order to have a greater pool of ideas to choose from. As mentioned earlier, they were also taught Osborn's "no criticism" rule, and as a result of this teams might not have been acting *fully* naturally. Furthermore, having an external individual observing their sessions, as well as knowing that these sessions were being recorded (and the recordings were possibly available for academics to listen to in the future), could have led the teams to act unnaturally – indeed, Group A referenced the fact that they were being recorded several times during their sessions. Saying this, Matthews found that engineers "rarely orient" to brainstorming rules, suggesting that this factor did not play as big of a role as previously thought (Matthews, 2009).

The final factor is that the teams still did not know each other well. Using Tuckman's stages of team development (see **section 2.3**), the teams were still at the "forming" stage. Because of this, the groups were still acclimatising to one another, and team members were on their best behaviour by being as polite, positive and restrained as they could be. A better time to observe criticism would have been later in the project, when the concept had been chosen and true development was underway. The need for criticism here, coupled with the fact that the groups would move into the "storming" phase (where disagreements start to emerge), would most likely have seen a larger proportion of criticism in all areas. In particular, one group reported a series of destructive criticism episodes later in their design process which eventually led to an "argument" between two team members. This is expanded on further in **section 5.2**.

Standard criticism was preferred to constructive (+164%), agreeing with Linley, who found constructive criticism to be the least common type. Whilst his study did not give any possible reasons for this finding, this research proposes several. The first is fairly obvious – individuals simply did not have a resolution to the issues that they had raised. However, by simply giving standard criticism, individuals might not even *try* to come up with a solution, instead relying on others to contribute instead. This is known as free riding, and was one of the factors hypothesised by Diehl and Strobe in their study on productivity loss. The third is that team members might not have the *confidence* to make suggestions. As the teams were still relatively new, members might want to look as competent as possible, and therefore might be reluctant to suggest solutions that could themselves be rejected, denting group confidence in said individual. Saying this, this is unlikely as these people have the confidence originally to criticise.

Overall, there was an increase of 7% in higher level questioning in critical episodes when compared to normal, caused mainly by a decrease in understanding questions being asked (L2) and an increase in analytical questions (L4). The decrease in understanding questions could be explained by critical episodes tending to have better ideas contained within them, which need less explanation as they are more developed (incidentally, a decrease in the frequency of lower-level questions asked in real-life situations was mentioned in the literature review). The increase in analytical questions is a result of team members wanting to *why* a solution works instead of *how*, and any issues raised by these types of questions can be thrown out to the group, who as mentioned earlier will work together to try and solve them. Furthermore, analytical questions are the highest level that require the person to explain without giving an opinion, and a result of this is that they are more likely to raise any issues with an idea if they exist. In turn, these are normally resolved, making the idea stronger.

Incidentally, the two groups led in their design stage by individuals with industrial experience in NPD asked a smaller proportion of lower level questions (six out of ten), lower than what the literature review observed among students (seven out of ten) and the group than did not have this (nine out of ten). Furthermore, chi-squared tests found this to be consistent for a range of alpha values, with the results indicating that there was a 0.5% chance of this relationship being a false positive. These results indicate that there is a relationship between the experience and level of questioning, but the size of this effect cannot be deduced without a control group to accurately measure the change. It is possible that spending time in industry, and being surrounded by people more likely to be asking higher levels of questioning have increased the critical thinking skills of these individuals, who in turn can apply this to their academic projects. Adding to this, people with industrial experience have more expertise with both participating in and leading brainstorming/NPD sessions, with an academic approach (based on learning) very different to an industrial one (based on effective results).

Overall, the teams aimed to resolve any issues arising from criticism or questioning, with advancement of an idea being their preferred response instead of no advancement over three times as much for criticism, and over five times for questioning. Again, Tuckman's model explains this, with the teams wanting to be as positive as possible by coming up with potential solutions to every issue. This also mirrors Sonalkar's results, who found that teams would more often than not attempt to resolve any blocks that were raised. Saying this, these results showed some of the smallest changes between normal and critical episodes using the raw data, which were proven again to be insignificant rises using chi-squared (although the increase in analytical questions leading to advancement would have been significant if the sample size was doubled). This constant insignificance shown by the chi-squared tests could also be a sign that using expecting significant differences between normal and critical episodes was wrong as, looking at the data, it is clear that the biggest change in language does not come from whether a normal or critical idea is being discussed, but rather the group itself. This is because each group will have their own unique style, influenced by the abilities and traits of the individuals within (such as critical thinking skills, interpersonal skills, experience and personality), and because of this they develop their ideas in a similar manner.

The results gained using quantitative data were not able to show a relationship between the types of either criticism or questioning and their effectivity in developing ideas. However, the hypothesis that low levels of criticism would be observed was proven to be true.

5.2. RESULTS GAINED FROM QUALITATIVE DATA

Three findings were made from shadowing the teams for the initial five week period. The first was that while it was rare for a block to be supported, support for the answer was common in all of the groups. Again, this can be put down to the stage of the project that the teams found themselves at, as well as the stage of the group development. In later sessions, support for the answer diminished as disagreements became more common (according to exit interviews), and the groups moved from forming to storming. Adding to this, support was not needed as much once brainstorming had finished, and decisions vital to the success of the project were made. At this early stage however, teams were comfortable making decisions they knew were reversible if needed (due to the relatively large amount of time they had left).

Also observed was the fact that one individual in each team did the majority of the criticism. This individual had no shared traits in each group; in one it was an individual with industrial experience in NPD, in another it was an individual with industrial experience (but not in NPD) while the third had no industrial experience at all. This suggests that personality is more important than experience when it comes to giving criticism. This finding also mirrors De Bono's six thinking hats method (De Bono, 1985). In this, he describes a tool for group discussion that separates thinking into one of six roles; optimism, judgement, intuition and creativity, as well as a person who calls for information and a person to manage the thinking process. It is interesting that two of the six hats have been filled naturally; as well as judgement, one individual (the project manager in all cases) always called for simply the facts. Furthermore, whilst not a hat, one member of each team participated notably less in group discussions that all other team members.

The team atmosphere also was a massive contributor to how freely criticism could be expressed. One group was very personal, and so the members felt comfortable with giving feedback early on in their sessions, while another group had a more results-based approach, and so suppressed criticism to a degree that it was not offered as freely afterwards. This effect, known as "evaluation apprehension" (see **section 2.1**), sees individuals being reluctant to share their suggestions due to fear of criticism – or in this case, being ignored. Interestingly, while all three teams reported destructive criticism after the initial observation period ended, the group that suppressed their critical individual's feedback experienced an argument later in the project. Saying this, individuals giving criticism and asking questions actively tried to phrase them in a way that was not aggressive or hurtful to other team members, and this was mostly reciprocated, going against the predictions the preliminary group analysis gave. Only when this did not happen (or when members felt they were being ignored) was there a change in the tone of their questioning. Indeed, a larger amount of rhetorical or leading questions would sometimes be asked, sometimes in a "passive-aggressive" tone (insinuating criticism while not directly giving it). This change in tone often changed the meaning of the sentence (Mullennix, et al., 2002). Adding to this, in all three projects, the person who was critical during the early stages was more likely to give destructive criticism later in the assignment.

While criticism and questioning led to advancement of an idea more than non-advancement, the majority of language that advanced ideas came in the form of statements made by team members. However, due to their lack of sub-categories, as well as the complexity of language that is contained within them, a decision was made to not include them in the final results. In particular, the *way* that statements were presented to the teams was critical in their meaning, as this changed depending on the tone, inflection and stress on certain words that they were asked with[4]. For example, the meaning of the sentence "I don't think that he would like this idea" (he referring to an external client) changes depending on which word (in bold) is tonally altered;

- **"I** don't think he would like this idea" suggests that the another group member does think the idea would be liked by the client,

[4] *Body language was also another factor that changed the meaning of statements and sentences. Adding to this, negativity could often be detected, even if it was not explicitly stated, through body language. However, since the majority of the data collected was in audio form, body language was unable to be analysed.*

- "I don't **think** he would like this idea" suggests that while the individual talking thinks the idea will not be liked by the client, it is worth presenting it anyway,

- "I don't think **he** would like this idea" suggests that the idea being discussed might not be liked by the client it was being presented to, but liked by another person close to the project (i.e. the project supervisor),

- "I don't think he would **like** this idea" suggests that the reaction of the client to the idea might be more extreme (i.e. loves or hates the idea),

- "I don't think that he would like **this** idea" suggests that while the idea being discussed might be rejected, there are alternatives that the client would approve of.

Statements were by far the most common type of advancement within the groups. The effects of this were magnified as, being at such an early stage in the project, teams were trying to build and improve on every idea independent of quality. However, it can also explain why so many bouts of criticism were presented in the form of a question, even though the words they consisted of did not suggest that they were. By not directly criticising, individuals were able to raise issues with an idea or concept without having to worry about offending anyone. Adding to this, it took the focus away from the individual saying it, meaning students that were not as confident in their abilities were still able to participate.

While the initial meetings were good for understanding the project and setting the tone, all three groups had final designs significantly different to what they selected in their feasibility reports. While ideas from the first five weeks were incorporated into each final design, the majority of these were not key. Instead, the critical elements of the final designs were created after the project proposal deadlines, with the aim before this to reach consensus on the *type* of solution agreed. All teams unanimously agreed that meetings with their supervisors or external clients helped progress their projects the furthest, as feedback was given on the groups ideas, and the key likes and dislikes of their proposals at each particular stage were identified. The biggest example of this is Group A, who ended up designing a knitting machine mountable onto a table top after stating numerous times that there was "no point in designing a machine if there are already good machines out there". Again, this agreed with the findings of Kluger and DeNisi, who said that feedback was most useful when task complexity was high, individual goals were difficult to achieve and feedback on changes was given – conditions that were present in the projects of all three groups. Indeed, most significant concept development was actually done *after* the mandatory technical proposals were handed in (and hence, the observation period had ended); meaning that, when combined with the teams being more familiar with each other, the highest levels of criticism and questioning would have been observed during this period also.

Saying this, all the groups reported brainstorming sessions that took place throughout the project, as ideas to solve the issues raised during the supervisor/external client meetings were created. However, they also stated that these tended to get shorter and shorter as the project progressed, most probably due to the fact that the magnitude of the changes implemented reduced after each meeting (as the final design got closer to what the supervisor or external client was happy with). Indeed, the groups found that the project progressed faster the further they got in.

5.3. RESULTS FROM INCIDENTAL FINDINGS

While there was only one incidental finding related to the literature discussed in **section 2.1.**, it adds to the evidence that individual brainstorming is more productive than group brainstorming (**Fig.2**). However, the method in which this result has been derived from is different to many of the other studies. All the groups followed had six members, and as a result the project was split into a series of sub-sections. Two teams split their projects into six systems (one for each person) who was then given complete control over that part of the project, including creating potential concepts, while the third split the project into three. While it was not possible to follow the sub-teams, group brainstorming gave a lower amount of ideas (per individual) than individual brainstorming. This results could due to external pressures, such as time limitations (as individuals do not want meetings to go on for too long) or a narrower scope of thinking. However, the various papers that have actively tested this go into greater detail on the reasons why.

Incidentally, this also negated the effect of anchoring that Thompson said existed in brainstorming sessions. She argued that ideas raised earlier in a session would be more likely to be incorporated into the final design, but all of the teams presented their ideas by sub-system. Saying this, the ideas created during the early stages of their projects were very different to what was proposed in the final designs of each team, due to the specification changing due to feedback from supervisors and/or external clients. In particular, the product design of Group A changed from an orthopaedic wrist cast manufactured using knitting technology to a knitting machine mountable onto a table top. With a project manager with industrial experience in NPD, Group C in particular met with their client as often as they could, taking their ideas with them to discuss and gain feedback. They were very wary of developing an idea that had not been agreed with their client too much, as if it was rejected their work would have ultimately been for nothing.

5.4. LIMITATIONS

While this study has some interesting findings, there are several limitations to it also. These are listed below.

- **Sample sizes too small to make significant conclusions**

  While the initial sample size for block frequency was 294, all other data-sets tested had sizes smaller than this. Smaller data-sets were more likely to show extreme percentage increases with minor changes in the raw data, but these increase the likelihood of type 1 or 2 errors in chi-squared tests (wrongly rejecting H_0 or wrongly accepting H_1). As mentioned earlier, the frequency change in analytical questions between normal and critical episodes would have been significant if the sample size had doubled with the same proportions.

  The necessary sample size (N.S.S) was calculated using;

  $$N.S.S = \frac{(S.D.)(1 - S.D.)(Z\ score)^2}{(margin\ of\ error)^2}$$

  With a standard deviation of 0.5, a 95% confidence level (which gives a Z score of 1.96) and a margin of error of ±5% gives a N.S.S. of 385, which is larger than any of the datasets tested.

- **Potentially unreliable results from Group A**

  The frequency of blocks (and support for a block) from Group A is substantially lower than Groups B and C. However, the development process that Group A went through was different to the other two teams. The time taken for Group A to understand their project was longer due to circumstances beyond their control, such as visits to experts, being scheduled much later than the other two teams. As a result of this, they started brainstorming and developing their concepts closer to the deadline, meaning that key decisions and meaningful discussion was delayed in favour of getting deliverables completed. It also must be noted that their final *design* actually consisted of a knitting machine mountable onto a table top, with their orthopaedic wrist cast design used in their business proposals as a possible product the machine could make.

- **A lack of control groups in the experiment**

  Due to external circumstances again, the groups could only be observed in a natural setting. A result of this is that no control groups were formed during this study, which meant that the independent variables (types of criticism and questioning) could not be tested for *effect*, but only for a *relationship*. Many people, including leading figures in business, argue that the development of an idea is more important than the initial idea itself, but this was not able to be tested. An ideal experiment would have given a series of teams the same concepts, looked at how they developed them in a fixed time period (focusing on language) before having an independent assessor mark their solutions. Even if setting up control groups was not possible, it might have been better to use projects that have the same aim in this experiment (such as the design and make project for first years).

- **No way of measuring the quality of the initial concepts**

  While the language used during the early NPD sessions was the most important thing to look at, due to uncontrollable external circumstances there was no way of actually measuring the quality of the concepts before and after the interactions. An assumption that all ideas that made it into the final design were high quality had to be made, but this is not always the case. Just because an idea has made it into the final design does not automatically make it a fantastic idea, but rather the best of all of the ideas that the group members had. Adding to this, interactions that lead to a low quality final design would not be wanted, but there was no way to resolve this issue without interfering with their results or gaining confidential information.

While the hypothesis that low levels of criticism would be observed was proven to be true, no relationship between the types of either criticism or questioning and their effectivity in developing ideas was found. Because of this, no rules for concept development can be recommended at this time.

6. CONCLUSION

How are concepts *optimally developed* from initial ideas?

This is the question that this project aimed to answer. By following three undergraduate engineering teams for five weeks during the early design stages of their group business and design projects, a greater understanding of the language used and its role in developing ideas into concepts was expected, as well as the possibility of developing rules for concept development. However, due to the results gained, no rules for concept development can be recommended at this time.

Having reviewed the recordings to identify the critical episodes, interaction dynamics notation was used to reveal the behavioural trends (block frequency and support for block) of each group during these. The results showed that as well as having the lowest amount of episodes out of the three (with 22), Group A had a lower block frequency for both normal and critical episodes (1.45 and 1.55 per episode respectively) than Group B (2.48 and 2.65 per episode) and Group C (3.45 and 3.70 per episode). However, for all three teams, the percentage increase between normal and critical episodes was approximately the same (+7%), indicating that this was a common behavioural trait. Support rates for blocks were low in all three teams, with only 33 instances logged – but again, these were constant among all the groups. However, chi-squared tests found the percentage changes to be insignificant for both block frequency and support for block.

Next, these blocks were broken down into criticism and questioning, with the frequency of each type asked analysed. The overall levels of criticism were found to be extremely low, proving a hypothesis that stated that this would occur due to the teams being unfamiliar with one another, and hence trying to stay as polite, positive and restrained as they could be. Other factors, such as the stage of the project and the fact that they had been taught Osborn's rules were also seen as possibly contributing to the result. In the limited instances of criticism logged, standard criticism was used 164% more than constructive, with an increase in critical episodes of 8% for standard and 20% in constructive criticism. Factors such as lack of knowledge and free-riding were given as possible reasons for why this occurred, but saying this, chi-squared tests again found these results to be insignificant.

The questioning rates within sessions were much higher than criticism, although these were spread depending on their cognitive complexity. The two extremes of questioning (remembering and creation questions, levels 1 and 6) were only asked nine times by all three groups combined (making up 3% of total questioning). The other types of questions were asked more regularly, with understanding questions (L2), application questions (L3), analytical questions (L4) and evaluative questions (L5) making up 39%, 21%, 21% and 16% of total questioning by the groups respectively. When summarising the differences between the three groups, a larger proportion of lower level questions was observed in groups that did not have industrial experience in NPD (nine out of ten) than in groups that did (six out of ten), a result chi-squared tests found to be significant. Adding to this, a larger proportion of higher level questions was asked in critical episodes when compared to normal. In particular, a dramatic increase in analytical questions (L4, +55%) was observed, as well as a fall in understanding questions asked (L2, -13%).

The group responses to criticism and questioning were defined as either advancement or non-advancement of an idea, with results showing that advancement of an idea was preferred instead of no advancement over three times as much for criticism, and over five times for questioning. Again, Tuckman's group development model explained these results, which also mirrored Sonalkar's finding that teams would more often than not attempt to resolve any blocks that were raised. Overall, while criticism showed no difference in both rates between normal and critical episodes, small increases were observed in questioning (advancement responses increasing by 5% and no advancement increasing by 10%). However, chi-squared tests proved these rises to be insignificant yet again. Adding to this, there was no significant increase for any particular type of criticism or questioning.

Qualitative data findings were also presented in this study. Interaction dynamics notation found that while it was rare for a block to be supported, support for the answer was common in all groups. Furthermore, one team member did the majority of the criticism in all teams, sub-consciously taking the "judgement" role defined by Edward De Bono in his six thinking hats model. This unconscious role taking was further backed up by the team leader (in all three groups) calling for facts as often as possible, which is the role of another hat. It was also discovered that most advancement did not arise due to questioning and criticism, but through statements and external feedback.

Overall, while several increases were observed in the raw data, the vast majority of them were proven to be insignificant using chi-squared tests – with only the relationship between a higher level of questioning and industrial experience found to be significant. Circumstances such as small samples sizes made finding significance difficult (indeed, some relationships would have been seen as significant if the sample size was larger), while potentially unreliable data from Group A could have also skewed the results in an undesired direction. To eliminate the effects of these, any future work must use control groups.

7. FUTURE WORK

There are several directions that future work in this topic could take.

A consistent finding of this study has been changes in the raw data that have not been significant enough to reject the null hypothesis when tested using chi-squared. Some results however, such as the increase in analytical questioning frequency, would have been significant if the dataset tested was larger. Adding to this, this research has only looked at finding a *relationship* – which it has, albeit a small one deemed to be insignificant.

One of the things that could be done in the future is to replicate elements of this experiment with the same task. By doing this, the differing ways teams approach the same problem could be observed and tested against some of the criteria (such as differing types of criticism and questioning). In an academic setting within the University of Bath, the "design and make" module taken by first years would be ideal, as this is the only design task that is done with a large group (six members).Furthermore, as this task is relatively short, the entire design process could be observed, meaning that episodes that contain destructive criticism could be analysed (if present). Then downside to this is that using first years will mean that the approach taken is purely academic, whereas results obtained in this dissertation have found industrial experience has a positive effect on the design capabilities of engineers.

Another interesting aspect to consider is whether group members naturally take on the roles defined by De Bono's hats. One of the most interesting findings in this research was that two of the hats (judgement, and calling for facts) were filled naturally by team members. Looking at whether all of the hats are filled by group members, whether individuals take more than one hat at a time or if these hats rotate during sessions (as De Bono recommends). Any research looking at this would most likely be observation based, making the design and setup of this experiment easier. This also opens up the possibility of observing engineers in a professional setting as well as an academic one.

Saying this, it is vital that any future work uses control groups. By doing this, not only can any potential relationship be established, but the strength of it can be calculated also. Using Cohen's effect sizes, the direction of this relationship can be established also. Furthermore, control groups minimise the effects of external circumstances skewing the results in an undesired way by removing as many of these as possible.

8. REFERENCES

Baron, R. A., 1988. Negative effects of destructive criticism: Impact on conflict, self-efficacy and task performance. *Journal of Applied Psychology,* 73(2), pp. 199-207.

Baron, R. A., 1990. Countering the effects of destructive criticism: The relative efficacy of four interventions. *Journal of Applied Psychology,* 75(3), pp. 235-245.

Blankenship, K. L. & Craig, T. Y., 2006. Rhetorical Question Use and Resistance to Persuasion: An Attitude Strength Analysis. *Journal of Language and Social Psychology,* 25(2), pp. 111-128.

Bloom, B. S., 1956. *Taxonomy of educational objectives : The classification of educational goals. Handbook 1, cognitive domain.* 1st ed. New York: McKay.

Carter, B. & Wynn, G., 2014. *The Cowbell Principle.* Scotts Valley, California: CreateSpace Independent Publishing Platform.

Christenbury, L. & Kelly, P., 1983. *Questioning: A Path to Critical Thinking.* 1st ed. Urbana, Illinois: National Council of Teachers.

Cohen, J., 1992. A power primer. *Psychological Bulletin,* 112(1), pp. 155-159.

Cross, N., 1997. Creativity in Design: Analyzing and Modeling the Creative Leap. *Leonardo,* 30(4), pp. 311-317.

De Bono, E., 1985. *Six Thinking Hats: An Essential Approach to Business Management.* 1st ed. New York City, USA: Little Brown and Company.

Dekoninck, E., Barrie, J. & Linley, A., 2016. *Comparing the usefulness of rules & no rules in engineering design brainstorming?.* Aalborg, The 18th International Conference on Engineering & Product Design Education.

Diehl, M. & Stroebe, W., 1987. Productivity loss in brainstorming groups: Towards the solution of a riddle. *Journal of Personality and Social Psychology,* Volume 53, pp. 497-509.

Dillon, J. T., 1982. Cognitive correspondence between question/statement and response. *American Educational Research Journal,* Volume 19, pp. 540-551.

Edinger, S., 2017. Sales reps, stop asking leading questions. *Harvard Business Review,* pp. 2-5.

Gallupe, R. B. et al., 1992. Electronic Brainstorming and Group Size. *Academy of Management Journal,* Volume 35, pp. 350-369.

Helquist, H. J., Santanen, E. L. & Kruse, J., 2007. *Participant-driven GSS: quality of brainstorming and allocation of participant resources.* 40th Annual Hawaii international conference on system sciences, IEEE Computer.

Howard, T. J., Dekoninck, E. A. & Culley, S. J., 2010. The use of creative stimuli at early stages of industrial product innovation. *Research in Engineering Design,* 21(4), pp. 263-274.

Kluger, A. N. & DeNisi, A., 1996. The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin,* 119(2), pp. 254-284.

Krathwohl, D. R., 2002. A revision of Bloom's taxonomy: An overview. *Theory into Practice,* 41(4), pp. 212-218.

Landy, F. J., Farr, J. L. & Jacobs, R. R., 1982. Utility concepts in performance measurement. *Organizational Behavior and Human Performance,* Volume 30, pp. 15-40.

Lewis, A. C., Sadosky, T. L. & Connolly, T., 1975. Effectiveness of group brainstorming in engineering problem solving. *IEEE Transactions on Engineering Management,* EM-22(3), pp. 119-124.

Liden, R. C. & Mitchell, T. R., 1985. Reactions to feedback: The role of attributions. *Academy of Management Journal,* 28(2), pp. 291-308.

Linley, A., 2013. *Design Team Study: Comparing the Usefulness of Rules & No Rules in Engineering Design Brainstorming?.* MEng (Hons) Mechanical Engineering, Faculty of Engineering and Design: University of Bath, UK.

Loftus, E. F. & Palmer, J. C., 1974. Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of verbal learning and verbal behavior,* 13(5), pp. 585-589.

Loftus, E. F. & Zanni, G., 1975. Eyewitness testimony: The influence of the wording of a question. *Bulletin of the Psychonomic Society,* 5(1), pp. 86-88.

Maginn, B. K. & Harris, R. J., 1980. Effects of anticipated evaluation on individual brainstorming performance. *Journal of Applied Psychology,* Volume 65, pp. 219-225.

Matthews, B., 2009. Intersections of brainstorming rules and social order. *CoDesign: International Journal of CoCreation in Design and the Arts,* 5(1), pp. 65-76.

Mullennix, J. W. et al., 2002. Effects of variation in emotional tone of voice on speech perception. *Langauge and speech,* 45(3), pp. 255-283.

Nemeth, C., Personnaz, B., Personnaz, M. & Goncalo, J., 2004. The liberating role of conflict in group creativity: a study of two countries. *European Journal of Social Psychology,* Volume 34, pp. 365-374.

Nguyen, D. Q., 1998. The essential skills and attributes of an engineer: A comparative study of academics, industrial personnel, and engineering students.. *Global Journal of Engineering Education,* 2(1), pp. 65-75.

Osborn, A., 1953. *Applied Imagination - Principles and Procedures of Creative Writing.* Oxford, England: Scribner's.

Parnes, S. & Meadow, A., 1959. Effect of brainstorming instructions on creative problem solving by trained and untrained subjects. *Journal of Social Behavior,* Volume 50, pp. 633-646.

Phillips, N. & Duke, M., 2001. The questioning skills of clinical teachers and preceptors: a comparative study.. *Journal of Advanced Nursing,* 33(4), pp. 523-529.

Pritchard, R. D. et al., 1988. Effects of group feedback, goal setting, and incentives on organizational productivity. *Journal of Applied Psychology,* Volume 73, pp. 337-358.

Reed, K., 2010. *Skill Sets Required for Environmental Engineering and Where They Are Learned.* ProQuest LLC, Ed.D. Dissertation: University of the Pacific, California.

Saeed, T. et al., 2012. Development of students’ critical thinking: the educators’ ability to use questioning skills in the baccalaureate programmes in nursing in Pakistan.. *Journal of Pakistan Medical Association,* 63(3), pp. 200-203.

Schulz, B., 2008. The Importance of Soft Skills: Education beyond academic knowledge. *Journal of Language and Communication,* 2(1), pp. 146-154.

Sellappah, S., Hussey, T., Blackmore, A. M. & McMurray, A., 1998. The use of questioning strategies by clinical teachers.. *Journal of Advanced Nursing,* 28(1), pp. 142-148.

Sonalkar, N., Mabogunje, A. O. & Leifer, L. J., 2012. *A visual representation to characterize moment to moment concept generation in design teams.* Glasgow, UK, The 2nd International Conference on Design Creativity.

Thompson, L., 2013. *Creative Conspiracy: The New Rules of Breakthrough Collaboration.* Cambridge, Massachusetts: Harvard Business Review Press.

Tuckman, B. W., 1965. Developmental sequence in small groups.. *Psychological Bulletin,* 63(6), pp. 384-399.

Tversky, A. & Kahneman, D., 1974. Judgment under uncertainty: Heuristics and biases. *Science (New Series),* Volume 185, pp. 1124-1131.

Voss, C. & Raz, T., 2016. *Never Split the Difference: Negotiating as if Your Life Depended on It.* 1st ed. New York City, USA: HarperCollins Publishers.

9. APPENDICES

9.1. INTERACTION DYNAMICS NOTATION

Symbol	Name	Description	Example
A	Move	A 'move' indicates that a speaker has made an expression that moves the interaction forward in a given direction.	A: I need to buy Legos (at) home. Think about how therapeutic it would be.
? A	Question	A question indicates an expression that elicits a move. A question projects onto the next response and constrains the content of that response because the next response needs to answer the question.	A: Where should we start?
B	Hesitation	Hesitation indicates an expression that is drawn out over time and is not completed. It denotes self-inhibition on part of the speaker.	B: Yeah or not erm (0.8s) there's something erm (1s) when we give (0.4s) yeah.
C	Block	Block indicates an obstruction to the content of the previous move. For a block to be felt, the coder needs to feel that the response in someways obstructed the flow that was established by prior moves.	B: Maybe have something which looks like a computer but you can just type your name or do a simple math, a calculator in the shape of a computer kind of. C: Er, but I don't know, I mean, considering the age segment we are targeting 3 to 7 years.
B	Support for move	Support-for-move indicates that the speaker understands and/or agrees with the previous move.	C: Safe and entertaining (bending forward to write). B: Safe and entertaining, yes.
C	Support for block	Support indicates an acceptance of a block by another person.	A: But that's also, I think that's already done. C: Yeah, its already there. B: Ok.
B C	Overcoming	Overcoming a block indicates that though a block was placed in front of a move, a speaker was able to overcome the block and persist on course of the original move.	C: Er, but I don't know, I mean, considering the age segment we are targeting 3 to 7 years. B: So 7 years they go to school, they would learn A,B,C right?
A B	Deflection	When a speaker blocks a previous speaker's move, that speaker or another can deflect the block with a move that presents an alternative direction for the interaction.	B: So when you say we need to divide the age-group, but you cannot have like 3, 4, 5. A: No, no of course not, but I mean you might have a few different (concepts).
X	Interruption	An interruption is indicative of a speaker being interrupted by another speaker or at times by himself.	B: Should we start generating some concepts now? A: Yeah (interrupted by X) X: 10 min are gone.
C	Yes and	A move is considered to be a 'Yes and' to the previous move if it accepts the content of the previous move and adds on to it.	A: What about... if we made a toy that incorporates girls and boys. Its like a house that has a car with it kind of like enables the guys to play with the girls? C: I think that's a good point to have some sort of a educational point in it.
C	Deviation	Deviation indicates a move that changes the direction of the conversation from the one implied by the previous moves.	C: But we need to remember it. C: This is not the buildable room (deviating from previous topic)
A,B	Humor	Humor indicates instances of shared laughter in teams.	A: I don't know I probably would have swallowed but (All of them laugh)

9.2. IDEAS GENERATED BY EACH TEAM

- Ideas incorporated into the final design are denoted with a "✓".
- Ideas that are still under consideration for integration into the final design are denoted with a **"?"**.

9.2.1. GROUP A

Idea #	Brief Description
1	Yarn feeds through gap, then tension is added to the yarn with a handle moving everything to the top.
2	Ribbing along radius and ulna with metacarpals.
3	Knitting with fibreglass elements included, before soaking in water to make it go hard. Also includes a urethane epoxy.
4	Needles controlled by actuators and kept in line with feed. Actuator then moves yarn feed, which goes across needle beds.
5	Inlaid elastomers included in a uniform knitted structure.
(✓) 6	UV curable glass fibres are inlaid into the final knit automatically by the machine.
7	Actuators control head movement. A yarn feeder then feeds yarn to a bed which is below, allowing the space for the knitted structure to grow.
8	A knitted structure that varies throughout the knit, which means a changing density and breathability.
9	An application of the UV epoxy tubes after the structure has been made.
10	A robotic arm that knits by moving in a circular pattern.
11	Yarn feeder that works like a traditional 3D printer (building from the bottom up).
12	Simple stockinette stitch that is seamless with elastomer fibres laid in.
13	Low temperature, perforated thermoplastic pellets or films.
14	A variation of idea 3, that uses UV instead of fibreglass wires to make the structure go hard. This also includes a gelatine coating for the structure.
15	Memory shape fibres that act one way (they make a shape, but cannot be reverted to their original state).
16	A variation of idea 15, but with memory shape fibres that act in two ways (they make a shape that can be reverted to its original state, making it reusable).
17	A symmetric rib that tucks stitches into the wale.
18	A mesh pattern that is laid in elastomers, allowing the overall knit to be more open.
19	1 x 2 ribbing and tuck stitches are incorporated into the design to create extra elasticity.
20	Thermoplastic pellets that are placed in hot water to make it go hard.

9.2.2. GROUP B

Idea	Time	Brief Description
1	1:32	Bamboo sits on rollers, meaning it can be held and rotated with one hand while the saw is operated in the other.
(✓) 2	2:44	Spring loaded "depression" collection system that collects the cut pieces of bamboo and holds them.
3	2:58	A variation of idea 2, but with a pulley system instead of being spring loaded.
4	4:40	A G-clamp holds the bamboo in position, while a circular saw cuts the culm to the size required.
5	5:33	An alternative way of cutting the bamboo that does not use traditional cutting methods.
(✓) 6	6:10	The use of a "guillotine" cutting method to cut the bamboo.
7	7:55	The use of a "crossbow" mechanism to cut the bamboo.
8	9:50	A mechanism that pushes the bamboo over the culm, not the other way around as the majority of ideas do.

9	11:00	The cutting tools are arranged in a radial pattern, with rollers on springs acting as devices to get the bamboo in the centre.
(✓) 10	14:45	A cutting device that splits the bamboo culms into two pieces after the initial cut is finished.
11	15:23	The bamboo is loaded at 45°. The machine must be naturally inclined for this to occur.
(✓) 12	18:18	The bamboo is stripped after cutting, before dropping into a container.
13	18:35	A variation of idea 12, with a waste chute instead.
14	19:24	A disposal system that acts independently to the main machine, instead of being incorporated within.
(✓) 15	24:04	A small, rotating ball belt conveyor system that moves the bamboo without a need to touch it.
16	25:40	A catch for the cylinder, with an open tray or a net.
17	26:40	An alignment system that uses a guide rail.
18	27:30	A guide rail, and a plate on an actuator. These combine with a roller system to align the bamboo.
19	34:28	An actuator that extends like a telescope, allowing alignment of the bamboo.
20	39:10	A large amount of feeding rollers that have blades lined up on the sides of the machine.
21	39:44	A table saw that is operated in the base of the machine that both cuts and aligns the bamboo.
22	39:55	A feeder roller that has a table saw acting in a vertical plane with an actuator operating it.
23	42:30	A double band-saw, allowing eight bamboo culms to be cut at once.
24	47:00	A separate, portable machine that deburrs the cut bamboo culm.
25	47:54	A container that is filled with boric acid that deburrs the cut bamboo culm.
(✓) 26	54:18	A deburring method that uses a sanding technique, such as sandpaper or a high frequency turner.

9.2.3. GROUP C

Idea #	Time	Brief Description
(✓) 1	1:27:00	The robot for movement is ceiling mounted using six bolts. Flanges hold this all in position.
2	1:27:58	A variation of idea 1, depending on how easy it is to drill into the ceiling. A plate is bolted to the concrete, with the robot being bolted to the plate.
3	1:28:44	A steel framework is bolted behind the wall to allow mounting.
4	1:29:05	A crane mostly mounted behind a wall, with a counterweight on the other side. The counterweight balances the weight of the robot.
5	1:29:54	A floor mounted robot that makes use of the available room by having a four jointed arm.
6	1:30:20	A variation of idea 4 that uses actuators instead of counterweights.
(✓) 7	1:30:50	Robot moves up and down in the vertical plane via an actuator. Only one axis of movement is available.
8	1:31:02	Heavy, six axis robot that has four contact points for movement.
9	1:31:48	Variation of idea 8, with the robot moving tin the “y” and “z” axes, not the “x”.
10	1:32:12	Variation of idea 6 that uses hinges instead, allowing only one degree of freedom.
11	1:32:40	Actuators combined with a spring dampener system, allowing four contact points.
(✓) 12	1:32:50	Variation of idea 11, using three contact points instead of four with no spring dampeners.
13	1:33:40	Two pin joints that allow the table to be moved closer to and further from the surgeon.
14	1:34:00	A combination of ideas 7 and 13.

15	1:35:00	Variation of idea 7, but floor mounted instead of wall mounted.
(✓) 16	1:38:38	A "jaw-end", USB style table loading system. Screwed in, but when unscrewed can be pulled out.
17	1:39:06	A weight distribution method that uses pins and latches.
18	1:40:03	A table mounted at the bottom on a large plate.
19	1:40:32	A manual hinge system on the bottom of the table, allowing two degrees of freedom.
20	1:40:50	A variation of idea 19, but with the hinge only allowing one degree of freedom.
21	1:42:00	An extendable table, allowing the shape to be changed.
22	1:42:43	Table with struts on corners to mount tools.
23	1:43:28	Table that is simply slotted into the robot, relying on friction to make sure it is secure.
24	1:45:50	A modular, rectangular table.
(?) 25	1:46:17	Supports on the edge of the table to hold animal limbs.
26	1:46:22	Table with raised lip, to prevent animals slipping off edge.
27	1:47:12	Positioning surface (i.e. memory foam or "pins") that put the animal in the position required.
28	1:47:30	Drain on the end of the table to remove unwanted fluids.
29	1:47:48	Table with warming plates included in the design.
(✓) 30	1:48:22	Table that is a non-standard shape to allow for ergonomics and aesthetics.
(✓) 31	1:58:58	Automatic control system based off Xbox Kinect.
32	1:59:30	An automatic control system that uses a joystick to move the table.
(✓) 33	2:00:00	An automatic control system based off leap motion.
34	2:00:40	An automatic control system based off light gates.
35	2:00:50	An automatic control system that uses a remote to move the table.
36	2:01:32	A manual control system that has control buttons on the table.
37	2:01:39	A manual control system that has control mechanisms on the table.
38	2:05:22	A manual control system that has a foot controlled system.
39	2:05:29	A manual control system that has a touchscreen UI.

9.3. CHI-SQUARED TESTS

The following abbreviations will be present in this section:

Symbol	Meaning	Definition
α	Significance level	The probability of the study rejecting the null hypothesis, given that it were true. At $\alpha = 0.05$, there is a 5% chance of wrongly rejecting the null hypothesis.
Critical value	Critical value	The number that the chi-squared value must be larger than in order to reject the null hypothesis.
D.O.F	Degrees of freedom	The number of values in the final calculation that are free to vary. Calculated using; $\boldsymbol{D.O.F} = (\boldsymbol{no.of\ rows} - \mathbf{1}) \times (\boldsymbol{no.of\ columns} - \mathbf{1})$
Expected value	Expected value	The value that would be expected in the experiment if it the experiment acted perfectly.
H_0	Null hypothesis	The hypothesis that there is no significant difference between specified populations, with any observed difference being due to sampling or experimental error.
H_1	Alternative hypothesis	The hypothesis that there is a significant difference between specified populations.
Observed value	Observed value	The value measured in the experiment.
χ^2	Chi-squared	A measure of how well the observed data fits the expected data (with smaller numbers indicating a closer fit). Calculated using; $\chi^2 = \sum \frac{(\boldsymbol{O} - \boldsymbol{E})^2}{\boldsymbol{E}}$

1. Blocks in normal and critical episodes (based on Table 4-1)

H_0: There is no significant difference in the blocking frequency in normal and critical episodes.
H_1: There is a significant difference in the blocking frequency in normal and critical episodes.

D.O.F = 2

Episode type	Group			TOTAL
	A	**B**	**C**	
Normal	*16 (16)*	*57 (57)*	*69 (69)*	**142** (48%)
Critical	*17 (17)*	*61 (61)*	*74 (74)*	**152** (52%)
TOTAL	**33**	**118**	**143**	**294**

χ^2 = 0
Critical value (at **α** = 0.05): 5.991 ⇒ insufficient evidence to reject H_0.

2. Support for blocks in normal and critical episodes (based on Table 4-1)

H_0: There is no significant difference in the support for blocks in normal and critical episodes.
H_1: There is a significant difference in the support for blocks in normal and critical episodes.

D.O.F = 2

Episode type	Group			TOTAL
	A	**B**	**C**	
Normal	*1 (2)*	*9 (8)*	*8 (8)*	**18** (55%)
Critical	*3 (1)*	*6 (7)*	*6 (6)*	**15** (45%)
TOTAL	**4**	**15**	**14**	**33**

χ^2 = 4.76
Critical value (at **α** = 0.05): 5.991 ⇒ insufficient evidence to reject H_0.

3. Standard criticism frequency in normal and critical episodes (based on Table 4-2)

H_0: There is no significant difference in the frequency of standard criticism in normal and critical episodes.
H_1: There is a significant difference in the frequency of standard criticism in normal and critical episodes.

D.O.F = 2

Episode type	Group			TOTAL
	A	**B**	**C**	
Normal	*1 (1)*	*7 (7)*	*3 (3)*	**11** (48%)
Critical	*1 (1)*	*7 (7)*	*4 (4)*	**12** (52%)
TOTAL	**2**	**14**	**7**	**23**

$\chi^2 = 0$
Critical value (at $\alpha = 0.05$): 5.991 ⇒ insufficient evidence to reject H_0.

4. Understanding question frequency in normal and critical episodes (based on Table 4-3)

H_0: There is no significant difference in the frequency of understanding questions in normal and critical episodes.
H_1: There is a significant difference in the frequency of understanding questions in normal and critical episodes.

D.O.F = 2

Episode type	Group			TOTAL
	A	**B**	**C**	
Normal	*5 (7)*	*15 (16)*	*34 (31)*	**54** (53%)
Critical	*8 (6)*	*15 (14)*	*24 (27)*	**47** (47%)
TOTAL	**13**	**30**	**58**	**101**

$\chi^2 = 1.996$
Critical value (at $\alpha = 0.05$): 5.991 ⇒ insufficient evidence to reject H_0.

5. Application question frequency in normal and critical episodes (based on Table 4-3)

H_0: There is no significant difference in the frequency of application questions in normal and critical episodes.
H_1: There is a significant difference in the frequency of application questions in normal and critical episodes.

D.O.F = 2

Episode type	Group			TOTAL
	A	**B**	**C**	
Normal	*7 (6)*	*13 (13)*	*6 (7)*	**26** (48%)
Critical	*6 (7)*	*14 (14)*	*8 (7)*	**28** (52%)
TOTAL	**13**	**27**	**14**	**54**

$\chi^2 = 0.595$
Critical value (at $\alpha = 0.05$): 5.991 ⇒ insufficient evidence to reject H_0.

6. Analytical question frequency in normal and critical episodes (based on Table 4-3)

H_0: There is no significant difference in the frequency of applying questions in normal and critical episodes.
H_1: There is a significant difference in the frequency of applying questions in normal and critical episodes.

D.O.F = 2

Episode type	Group			TOTAL
	A	B	C	
Normal	*2 (1)*	*12 (10)*	*8 (11)*	**22** (39%)
Critical	*1 (2)*	*14 (16)*	*19 (16)*	**34** (61%)
TOTAL	**3**	**26**	**27**	**56**

$\chi^2 = 3.156$
Critical value (at $\alpha = 0.05$): 5.991 ⇒ insufficient evidence to reject H_0.

7. Evaluative question frequency in normal and critical episodes (based on Table 4-3)

H_0: There is no significant difference in the frequency of evaluative questions in normal and critical episodes.
H_1: There is a significant difference in the frequency of evaluative questions in normal and critical episodes.

D.O.F = 2

Episode type	Group			TOTAL
	A	B	C	
Normal	*0 (0)*	*5 (5)*	*15 (15)*	**20** (48%)
Critical	*0 (0)*	*6 (6)*	*16 (16)*	**34** (61%)
TOTAL	**0**	**11**	**31**	**42**

$\chi^2 = 0$
Critical value (at $\alpha = 0.05$): 5.991 ⇒ insufficient evidence to reject H_0.

8. No advancement frequency in normal and critical episodes (based on Table 4-5)

H_0: There is no significant difference in the no advancement frequency in normal and critical episodes.
H_1: There is a significant difference in the no advancement frequency in normal and critical episodes.

D.O.F = 2

Episode type	Group			TOTAL
	A	B	C	
Normal	*4 (3)*	*6 (7)*	*9 (9)*	**19** (48%)
Critical	*3 (4)*	*8 (7)*	*10 (10)*	**21** (52%)
TOTAL	**7**	**14**	**19**	**40**

$\chi^2 = 0.869$
Critical value (at $\alpha = 0.05$): 5.991 ⇒ insufficient evidence to reject H_0.

9. Advancement frequency in normal and critical episodes (based on Table 4-5)

H_0: There is no significant difference in the advancement frequency in normal and critical episodes.
H_1: There is a significant difference in the advancement frequency in normal and critical episodes.

D.O.F = 2

Episode type	Group			TOTAL
	A	**B**	**C**	
Normal	*11 (11)*	*41 (41)*	*56 (52)*	**108** (49%)
Critical	*12 (12)*	*42 (42)*	*60 (54)*	**114** (51%)
TOTAL	**23**	**83**	**106**	**222**

$\chi^2 = 0.974$
Critical value (at $\alpha = 0.05$): 5.991 ⇒ insufficient evidence to reject H_0.

10. Advancement frequency for each question type (based on Table 4-5)

H_0: There is no significant difference in the advancement frequency depending on the type of question.
H_1: There is a significant difference in the advancement frequency depending on the type of question.

D.O.F = 5

Response	Type of questioning						TOTAL
	Remembering	Understanding	Applying	Analysing	Evaluating	Creating	
No advancement	*1 (1)*	*16 (13)*	*10 (12)*	*10 (7)*	*3 (5)*	*0 (0)*	**40** (13%)
Advancement	*7 (7)*	*85 (88)*	*79 (77)*	*46 (49)*	*39 (37)*	*1 (1)*	**257** (87%)
TOTAL	**8**	**101**	**89**	**56**	**42**	**1**	**297**

$\chi^2 = 3.557$
Critical value (at $\alpha = 0.05$): 11.070 ⇒ insufficient evidence to reject H_0.

11. Industrial experience and its effect on the level of questioning asked (based on Figure 7)

H_0: There is no significant difference in the level of questioning asked depending on the industrial experience of the design phase leader.
H_1: There is a significant difference in the level of questioning asked depending on the industrial experience of the design phase leader

D.O.F = 1
$\chi^2 = 10.358$

Type of questioning	Experience level		TOTAL
	No industrial experience in NPD	Industrial experience in NPD	
Low level	*27 (19)*	*136 (144)*	**163** (62%)
High level	*3 (11)*	*96 (88)*	**99** (38%)
TOTAL	**30**	**232**	**262**

Critical value (at $\alpha = 0.050$): 3.841 ⇒ sufficient evidence to reject H_0.
Critical value (at $\alpha = 0.025$): 5.024 ⇒ sufficient evidence to reject H_0.
Critical value (at $\alpha = 0.010$): 6.635 ⇒ sufficient evidence to reject H_0.
Critical value (at $\alpha = 0.005$): 7.879 ⇒ sufficient evidence to reject H_0.
Critical value (at $\alpha = 0.001$): 10.828 ⇒ insufficient evidence to reject H_0.