



ESTIMERING AV SYSTEMUTVIKLINGSARBEID

Stein Grimstad (steingr@simula.no)

Oslo, 15. mai 2008

Om oss

- Simula Research Laboratory, IT-Fornebu
 - Grunnforskning innen communication technology, scientific computing og software engineering.
- BEST-prosjektet (Better Estimation of Software Tasks)
 - "...improve expert judgment-based software effort estimates and cost uncertainty assessments through better processes, process support and better learning/training processes"
- Mer informasjon, inklusive forskningsartikler, er tilgjengelig på www.simula.no/best

Hvorfor forsker vi på estimering av systemutviklingsarbeid?

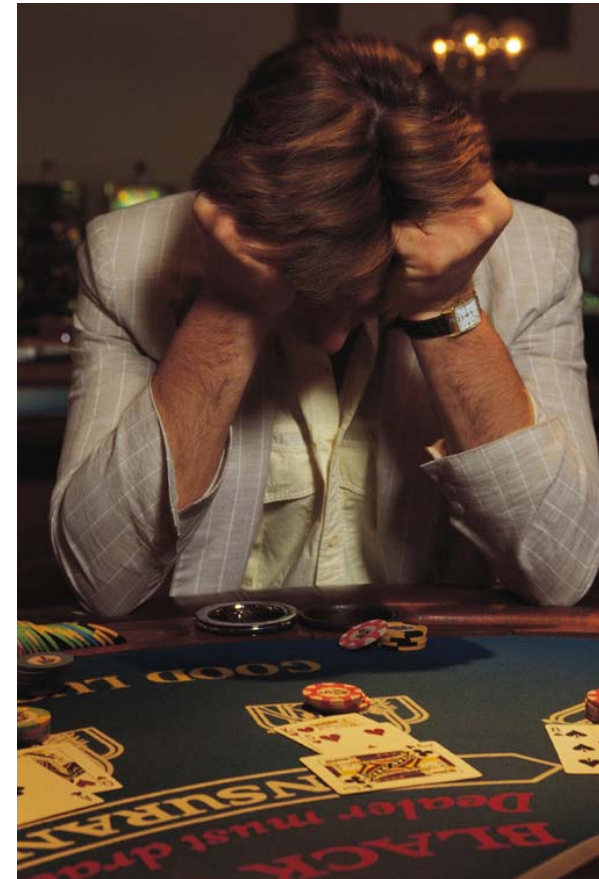
- I 2007 er estimatene tilsynelatende like unøyaktige som for 30 år siden
 - Undersøkelser viser at, da som nå, er gjennomsnittlig overskridelse på rundt 30% av estimert arbeidsmengde [*]
- Unøyaktige estimerer kan ha svært negative konsekvenser både for kunde og leverandør
 - Store økonomiske tap, software av lav kvalitet, forsinket time-to-market, konkurser, stress, m.m.



[*] "A Review of Surveys on Software Effort Estimation", Moløkken and Jørgensen, ISESE, 2003

Hvor stort er forbedringspotensialet?

- Det er urealistisk å forvente perfekte estimerer fordi:
 - Det er en kompleks og dynamisk sammenheng mellom faktorer som påvirker utfallet av systemutviklingsprosjekter
 - Det er en iboende usikkerhet i systemutviklingsprosjekter



Det er et potensial for forbedring

- Vi er inkonsistente når vi estimerer [*]
- Det er systematiske skjevheter i estimatene [**]



[*] "Inconsistency in Expert Judgment-based Estimates of Software Development Effort", Grimstad and Jørgensen, JSS, 2007

[**] "Practical guidelines for better support of expert judgement-based software effort estimation", Jørgensen, IEEE Software, 2005

OPPVARMING:
EN LITEN OPPGAVE

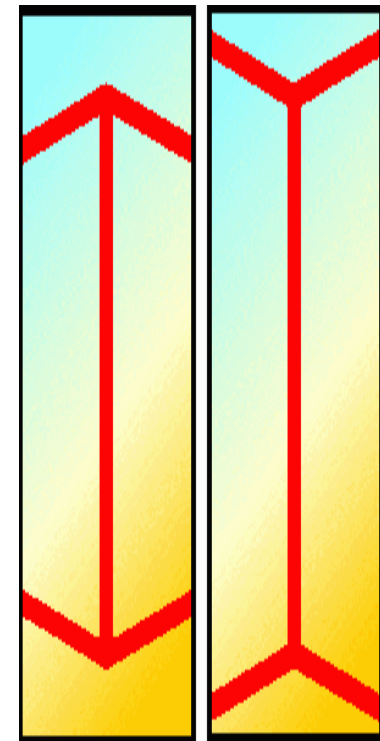


En viktig årsak til estimeringsfeil:

Irrelevant og misvisende informasjon

Irrelevant og misvisende informasjon

- Studier innen andre forskningsfelt har vist at å inkludere irrelevant informasjon i input til estimeringsprosesser kan være en viktig kilde til unøyaktige estimater og vurderinger
- Kravspesifikasjoner og annen informasjon som brukes for å estimere systemutviklingsarbeid inneholder typisk
 - Mye estimerings-irrelevant informasjon
 - Noe misvisende informasjon (tilfeldig eller for å manipulere)
 - Mye informasjon som har lite relevans for estimeringsarbeidet



Våre studier

- Vi har 30+ studier som undersøker effekten av irrelevant og misvisende informasjon
- Størrelsen på studiene varier
 - Små eksperimenter med under 20 deltakere
 - Store eksperimenter med flere hundre deltakere
 - Små oppgaver (del av seminarer, etc)
 - Omfattende oppgaver (flere ukesverk per deltaker)
- Subjektene er vanligvis utviklere, arkitekter og prosjektledere
 - I noe tilfeller har vi studert hele organisasjoner
 - Vi har også brukt studenter i noen studier
- Estimeringsmaterialet i studiet tar som oftest utgangspunkt i reelle kravspesifikasjoner
 - Har vært benyttet i systemutviklingsprosjekter
 - Forenkles / modifiseres



Inkludere bakgrunnsinformasjon med lav relevans for estimeringsjobben

- Gruppe A fikk den originale spesifikasjonen.
- Group B fikk den samme spesifikasjonen, men vi la til estimerings-irrelevant informasjon om sluttbrukernes kontorstøtte-programmer, webdesign, brukeres passord, osv..

Simula organizes several seminars each year. At present, participants register for the seminars by sending an email to a given contact person.

Simula wants you to develop a simple web system (one web page) for registration of participants. The system will handle registration of all Simula's seminars. The participants will register on the web by submitting their email address and a registration code that uniquely identifies the seminar (this code is sent to them by email). The only functionality in the system is to store the email address and the registration code in a database. All queries will be done manually (in sql), and there is no need for any validation of submitted data. There are no security requirements.

The system will run on a webserver that has Tomcat, Java and MySql installed and running. Assume that you are familiar with the relevant technologies. You are free to use development tools of choice.

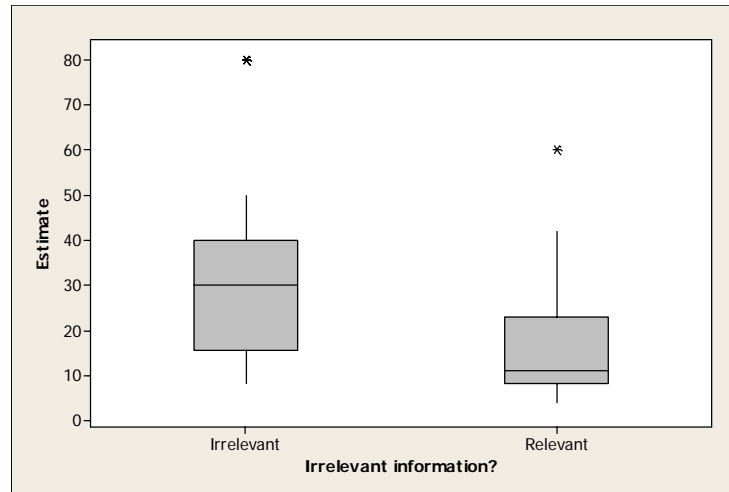
Simula organizes several seminars each year. At present, participants register for the seminars by sending an email to a given contact person.

Simula has ordered a web-based registration system. This system will let participants choose seminar from a list, and then register name, company and contact information. An email that confirms the registration will be sent to the participants. The system will keep track of the number of participants for each seminar and automatically close registration when the seminar is fully booked. There will also be a management module where employees at Simula can log in and query and manipulate conferences. This system will run on a Weblogic server application server, use a Sybase database and be fully integrated with Simula's other web pages. However, due to economical priorities, this system will not be developed before the end of 2006. Simula therefore needs an intermediate system for registration. This system will be thrown away when the new system is ready.

Simula wants you to develop a simple web system (one web page) for registration of participants. The system will handle registration of all Simula's seminars. The participants will register on the web by submitting their email address and a registration code that uniquely identifies the seminar (this code is sent to them by email). The only functionality in the system is to store the email address and the registration code in a database. All queries will be done manually (in sql), and there is no need for any validation of submitted data. There are no security requirements.

The system will run on a webserver that has Tomcat, Java and MySql installed and running. Assume that you are familiar with the relevant technologies. You are free to use development tools of choice.

Resultater



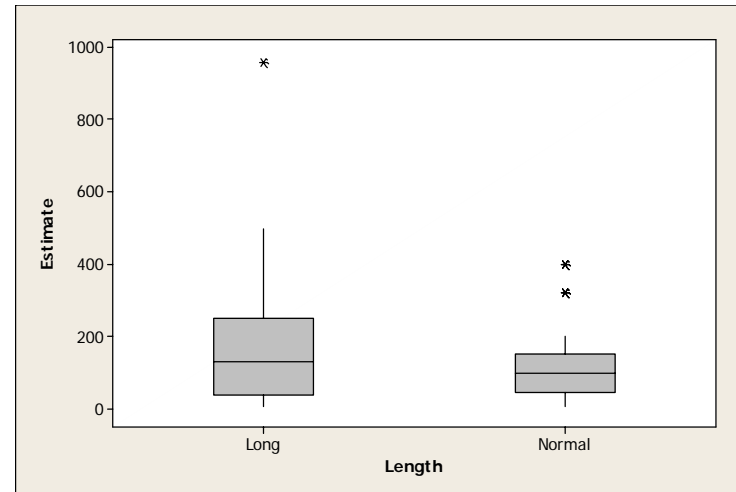
	Irrelevant	Relevant	Difference
Mean	34	17	100%
StDev	20	14	43%

Manipulere størrelse

- IFI-studenter estimerte arbeidsmengde til den samme programmeringsoppgaven
 - Gruppe A: Fikk den originale spesifikasjonen, som var en side lang
 - Group B: Fikk en versjon av spesifikasjonen som hadde identisk tekst, men var på syv sider. Økningen i lengde skyldes dobbel linjeavstand, vide marger, større font-størrelse og mer avstand mellom avsnittene



Result

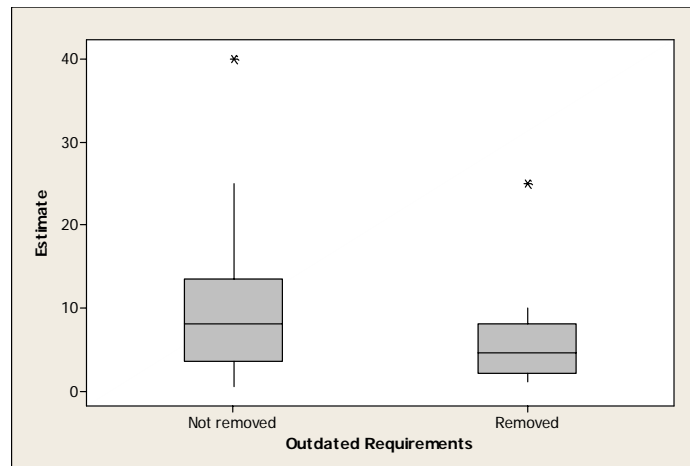


	Long	Normal	Difference
Mean	170	117	45%
StDev	173	98	77%

Ikke fjerne utdaterte krav

- Eksperiment:
 - Gruppe A fikk krav R1-R4, og estimerte arbeidsmengde for å implementere dem
 - Group B fikk krav R1-R5, men de ble fortalt at R5 ikke skulle implementeres. Med andre ord, de skulle estimere arbeidsmengde for å implementere R1-R4 (samme som gruppe A)
- Det var ingen åpenbar sammenheng mellom kravene

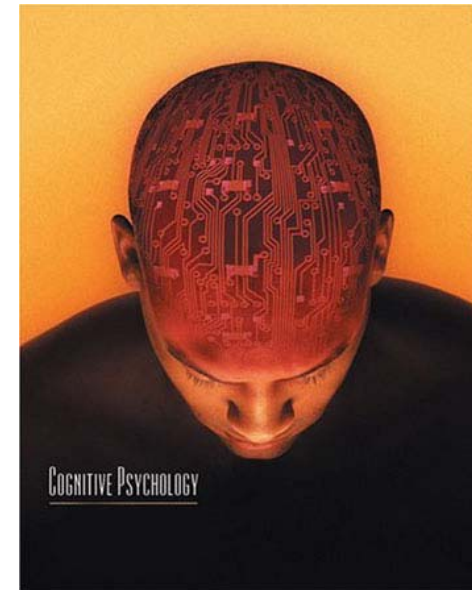
Resulter



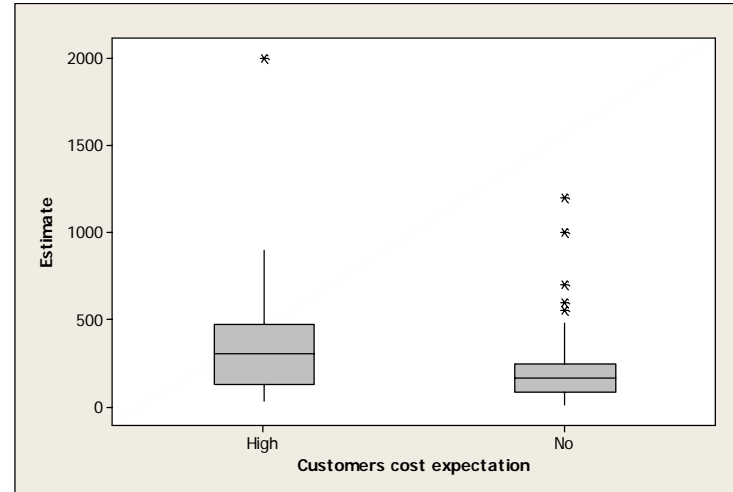
	Not removed	Removed	Difference
Mean	9,6	5,6	50%
StDev	9,9	5,4	83%

Informasjon om kundens forventning til utviklingsarbeidets kostnad

HIGH group: *“The customer has indicated that he believes that **1000 work-hours** is a reasonable effort estimate for the specified system. However, the customer knows very little about the implications of his specification on the development effort and you shall not let the customer’s expectations impact your estimate. Your task is to provide a realistic effort estimate of a system that meets the requirements specification and has a sufficient quality.”*



Resultater



	High	None	Difference
Mean	385	233	65%
StDev	376	242	55%

Kravenes rekkefølge

- Gruppe EASY-HARD fikk en spesifikasjon som inneholdt fire krav. Kravene var sortert etter stigende vanskelighetsgrad (lette først, vanskelige sist)
- Gruppe HARD-EASY fikk den samme spesifikasjonen, men kravene var sortert i motsatt rekkefølge (vanskelige først, lette sist)

$(A+B) = (B+A) ?$

Resultater

- Group EASY-HARD : 46,5 work-hours
- Group HARD-EASY : 65 work-hours
- Difference: 40%

Oppsummering av resultatene

- **De som estimerer har store problemer med å ignorere estimeringsirrelevant informasjon**
 - Dette trass i at de har profesjonell erfaring med lignende oppgaver og det er lett å skille på relevant og irrelevant informasjon.
 - De er derimot gode til å skille på relevant og irrelevant informasjon når de blir spurt om det.
- **De som estimerer er i liten grad klar over at de blir påvirket.**
 - I noen tilfeller har vi i etterkant spurt deltakerne om de tror de er påvirket – og eventuelt hvor mye. Bare et mindretall tror de har blitt påvirket og de anslår effekten til å være liten. Dette indikerer at deltakerne ikke bruker den irrelevante informasjonen bevisst, dvs. at effekten hovedsakelig er ubevisst.
- **Det er vanskelig å forutsi størrelsen av effekten på forhånd**
 - Det virker å være en kompleks sammenheng mellom størrelsen på effekten og faktorer slik som mengden irrelevant informasjon, ekstremiteten, plasseringen, osv.
- **Irrelevant informasjon øker sannsynligheten for uenighet mellom de som estimerer**

Finnes effekten av irrelevant informasjon kun i laboratoriet eller har den også praktisk relevans?

- Eksperimentene har høy grad av realisme i viktige dimensjoner
 - Kjente oppgaver (ekspertise)
 - Oppgavene er basert på reelle kravspesifikasjoner
- Funnene valideres i en pågående studie i en reell kontekst
 - 40 bedrifter betales for å delta (fra typiske outsourcingland)
 - Gruppe A: Leverer estimater basert på spesifisering med irrelevant informasjon
 - Gruppe B: Leverer estimater basert på spesifikasjoner uten irrelevant informasjon
- De foreløpige resultatene indikerer at effekten av irrelevant informasjon er stor også i realistiske settinger



Noen mulige fremgangsmåter for å redusere effekten av irrelevant og misvisende informasjon

- Bruk mer strukturerte estimeringsprosesser
 - Uthev relevant/fjern irrelevant informasjon
- Bruk kompetente personer til å estimere



Bruk mer strukturerte estimeringsprosesser

- Bruk av strukturerte estimeringsprosesser kan gjøre det lettere å skille mellom relevant og irrelevant informasjon
 - Strukturen bør være rettet mot å vektlegge relevant informasjon / fjerne irrelevant informasjon
- Det er ikke sannsynlig at bruk av strukturerte estimeringsprosesser vil fjerne påvirkningen, men det kan muligens redusere den

Uthev relevant/fjern irrelevant information (I)

- Studert i et eksperiment hvor vi la til irrelevant informasjon i kravspesikasjonen:
 - Group IRRELEVANT: Irrelevant informasjon
 - Group HIGHLIGHT RELEVANT: Deltakerne ble bedt om å highlighte relevant informasjon før de estimerte (gul tusj)
 - Group REMOVE IRRELEVANT: Deltakerne ble bedt om å fjerne irrelevant informasjon med en svart penn før de estimerte
 - Group CONTROL : Ingen irrelevant informasjon



Uthev relevant/fjern irrelevant information (II)

Group	Estimate
Group IRRELEVANT	40
Group HIGHLIGHT RELEVANT	40
Group REMOVE IRRELEVANT	35
Group CONTROL	25

Bruk kompetente personer til å estimere (I)

- Vi undersøkte om kompetente utviklere var mindre påvirket av irrelevant informasjon i et studie hvor vi varierte formuleringene:
 - Group LOW: Noen ord som vanligvis assosieres med små oppgaver
 - Group HIGH: Noen ord som vanligvis assosieres med store oppgaver
 - Group CONTROL: Nøytrale formuleringer
- Etter estimeringsarbeidet, så anslo utviklerne selv kompetansenivået sitt



Bruk kompetente personer til å estimere (II)

	Low Skill	Medium Skill	High Skill
Group LOW	65	40	27
Group HIGH	250	80	50
Group CONTROL	180	55	24

Vår anbefaling for å unngå å bli påvirket av irrelevant og misvisende informasjon

- Vi har ingen god metode for å nøytralisere effekten når man først har vært eksponert for irrelevant og misvisende informasjon
- Vi anbefaler derfor at følgende elementer inkluderes i estimeringsprosessen
 - Trinn 1: Preparering av estimeringsinformasjonen
Irrelevant informasjon fjernes
 - Trinn 2: Estimeringsarbeid
Av personer som ikke er involvert i trinn 1
 - Trinn 3: Justeringer
Evaluering av personer som var med i trinn 1, re-estimering av personer som var med i trinn 2



Andre faktorer

- Uklar estimeringsterminologi
 - Skiller ikke på "pris", "budsjett", "plan" og estimat av mest sannsynlig arbeidsmengde
- Lite bruk av historiske data
 - Mangler ofte data
- Lite fokus på læring
 - Sjeldent satt av ressurser

Er du interessert i å få vite mer?

How to Avoid Impact from Irrelevant and Misleading Information When Estimating Software Development Effort, Jørgensen og Grimstad, kommer i IEEE Software.

The Impact of Irrelevant Information on Estimates of Software Development Effort, Grimstad og Jørgensen, sendt til JSS.

The impact of customer expectation on software development effort estimates, Jørgensen og Sjøberg, publisert i Journal of Project Management

Vi søker deltakere til et eksperiment!

- Web-basert
 - Du kan gjøre oppgaven hjemmefra, og når du vil!
- Lite omfang
 - tar under 5 minutter!
- Vi trenger ca 20 deltakere
 - En heldig deltaker vinner en middag for to personer på Eik!





Gruppe-estimering

Eksperiment: individuell vs gruppe-estimering



- Tyve fagpersoner fra samme firma estimerte hver for seg arbeidsmengden for det samme systemutviklingsprosjektet [*]
 - Deltakerne hadde forskjellig bakgrunn
 - Prosjekt var et reelt prosjekt som var implementert
- De delte seg deretter opp i fem grupper. Hver gruppe ble enig om et felles estimat
 - Gjennom diskusjon og kombinasjon av kunnskap

[*] Moløkken-Østvold and Jørgensen (2003): Software Effort Estimation: Unstructured Group Discussion as a Method to Reduce Individual Biases. In The 15th Annual Workshop of the Psychology of Programming Interest Group

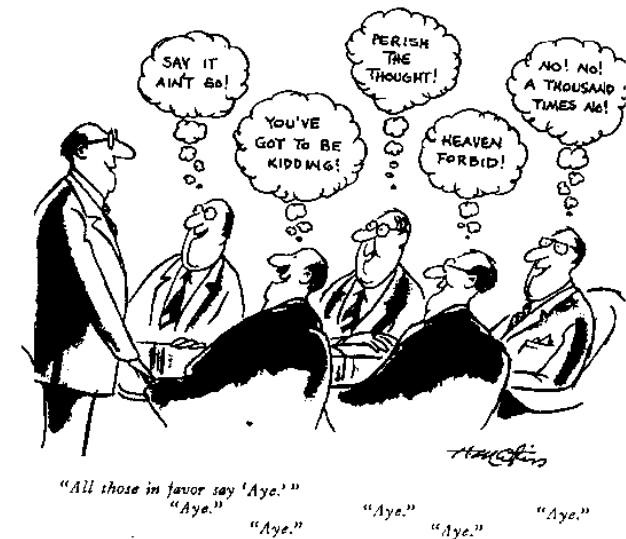
Resultater

- Estimatene som var basert på gruppe-diskusjon var nærmere den faktiske arbeidsmengden enn gjennomsnittet av de individuelle estimatene
 - Mulig forklaring: Gruppens evne til å identifisere flere prosjekt-aktiviteter
 - Mulig forklaring: At de i gruppen måtte begrunne estimatene sine kan medføre at realisme øker
- Vi fant lignende resultater i et eksperiment hvor vi undersøkte usikkerhetsintervall [*]
 - Gruppediskusjoner medførte at man anga mer realistiske usikkerhetsintervaller

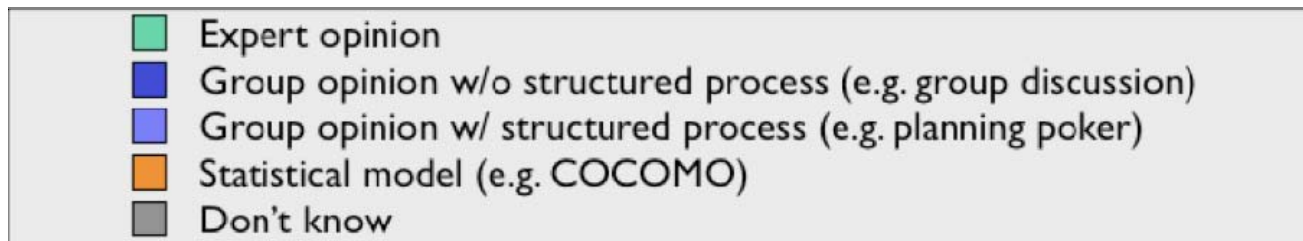
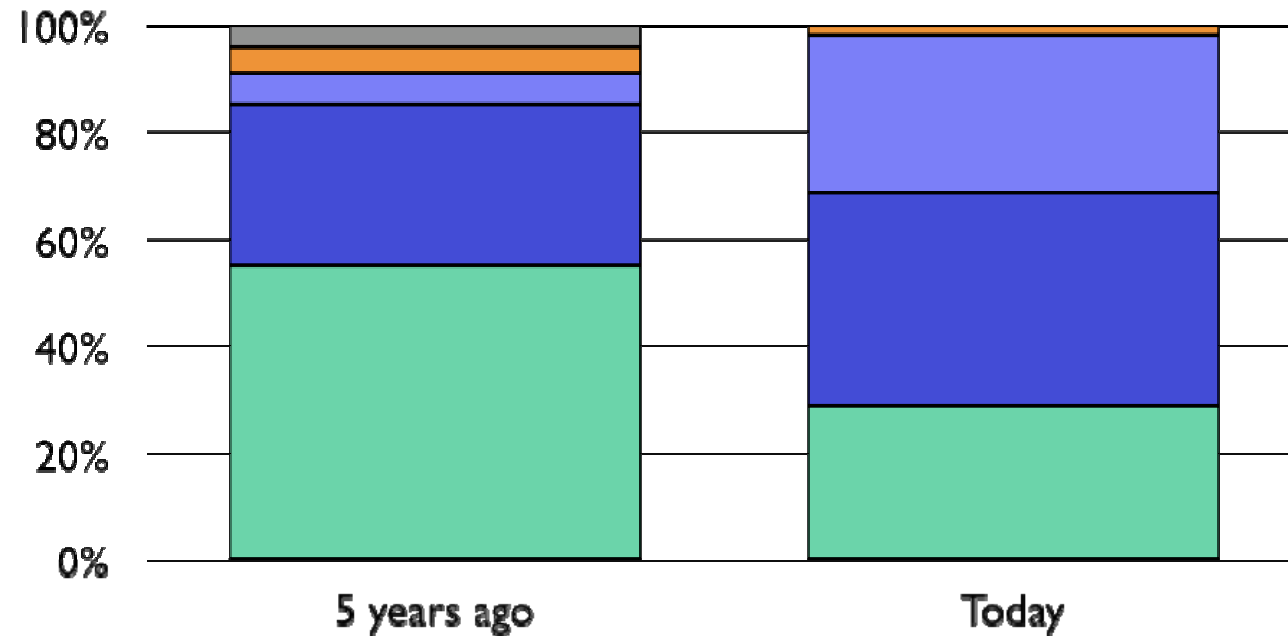
[*] Combination of software development effort prediction intervals: Why, when and how? Jørgensen and Moløkken, SEKE 2002

Forskning på gruppe-estimering

- Få studier innen Software Engineering
- ...men mange relevante studier innen andre forskningsfelt (psykologi, business forecasting, etc)
- Resultater
 - Kombinering av estimater forbedrer estimeringen (spesielt når de som estimerer har forskjellig bakgrunn)
 - Struktur kan forbedre estimeringen (for eksempel: redusere påvirkningen fra irrelevant informasjon)
 - ”Flere hoder husker mer”
- Ulemper
 - Ressurs-krevende (dyr) sammenlignet med individuell estimering
 - ”Group think” kan forekomme (for eksempel: alle er enige med sjefen)
 - ”Group polarization” kan forekomme (for eksempel: gruppen er mer optimistisk enn gjennomsnittet av individene)



Gruppe-estimering vinner frem i norsk IT-industri (undersøkelse på JavaZone 2007)



Strukturert gruppe-estimerings påvirkning på opplevd estimeringsnøyaktighet (JavaZone 2007)

- 50% opplever at estimeringsnøyaktigheten var forbedret
- 30% opplever at estimeringsnøyaktigheten var uendret
- 10% opplever at estimeringsnøyaktigheten var forverret
- 10% visste ikke

Metoder for struktureret gruppe-estimering

- Planning Poker
- Wide-band Delphi



Planning Poker

- Smidig ("Agile") estimeringsteknikk
- Beskrevet av Grenning [1] og Cohn [2]
- Kunden forklarer "user story"
- Teamet diskuterer hvilken jobb som må gjøres
- Alle velger et kort som representerer estimatet
- Alle viser estimatet sitt samtidig
- De med lavest og høyest estimat begrunner
- Teamet diskuterer estimatene
- Gjenta fra steg 3. frem til estimatene konvergerer
- Teamet blir enige om et estimat

[1] J. W. Grenning, *Planning Poker*, 2002

[2] M. Cohn, *Agile Estimating and Planning*, 2005

Planning Poker:

”Hele teamet deltar i estimeringen”

Research:

”Kombinering av estimater kan forbedre kvaliteten på estimeringen, spesielt når de som estimerer har forskjellig bakgrunn”

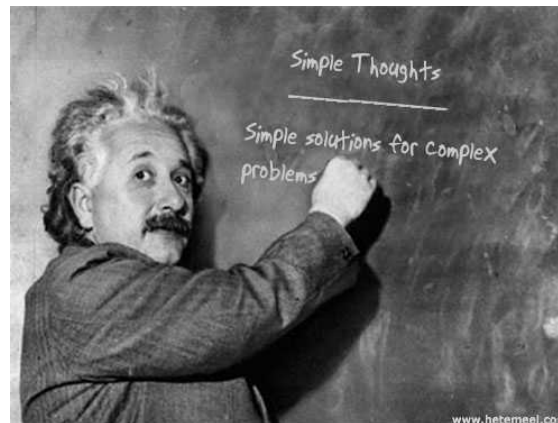


Planning Poker:

”Først diskutér arbeidet som må gjøres, deretter estimere”

Research:

”For komplekse oppgaver er det ofte best å først jobbe alene, for deretter å diskutere oppgaven med de andre”



Planning Poker:

”Den med lavest og den med højest estimat begrunner sine estimater”

Research:

”Mangelen på anonymitet kan øke *group think* (f.eks. alle sier seg enige med lederen)”

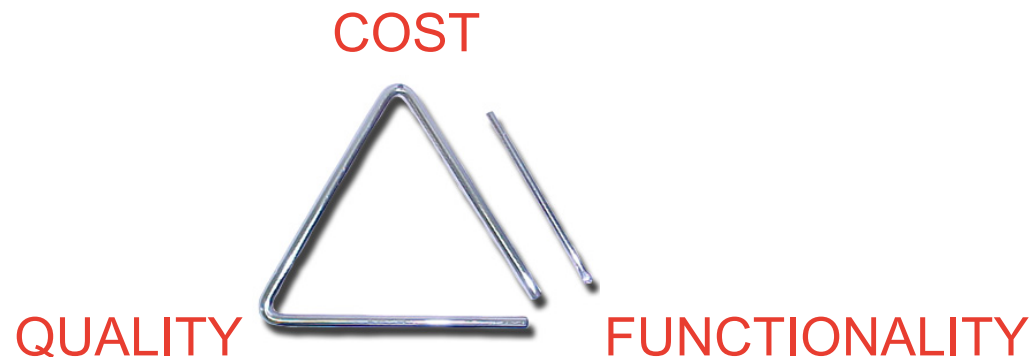


Planning Poker:

”Teamet diskuterer implementasjonen, dvs blir enige om funksjonalitet og kvalitet”

Research:

”Dette begrenser fleksibiliteten, og kan således føre til mer unøyaktige estimater på kort sikt. Men! Gode historiske data muliggjør effektiv læring”

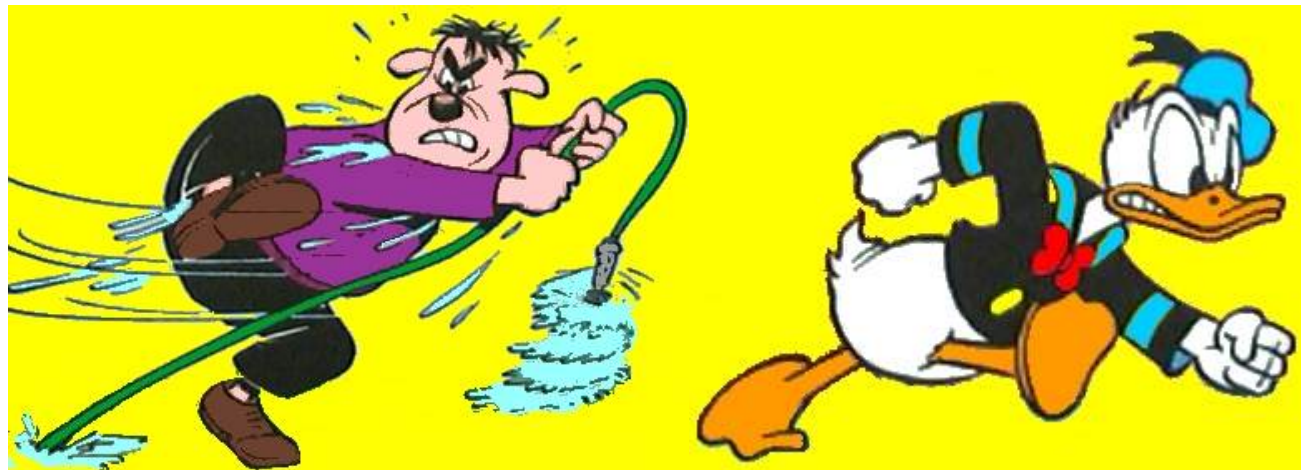


Planning Poker:

”Estimér helt til estimatene konvergerer”

Research:

”Avslutt når det ikke fremkommer ny informasjon i diskusjonen. Dersom det fremdeles er uenighet om estimatet, bruk gjennomsnittet.”

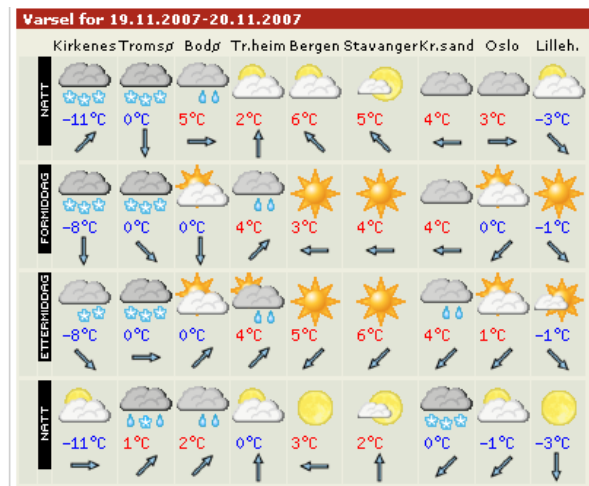


Planning Poker:

”Yesterday’s weather”

Research:

”Estimater av systemutviklingsarbeid er som oftest overoptimistiske. Bruk av historiske data kan øke realismen.”



Planning Poker:

”Planning Poker er morsomt”

Research:

”Motivasjon er en viktig, ofte glemt, suksess faktor innen SPI”





NB! Det finnes få forskningsstudier som undersøker
Planning Poker. Resultater fra andre
forskningsområder er ikke nødvendigvis overførbare.

Industrial studies

	Planning poker vs. unstructured group	Planning poker vs. individual expert
Planning scale	Release planning (2-3 months)	Sprint planning (2 weeks)
Team	8-12 developers	4-6 developers
Automated acceptance tests	Yes	No
Pair programming	Yes	No
Progress visibility	Story cards on wall	Jira
Customer view in session	Business analyst	Developers

Felles for begge studiene

- Moro! Begge teamene fortsatte med dette!
- Mer effektiv estimeringsprosess
- Økt eierskap til estimatene
- Økt ansvar for projektets progresjon

- Men estimeringsnøyaktighet synes ikke å være forbedret