웹 소프트웨어 신뢰성

✓ **Instructor: Gregg Rothermel**
✓ **Institution:** 한국과학기술원
✓ **Dictated: OES** 봉사단

Well we begin this I don't want to be in to soon.

But, it's probably okay.

All right well.

I hope pick in here enough as about as much, volume is I can get out my voice yet.

So, the object of this class and the class that we will hold make up class a week from today.

It's to talk about this material on experimentation and computer science.

That should be useful to you in many ways.

First of, some of you will ,well, in the short term, some of you will be reading about studies for this project.

Some of you need to evaluate them, some of you might be conducting small studies for this project.

And you'd like to know how to do that.

In longer term, most of things we do in software engineering involve what I would call heuristic, there unlike algorithm which can produce exact solutions.

There are not things you can valuate it just with big O notation and showing that some safety property met there are things that you usually have to look at and say how well did they do on real workload.

And that ends up needing empirical studies.

And I mentioned one way many different quailities we can study, usability is one you can study empirically that you couldn't look at any other way.

Effectiveness is detecting the faults, we can compare testing techniques by coverage higher criterion by subsumption for how much they cover but they does say how well they do in detecting fault.

Various things we can try to detect them.

Now, the first thing we are going to think about is that we are empirical studies are concern their several different strategy almost I talk about those.

And then we talk some about measurement.

And then experiment process which is probably next week.

So, there are three overall excuse me this comes later, two overall let's see research paradigms that you will see in capsulated in studies.

Qualitative and quantitative.

Qualitative study look at things in a natural settings.

Looking at people's explanations or things they describe or things they notice so they don't involve so much measurement usually.

Quantitative look at things with much more, quantitativeness comparing two or more groups or treatment, try and identify cause and effect relationships.

So some examples of that qualitative studies might in some sense our um web macro study is qualitative in that it was having people do things and talk about what they were doing in so as the mash up one they will think a lot studies.

Now You can still get some quantitative data sometimes study have concept of both quantitative and qualitative but the qualitative focuses on things that are less easy quantify but still that can be reach in terms of what they give you.

So, will see differences of those probably in different papers that you put force to.

In software engineering in testing and analysis mostly we see quantitative studies.

And that's even more possible when you are not involving humans, if you are looking how do to tools do, how to techniques to do.

And then actually things to compare them quantitatively, but when you move toward how to human use how human do at using approaches that's where you could look at quantitatively or qualitively or both.

And they could be useful for different things.

So we say they are complimentary.

One way to put it is quantitative work looks try to access the effect, whereas qualitative tries to look at belief and understanding.

So there are two broad way looking at studies, there are other different types of investigation you can conduct.

And this is for survey and case study, experiment and what we call quasi experiment.

Now Survey's you've all been involved surveys probably.

They involve interviews, questionnaires.

Don't be some slice on each of these follows.

I'll be brief on this slide.

You probably got on this.

It's a questionnaire and interviews, and people ask bunch of questions you fill them up.

A case study is involves actually being in situation observing things happening and taking some measurements there.

An experiment is typical control study often done in the laboratory where we manipulate variables and view their affects and quasi experiment is like an experiment but has one component missing and I'll get that later.

Typically, experiments are quantitative, then they can get some qualitative data.

Typically, survey and case studies could be quantitative or qualitative, you can image the survey that the.. Well that's really try to just get numbers of thing, number of people, say actually number of say why or trying to get your feeling how the interface work for you.

So, in terms of types, we talk about four types studies here.

For those of you just came back, surveys, case studies and quasi experiment, Surveys typically meant to be descriptive that is to talk about what the care.. find out what the characteristics of something are or to explain thing or to explore things.

But the process for survey is the select some variables you still trying to measure values of some variables just they might be things more easily gotten through to surveys select the sample of people to survey, let me check the something what comes after this, okay good, collect data and analyze it and generalize things, so an example, we could survey ten percent users in a web of the web application uses in a community other opinion of a new web technology.

You infer they overall opinion across the developers they should be developer's developer and understand why they do not use this technology.

So it's important we try to understand why they don't use something, we need to get little beyond do you use it or not we try to get better why so just do you use is less interested in us than why and so the survey would try to have questions that pertain to that.

Now, so variables in this case, am, opinion of some technology why you reason for liking or not, liking the technology and you might find some others to phrase like that, but you figure out what those are and then you frame those questions and get out the variables, then you have select the sample, that's okay, in the survey, keep in mind what you are usually trying to do.

It's characterize an entire community like in this case it says, all developers in the community, you'd like to know above all the people who use this.

Here what they think of course you can never ask everyone there's too many someone respond so you select a subset gathered their answers and you hope that that will generalize all community and that's done all the time in political polls, too.

Survey a subset trying to generalize over the community.

Now, one way that goes wrong is if you haven't picked up good subset obviously.

So, if I go to the young Republican club in US ask who they want to be president I going to get 99% of republican.

That's not obviously, that's a biased sample okay?

You have to select appropriate a sample, and you want to be random across that sample to make sure you got something you can generalize the whole community one.

That's large enough.

There are tools, there are formulas by which you can judge this.

Then you get your survey back and the deal with the data.

Other few ways do this cheat among them is questionnaires and [here] is interviews.

Questionnaire is something on paper on the web these days there is online survey things that I've asked to take various web sites for you click things.

And obviously those are fairly easy to distribute and execute on even before the computer you can easily slip on survey on envelop with returned envelope stamped send to about a thousand households and await them for comeback now you won't get 100% back, and you got a come that in as a factor but you take those back and deal with date you get.

So that's cost of those stamps, it's relatively cheap it to do compare to the next one which is interviews.

Interviews is setting down with people and asking them things.

Then a lot of the personal time face to face time.

So, obviously that's going to be more expensive in them of your time, then coming up with good survey and standing up heart.

But, you do in an interview your higher response rate everyone attends an interview is going to give you acceptable sometimes they may accept the next to interview then you got a something discount them.

Also, you can have different forms of interviews, some interviews you simply ask the questions I've interviewed like before ask the question and I don't understand the questions about, I don't understand blablalbla, And they're not allowed to rephrase beyond a couple of words, let me say again, let me say again, sometimes they stock on them.

But, sometimes they allow to remove in ambiguity into clarify things and make sure you answer the question all the depends on what format they set up.

So, that's a just little brief stuff what surveys happen you start to find them in soft engineering papers.

Um, they are more large used today.

It's in trying to sometimes trying to set up motivation for certain work.

You could survey people to find error rates and types of technology they used and based on the answers you have some motivation for something.

But they are used to other things and there are the entire books this is just one there are dozens and dozens of books because more in the social science sociology in stuff that's the main tool they have, surveys.

So lots of details how to those better in those books if you have need to do them.

Case studies!

That's one actually observe a phenomenon.

Phenomenon lose work meant to be an activity or something people are doing or any specific time in space and trying see what's happening so we could investigate people using a particular develop methodology on a particular project.

Connect Global, Create Future **KERIS**

These people were first one to user UML modified with this certain construct.

How do was that go and you watch them So, it can be useful for dynamic meaning things very over time, or larger study or longer term evaluations we are going to watch a long time or industry evaluation we are going into company, watching out do thing into practice get also be used it, lab things when you got a single project going on among group student or something but they can be used it longer things.

And you can use compare approach if you two testing approaches, approach A and approach B, you can watch group A at the company you approach A and group B you approach B, and see what they do and gather the information.

Um. We'll see how that compare to control experiment as we go on here, they can be not necessarily they can be easier to plan control experiment because we see it they want usually they are not as much control this things we want.

Process if you go to company and your studying some process they are using, you can tell them to keep doing at both make access we quite there is lack of control of there over things or if management comes in tell them why we need step aside some else for a while.

Not you can do.

But, okay so harder to control, or another way in harder to control, if you are working in company you had group A, group B, they can have people living in entering the groups you can do anything about that, groups may differ in their composition in terms of education of the members.

So, I'll control.

So sometimes less easy to that A implies B.

An example: observe the application of a web application development method within an industrial context using two methodologies, over a long period portion of the lifecycle of two development projects.

Now you can put almost anything in there.

Web application testing technique within observe.. a web application testing technique within an industrial context over two projects and see what happens.

And we will find those literatures too in those papers often can be quite interesting report.

Here's the technique we were interested in studying.

Here's the way the company used to do things.

We study both of those.

If there's new one over time they have ten percent fewer defects and you get some interesting data one of that.

If get much harder to say that it generalizes cause you talking about specific company.

There are whole books about this too and this is one reason what I want well.. reason very relative reason.

Reasons about the science and do surveys and that's not our case studies.

And then we get to experiments.

And this is what I focus on the rest of slides here but if your projects.. well

I don't know anyone is doing the surveys but you went out some smoke case studies of some forms.

Okay. Experiments tend to be the most controlled form of study or we set them up carefully manipulate the independent variables in order to view the independent variables like try different techniques to see how that effects fault.

We control the environment.

We try and make it so that nothing else can cause difference between an implication between A and B except for the independent variable.

Experiment should be reproducible and one big issue in this used randomization because that's part of.. good part of controlling things so that you eliminate unwanted effects on independent variables.

And then you usually comparing to the baseline and medical experience, study experience that usually the placebo group, the group that gets no treatment or sugar pills or something.

So more about that in a minute.

But this is a book on that.

That's newer edition that I didn't put that down here if you put it that tile you might.. a later edition.

That I was told it would cost less in this one was ridiculous.

But I haven't checked the price yet.

So you can use experiment to test theories.

That's a big use for them in physics, right? Einstein's theory of relativity, that was just a theory until someone finally found the experiment by which they could provide evidence that theory was probably true.

So you can confirm or reject theories.

You can also consider conventional wisdom.

Conventional wisdom. What is that? My grandmother told me that chewing the bark of walnut tree will help my headache go away.

As it happened she probably right because the bark walnut trees contain an ingredient of aspirin. Okay?

But there's a lot of times grandma says these things where there's really no foundation and true they just need.. I mean I don't know is chew is the best thing when you are sick? I mean I tried it.

I don't think I got better and any fast. I was good.

Test pre-conception things, that people always assumed they are true in used experiment with these things.

Also as we are talking about does A cause B, what effect does A have on B?

It looked relationships between things.

There used to valid actite models.

Here I got model system.

Let's look it whether things protected by model hold.

To validate measures if you got various ways measuring things and those the accurate measures.

Not to some extent a case study can do this too just not with quite so much strength you could ask the same questions in the case study.

As you ask in a controlled experiment just you might not have statistical significance. It's just relative to watch your study.

Ah, text pre-conceptions. A pre-conception, it's kind a like conventional wisdom.

A pre-conception is a thing that you believed to be true.

'Pre' and 'Conception' is an idea.

Kind a like inception if you saw the movie like conception.

Among my conceptions were the idea that such and such.

So a pre-conception is a pre-held idea and in which are we may find that, you know, assuming that the world is flat then I can explain the movement of the planet like this.

Well, what about that pre-conception maybe the cause and effects told analytically.

But what about that pre-conception? Some pre-conception turn out to be valid.

Some turn out not to be.

In some times.. it's interested.. Sometimes when you are looking at a piece of research.. Is there anybody.. you know, it all holds together if they had shown that B falls from A but I don't believe A and so you can look at A and that might be all something useful.

So here's a controlled experiment and I've just tooled these as web testing type things or development things.

Randomly assign a set of web app testers from company C to two groups: one using technique A and one using technique B, and ask them to apply these techniques to a randomly selected set of C's web apps containing seeded faults, and measure the fault, detection effectiveness.

So you have two groups of, now this is a study of people.

Two groups of people doing this.

Now why isn't that a case study? Well, I am not trying to do it over time in their context.

I am.. I'm going to take amount of their you know the work situation put them in the lab for three hours I am going to have all do the exact same thing.

That's what made controlled experiment and I am going to trying you know make the group C equal in terms of their background in experiment.

Let me start with examples first, okay? here's in the software engineering realm and I work from this one.

Just trying get a difference.

So you got company.. it's company C and I have got two groups.. one..two groups in the company.

So these could be two different.. Ah well, let me do it this way.

Let me talk about the case study end of this.

So I said I am interested in two techniques A and B.

These are two testing techniques, okay? And so I go into a some company and I get very interesting and exploring this.

So I get company C.

So I get group C1 who is building some web app, I don't know a name for it, I am just going to call it App1 for doing something, okay?

I don't know what they use before for testing techniques so what they use now but I am going to train them and using A and then over you know this three months period it doesn't have to be but I will put it this way.

They are going to apply that.

We are going to end up and we are going to look at faults detected, okay?

Now C2 there's developing some other web app, App2.

I am going to them and say okay you're going to use I am going to train them B and have them use that for certain period of time.

Maybe there's going to be two months I don't know.

I will measure the faults and we will look at this development each other but it's a kind of.. it's on.. it's in.. it's online thing and it's in middle of their normal work flow.

They are trying to this out.

It's not in the lab.

Things could happen in all of those period months might be entering and leaving the group and more over I mean these.. there's two different apps here involved with two different techniques and the groups have made different training.

So there's lots of reasons why it's hard to comparing contrast the results here in a very scientific way because if they're seeing that B is better than A Well, who knows it might be B's better for applications of this type strongly better than that A is for those.

So there's lots of different things that could cause results.

That I haven't controlled for.

Now the controlled group of that is I go to see, I can't these two groups C1 of those C2 but instead of then being two different working groups, one in building A, one in building B or on different projects I get a subset of randomly chosen.

So I got two random groups so there randomly select so it should be randomization it worked equivalent come up balanced and then I don't give them.

Well if I have to give them the same app, if I have to give them different app so I try to make them very close to this same difficulty and functionality but I ideally I give them the same app.

A and then I am overloading my letters and then I have them try techniques some app for the same amount of time.

I trying to settle variables I can as equal.

Now if I get differences in results I know it's not doing difference in the app.

It's not do the differences in the content of the group as long as my post studies determine my distributed.

Well, it's not do the difference in time.

I get to much greater claim to choose asserting a cause of connection.

So the big difference is control.

Controlling the factors that might intervene in our..

We are attempting to conclude the connection between independent and dependent variables.

We are trying to control any factors that we don't affect that.

That's a controlled experiment and sometimes that's really hard to do and people fall back on the case study where they say "Well, I want to study A verses B I couldn't figure out anyway to do it in the controlled lab.

So here's the next best thing.

Here's my results.

Keep in mind things the group might differ.

They are all different applications." So which really amount of control involved.

There is one another type case study people do sometimes even without people involved.

Sometimes on testing techniques our subjects if you want our programs.

We got two techniques.

We are not interested in how humans' use in applying them to many programs to program K.

If I have a lot of programs that represent the sample, A looks better than B, that tells me something rather strong about A.

Sometimes people will look.. if I just have one maybe large industrial program I can look at A verses B but I must have less power to generalize because just one program.

So some people will call that a case study too particularly if it's harder to control some other aspects, other differences.

So think of case study as.. it's not just very small or some case studies going through long time.

They are medical case studies that study people across decades.

Okay. That's experience, excuse me, quasi experiment technical speaking it's the same experiment but where we have a lack of randomization over subjects or objects.

Explain that more but it's like a testing technique we are getting a lot of test suites not only a single web application that your group created sounds like a case study too but statistic underlying all of these experiments are statistics.

If any of you has statistic courses well, at least statistics somewhere in there you might got in this whole notion in this experiment random variables and independent variables and determining relations and correlations and all and in most of that work in randomization is a key element and they say so one of the first time experimentations software engineering is relatively new compared to other things and I know one of the first times that we did the study, we studied two techniques on seven programs P1 through P7 and we had lots and lots data thousand test suites on each many many data points on how well do the test suites do or cross five hundred bolts.

So lots of data but we just need seven programs that we went randomly selected.

They are programs we could get that our tools functioned on.

They are from many bench mark or something so we went to a statistician and they

said you can't do statistic on these your subjects won't randomly selected.

We don't really know our community how randomly select or not just not randomly, you know you have no idea whether there are representatives set whether there are randomly selected from representatively.

Now with people start medical experiment study a lot how do we get a good representative set.

Understand it's still not easy but it's been studied but we don't really know what it takes to get a representative set a programs I don't know what a representative set is and so in that case it's some extends of some the statistical the assumption is violated and it becomes deadly speaking of a quasi experiment.

This terms started be used for that cause it's the best we can do but we still get data out of it and okay so that we have to be care of generalizing data.

So now having said that I don't see anyone writing experiment saying do investigate this we conducted quasi experiment they just say experiment if you see a stastition that's what we matter.

So these things have different powers and characteristics really got survey studies and experiments.

Execution control means how much you can assort how much control you can excel over how the things going and whether goes the way you want how we turns the result whether the experiment protocol is add here to.

Measurement control is how well you can accurately measure things. Investigation cost how expensive it is.

Is a replication how easy can people do the same thing?

So a survey well as I said the lowest cost and it's easy to replicate cause you just take the same survey document take another sample and send it to them.

But in terms of an execution control once it's out there you have no further control over it and measurement control relatively low because it's hard to find but you can put you can do quantitative measurement on survey responses but it's little harder than in when making quantities measurement of things.

Let's just now look at case studied experiment, the experiment has high execution control that's one of the key elements about it where we control things very carefully well it's case the study it's harder to control in that company situation it's harder to make sure things keep going the way we want.

They both can be very carefully in terms of the measurements they take along the way.

It tends to now this generalization but it tends to be a lot more costly to do experiment in a lab exerting all this control and it does to go into some place and they observe something overtime and its far as replicating if you've once you've written down all the rules used for making study people can replicate whereas how you've replicated this because you know in a case study there are lots of possible factors that you have no control over and therefore you can replicate them.

But you do see that the a cost, a cost things being a important thing here low medium high if you have an idea that you are trying to look at impractically?

And it's a first time and any ones look at it.

It may not be worth the cost of an experiment initially maybe better to do some smaller scale case of study and trying to see whether the data points the near direction before spending essentially more resources.

It's been often say human studies if you see this on both of the papers that I talked about where did small samples of humans more it always did some states there are really more in the case studies site are and say individual with people might argue with that they are certainly quality of data but the cost was lower than trying to get 30 people to sign up and come in for a two hours in a block and make sure all the instruments working correctly so we start it with those.

So how you choose the strategy all the type of research questions, some questions might inter-quantitative of you know what to people think about something is quantitative how much control you will require how much time in space are available and what's your resources are and from that you trying to choose studies what will best help you get out of your question in some meaningful way all studies or types of studies can provide something useful just depend what is you are looking for so it is all your strategy.

Let's get measurement.

Studies are observing things in the world and trying to back into much more quantitative of conceptual things or theories in substance and so measurement when we take something in the world and trying map into some sort of concept.

A measure is a particular number of symbol assigned by this mapping.

A valid measure a measure that captures necessary properties of the attribute so I got a example I can verify.. slow good okay.

Measurement, as a good measurement measure here so it suppose your goals to measure programmer productivity there really is one of first goals of people have in software engineering we are building these system some people built in faster than others.

Who should get the raises, who should be moved into something other development. More map how long we will take to build the next system some people belonging for long time for ways to measure how productive programmers are.

So how productive a programmer is that's kind of a concept.

How productive are you, how productive are you.

But, now I want to be able to say that you know 150% more productive than you?

I need to quantify somehow.

I need some sort to measure. One of the first things people came up with was code size how many lines of code you write per day.

So that's one possibility so if you write 50 per day and you write 100 hey you are 1 half times productive problems.

Now, well when might be effort if two year coding and you know you yours take 3 months is take two year and one and half times productive there are problems with that.

Lines of code is only part what we do previous code and if your code is twice has been bugs in it, it is I'm not sure twice productive there's also other outputs became a coding as well maybe there is specifications documents etcetera so lines of code doesn't see what capture everything and you are one could see we really in ultimately when we ask about productivity we are just in deliver benefit not just you know how fast they produce something what about but how much benefit is delivered to company so that's lines of code and effort equivalent some employees what did I mean by that.

I don't know what that means well obviously if you measuring one you might if you consider how long something takes per part time weeks and they are not working for weeks there are problems with effort and some measure problem but well just looking at.

I get effort in person months somewhere or does it faster but produces more bugs again that's not a great measure okay?

So we say that these measurements if you are interested in program productivity we should be looking at deliver benefit and these simple measures don't succeed and do that and backing up to here we have that four more relational world we are interested in deliver benefit and the mapping involves finding a measure in the world that relates to that and lines of code doesn't do that well so it's not real valid measure it doesn't capture necessary properties.

Another terms that's used for relative measures are objective and subjective.

Now this is kinds to find kind of circularly an objective measure involves of no subjective judgment about the measurement value.

But, still I can use a ruler to measure a length and there's not much subjective judgments for example you know might turn and decide where is the hash mark is but its subjective does involve subjective judgment which can depend on either objectives title? of viewpoint.

So, on objective measure productivity lines a code is an objective measure of productivity it just not a valid measure a delivery date which as do a effort that subjective or there's no subjective value in that right?

It's when it is but it gets valid.

Personnel skill and usability, it's very hard to find an objective way to qualify that people look for we will talk about measures in the next slide or something we will talking about that more.

Usually we prefer then objective measure sometime we find there just isn't one and we have ability of subjective measures and that reduces the I think something of construct threat ability to our measures that we have to acknowledge technology.

One other we talk about in measures is direct versus indirect.

A direct measure of an attribute doesn't involve measurement of other attributes.

An indirect measure does involve often only indirect measures are available.

Till straights that something that I like to talk about a study of a there's a study done a few years ago when I was a bachelor to be read this study it's a long long ago but they were trying to determine whether I think it's whether photos of certain photographs and certain situations would make people angry are you and had probably had to do that extra or something.

Umm so they wanted to show a bunch of people they had a bunch of people they're going to show some of them this group of photos or some of this group of photos and see whether people experience anger but how do you measures anger?

You could ask are you angry are you angry now you are asking you're self-reported ok?

But if you don't want to do that cause that has some drawbacks when the people consider sort of angry do you want to look at physiological anger responses how would you do that?

At that time there was some evidence that people were angry their palms were sweat you get perspiration in your palm.

So in this study they hope to people up to this measurement and they show the 2 groups the groups of photos and they measured the palm temperature in perspiration and they started what degree in the two group of angry.

This all depends on whether this indirect measure of anger palm perspiration is a good tracks well with anger is a good indirect measure cause it's not a direct measure angers so there are always psychological things that there are no direct way to measure happiness how do I measure that so there we need indirect measure and that can often be true so that's often only indirect ill what you want to be aware of that because once you are indirect your degree away from what you are trying to measure anyone do you have a proper measure question a direct measurement suffer the effect of testing of direct measurement could be number of defects it's if you are able to classify the number of defects found where the technique that's a direct measure or and if you want to use that to talk about deliver benefit that's a part of a deliver benefit.

Part of deliver benefit indirect measurement programmer productivity so it's a lines of code is objective but indirect it's not really a thing we are trying to measure and part of its indirect in addition to be indirect it's also just not valid.

Measurement essentially needs things where in studies we are trying to draw some let me look at it oh good we get to this in a minute draw some conclusions about the world.

Draw some conclusions about the world based on measurements.

If we don't do that well then that's the problem for us.

Ok experiment process… What we are looking for.. with the properly designed experiment is way to control of our subjects or objects.

Subjects are used to refer the people.

Objects might be thing like program, test cases and instrumentation.

So that we can draw more general conclusions or more correct conclusions about causality.

Properly experiment also provide ability to use statistical testing or hypothesis testing where we can assert that results really are statistically meaningful.

And if the properly designed, they allow replication by others.

I would just say something about replication means.

It doesn't necessarily mean doing exactly the same thing in exactly the same way.

And seeing if you can get the same results.

It could mean that.

Some years ago there was group of people I think in Utah who said they did the running experiment that it proven they can achieve cold fusion.

Cold fusion.

Fusion is supposed to be the greatest thing to replace the vision if we can manage it.

I don't know too much about this but cold fusion that was like Oh wow!

They did that! The whole world is saved.

So people try to do exactly what they did to the letter and they couldn't replicate it.

That was the case they are really try to do everything exactly the same in other laboratory and they couldn't do it.

So there must be something that they didn't report right or didn't hold constant but science does this replicates things.

But another type of repl….. that's… there is another word for that … well it's kind of like repetition.. another form of replication though really means running substantially the same experiment but with a few different qualities.

A medical example works well.

In the medical experiment I've said we might …One of our first human study might be a small group of humans.

And we try and get…

Let's just say it's small group of graduate students working with the hair loss thing right?

Graduate students are willing to earn a few 100 box to experiment with their hair loss thing so..

So we get two dozen graduate students experiment with size, hair loss, remedy and see what happens.

Now, if we get some results out of that, we've only studied couple dozen graduate students.

We want to be able to generalize.

Applying the same study protocol to a whole different group of people is a replication

too.

I mean, it's not the same people.

So it's not a repetition.

You can't do it.

You know their hair didn't.

But it's all different group of people.

Maybe they are not just graduate students now.

Many people under different circumstances.

So, in our case replication might be well our first study was 7 Seimmence programs, next one was on 10 Linux utilities, Nest one was on too large something elses.

So that's the way of do replicational.

And not just subjects.

A first one could be… you can also change the conditions under which things old.

Our first thing was on Java programs, our second ones were C programs.

But now most rest of what I'm talking about will be how to get a proper experiment design.

How far I go here, just few more slides today.

This is from the volume book that sited few slides ago.

And I like this. I'll come back to it little later but it's trying to explain what's going on.

On a high view, theoretical view when we do a experiment.

Because as I said before, it's not the same as the theoreticians do where they come up with the theory and prove it and they are done.

It's not a mathematical thing.

It's done by observation in the world.

So someone has a theory that A causes B and someone has an observation in the world and and… If you got a theory and you want to investigate whether it holds in the world.

You get to do it based on some observations.

Oh this only works in the examples.

I try to remember my examples.

I think the theory of relatively…says that hmm…. Well it price lots of things that I mass this up well there's no physicist.

Physicist listening but I think theory there is that the gravity can cause light to bend.

Among other things which was not thought to be true before. But gravity is pretty difficult..

Pretty theoretical construct and how you going to assess that and so to … you know…I mean I observe that Who can tell me the theory of relativity? Anyone? Better knows it than me? Alright.

Main thing I remember was… The theory of relatively did imply some several things in hold if this theory holds then these things will hold.

And one thing it said was if the theory holds, then gravity will cause light to bend.

And so in the experiments.. So there was this theory essentially said something…

One component of theory said that light could be cause to bend when people experiment with ..

Well actually, there's a movie about this but they actually looked..

How are you going to observe like bending well they determine that if you could view the…. at the time of solar eclipse.

If you could view and that's when usually we can look real close to the sun with the telescope.

And there's star in the distance real close to the sun. If it's close enough, and if the theory holds, and if you takes the photographs in right period of time during the eclipse, the actually appearance of the star will be different place than you physically know where it is.

And you already know that from other information and that essentially proves that the sun is caused light to bend. Ok?

So the experiment is went to ten years or something like that.. somewhere unexpected..

The time of expected eclipse, it was cloudy and cloudy and cloudy and just the moment and they be able to get in there.

And get the photos and they observed this fact.

So what happened there.

In the world the independent variable a here had to do with essentially this is… I'm not sure what it is right now…

The presence of the sun next to the star because another one might not.

The dependent variable was whether the appearance of the star was in the different place but by looking at this effect, they will then be able to prove the theory.

Now let me try and make this…

Bring this more into a realm where we are familiar with. So… ok…

I might have a theory that says Code-based test adequacy criteria improve software dependability.

That's a theory.

It's pretty general.

Not specifying the particular one.

Now I suppose you want to investigate with that theory holds in the real world. Well, you are going to look at some specific cases.

You are going to have to pick a code based criteria.

You are going to have way of… way of assessing dependability, so what that is saying.

I'm going to have to change size of this now.

The original theory says that cause construct is useful code-based criteria and the theory is asserting that that will have an affect on dependability.

That's all above the world of observation that's the very general theory.

Now you look at the world in particular and you got a pick a particular treatment like a particular criteria.

That's not your independent variable, you can try couple of them.

But that's your criteria and you are going to looking at whether that improves dependability in terms of some measure.

Now you could look at faultiness.

You could look at mean time to failure.

There's different measures you could choose.

And so you apply the treatment.

And look at the affect on the outcome and if you find the cause of relationship now you can assert with the theory appease to the hold.

Now what you got a keep in mind is if you only try the couple of criteria, you can't generalized all criteria.

So you'd like to prove the theory.

But using only one experiment you could only partially gain support for the theory.

So, we will talk later about what that means but that's going to be an external of validity how much can I generalized.

I'm trying to assert a theory by looking at limited set of observations.

So how much can I really prove..

How much can I generalized, it's going to be external of validity.

My measurement is fault detection effect of this is that truly what I mean by dependability? How good the measure is that? So have I….

How well I've done in the asserting my theory depends on how I set of my experiment and how well it models the theory.

That's what we trying to do there in the studies.

So little bit of terminology before we end today.

We viewed this before here in quite a bit.

So a variable in experiment..

Variable is entity that can change and take on different values same as statistics right?

Independent variables are the one that are manipulated and controlled for.

There are some different terms here.

Factors also known as factors.

I'm going to use this terms little bit different. I'm going to say independent variables are variables that are manipulated and that any other variables that could affect the dependent variable are factors.

But they are not independent variable.

Dependent.

Other ones we are viewing affects on.

So you vary independent variables to see the effect on the dependent variables and if there other factors may affected it.

You need to do something about those.

So if we have independent variable, particular value of independent variable is treatment.

So well we'll see more examples later but you going to various..

Independent variable is … ok. Independent variables is testing technique, you could use multiple different testing techniques each of those are treatment and view the result on some dependent variables.

End up the signing the treatment subjects and objects.

Subjects usually uses for people.

Objects for artifacts.

And we saw this and studies where we talked about little bit about between and within subjects studies where or you could have everyone use the treatment so you could sign.

You could use it as tools you sign that to one group and you sign that no tool to the other. You 100% sign the treatment to them.

And test or trials combinations of treatment apply to the subjects and objects. And that's going to be end of that for today thankfully… out of my voice.

Next week we will start with the… We will finish the second and third sets of slides.

So come with this that go into the rest of experiment process.

So, between now and next week you've got two things to do.

Send me the paper information into Friday which paper you are going to present and preferences on when and the proposal ok? Questions and Comments? Yes please

Q: When is the due date? due date is Tuesday.

Sent just by e-mail Tuesday.

If you have earlier, send it earlier talk for someone you know what you are going todo so… Others?

Ok Thank you.