

**IN2002: Software engineering med prosjektarbeid,
27. februar 2024**

Forskningsmetoder / Evaluering av IT-systemer

Dag Sjøberg og Gunnar Bergersen



About Me

- Current position: Professor at University of Oslo
 - Software Process Improvement, Agile and Lean Methods, Software Quality, Empirical Research Methods
- Education:
 - MSc, University of Oslo
 - PhD, University of Glasgow, Computing Science
- Prior work experience
 - National University Hospital (Rikshospitalet)
 - SINTEF ICT
 - Simula Research Laboratory
 - Statistics Norway (SSB)
- Startup
 - Member of steering committee and co-owner of four startup companies



About me

- Current:
 - Associate professor in Software Engineering
- Education
 - MSc (2001) and PhD (2015) from University of Oslo
 - PhD thesis: “Measuring programming skill”
- Prior work experience
 - Programmer
 - IT Project leader of two companies
 - CEO of three companies and Chief Product Officer in Greps (skill testing of developers)



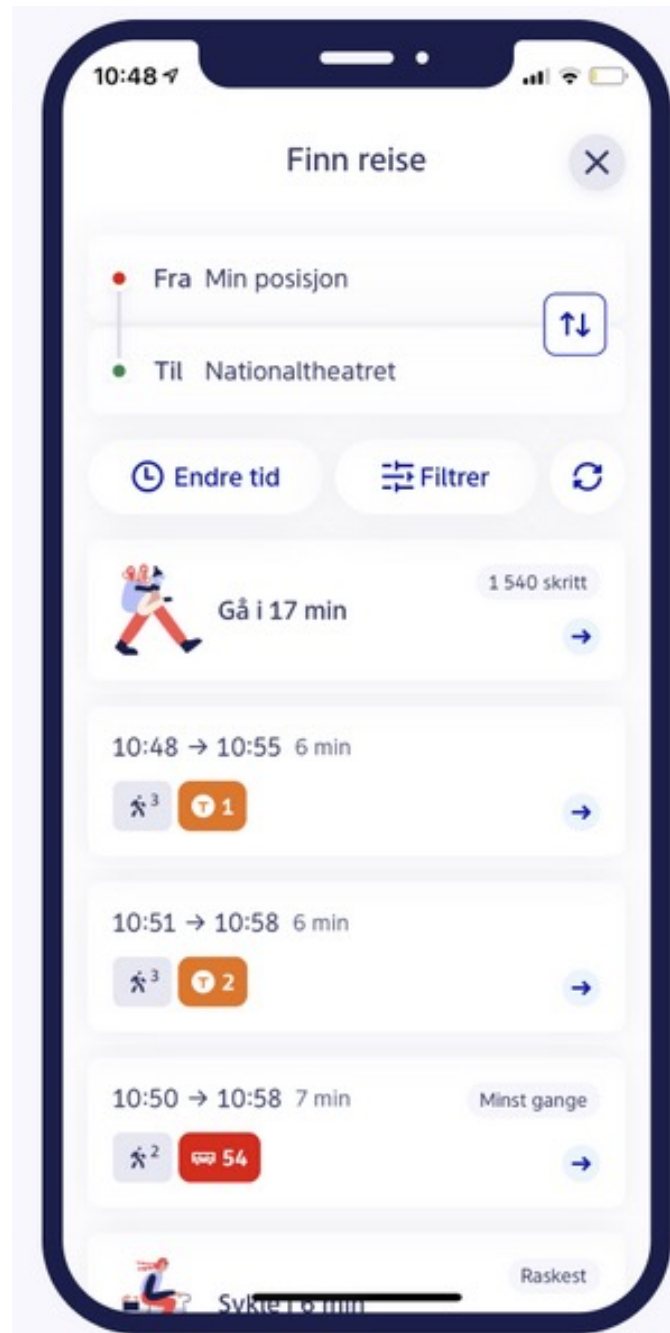
Hensikt med forelesningen

- Lære om evalueringsmetoder generelt
- Få en god app i deres prosjekt ved å anvende metoder for å evaluere app'en
- Få en god rapport som dokumenterer hva dere har gjort og hvordan dere evaluerer det


Ruter-appen

Hva synes du om Ruter-appen?

- Hva er bra?
- Hva er dårlig og hvordan kunne det vært bedre?



Plan

- IN2000 Case 
- Behov for metodekunnskap
- Evaluering av IT-systemer/apper
- Typer forskningsmetoder
 - Eksperiment
 - Case-studie
 - Etnografi
 - Spørreskjema-undersøkelse (survey)

Case-beskrivelser IN2000, 2004

- Case 1. Raketttoppskytning
- Case 2. Småflypiloter
- Case 3. Vær- og farevarsel for yngre brukere
- Case 4. Havvarsel
- Case 5. Åpent case


Rapportkrav IN2000: Produktdokumentasjon

Produktdokumentasjonen forutsetter at leseren har grunnleggende kjennskap til løsningen, slik den er beskrevet i presentasjonen. I tillegg vil det være en fordel om leseren har satt seg inn i funksjonaliteten til løsningen, slik denne er beskrevet i brukerdokumentasjonen.

- Beskriv de viktigste kvalitetsegenskapene ved appen (for eksempel **funksjonalitet, brukskvalitet, pålitelighet** (fravær av feil) og hvordan disse er evaluert (jfr. forelesning evalueringsmetode – forskningsmetode). Redegjør eventuelle brukerundersøkelser og metoder for datainnsamling og analyse.

Hvordan evaluere kvalitetsegenskaper?

Plan

- IN2000 Case
- Behov for metodekunnskap 
- Evaluering av IT-systemer/apper
- Typer forskningsmetoder
 - Eksperiment
 - Case-studie
 - Etnografi
 - Spørreskjema-undersøkelse (survey)

Forskningsmetode = måten kunnskap dannes på

God metode → **Pålitelig kunnskap**

Dårlig metode → **Upålitelig kunnskap**
(mest ekstreme: Fake news)

Hvordan velger man metoder, teknikker, praksiser, verktøy, språk for systemutvikling?

- Ofte basert på moter/"hype" og synsing fra "guruer"
- Velger ofte det man kjenner best
- Ideelt: Basert på empiriske undersøkelser (empirisk = erfaringsmessig, det som baserer seg på hva som fungerer i praksis)

Hvordan finne ut om noe (i IN2000: en app) er bedre enn noe annet?

- Undersøk hva som fungerer best i praksis, dvs. gjennomfør en empirisk studie
 - Eksperiment
 - Spørreundersøkelse
 - Case studier, aksjonsforskning, etnografiske studier, m.fl.

Nytteverdi

- Nyttig å kjenne til systematisk evaluering og forskningsmetoder
 - Forstå og nyttiggjøre forskningsresultater
 - Stille kritiske spørsmål til påstander, inkludert AI-genererte påstander
 - Nyttig både i privat og offentlig sektor
 - Nyttig i alle fag, ikke bare informatikk

Hva betyr det at noe er bedre?

- Hvor mye?
- Hvor mange?
- Hvor stor?
- Hvor gammel?
- Hvor lang?
- Hvor høy?
- Hvor varm?
- Hvor bred, fast etc.

Kvantitative data

- Data uttrykker kvantitet
- Data kan tallfestes
- Brukes i statistikk

Kvalitative data

- Data uttrykker kvalitet på en eller annen måte
- Tekst, bilder og andre formater unntatt tall
- Brukes ikke i statistikk

Plan

- IN2000 Case
- Behov for metodekunnskap
- Evaluering av IT-systemer/apper
- Typer forskningsmetoder
 - Eksperiment
 - Case-studie
 - Etnografi
 - Spørreskjema-undersøkelse (survey)



Hvordan evaluere kvaliteten på app'en?

Hvilke kriterier kan man vurdere tjenesten utfra?

- Tenk på det dere har lært om kravhåndtering tidligere
 - Funksjonelle krav
 - Ikke-funksjonelle krav
- Tenk over kriterier som kan variere mellom ulike interessenter (stakeholders)

Aspekter ved evaluering av IT-systemer: ISO 25010

- *Functional suitability*: Gjør systemet det det skal?
- *Reliability*: Er det feil (bugs) i systemet?
- *Usability*: Er systemet greit å bruke?
- *Performance efficiency*: Hvor raskt og hvor mye plass trenger systemet?
- *Maintainability*: Er systemet greit å vedlikeholde og utvide?
- *Portability*: Kan det enkelt flyttes til ulike plattformer?
- *Compatibility*: Kan det greit fungere sammen med andre systemer?
- *Security*: Hvor sikkert er systemet?

Hvilke av disse produkt-egenskapene er relevante for dere?

Hva slags arbeidsprosesser (hvordan dere jobber) bør dere ha for å oppfylle de produktkravene dere ønsker?

Funksjonelle krav

Hva systemet skal gjøre

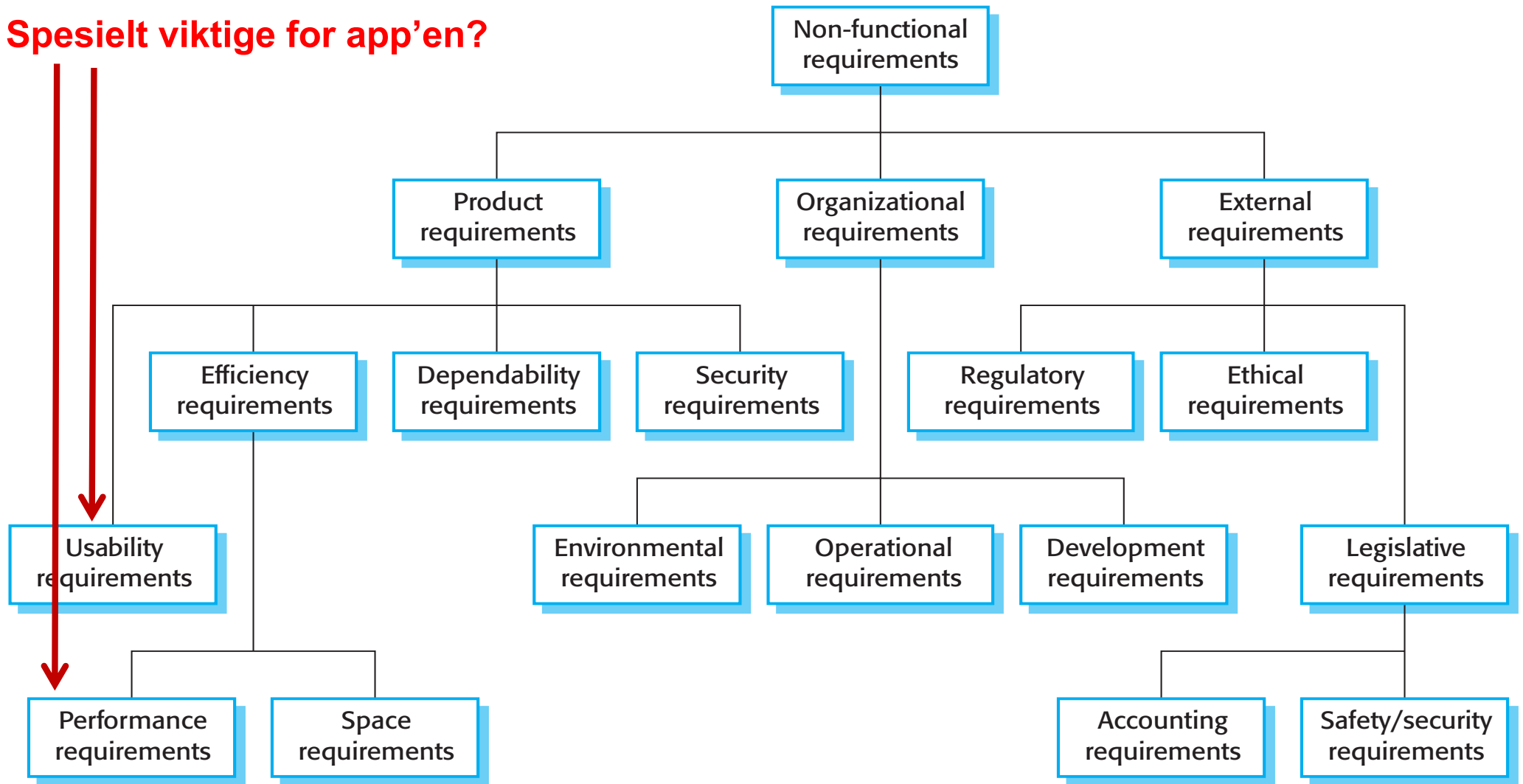
- Hvilke tjenester (funksjoner) skal systemet tilby?
- Hvordan skal det reagere på ulike typer input?

Ikke-funksjonelle krav

- **Hvordan** systemet skal implementere de funksjonelle kravene

Typer av ikke-funksjonelle krav

Spesielt viktige for app'en?



”Usability requirements”

– er systemet lett å lære og bruke?

- Brukskvalitet/Brukervennlighet
 - avhenger av krav til opplæring
 - varierer for ulike brukergrupper
 - universell utforming
- Kan måles:
 - Hvor lang tid tar det å lære systemet for nybegynnere?
 - Hvor mange “brukerfeil” oppstår med erfarne brukere?
 - Hvor ofte får brukerne meningsløse tilbakemeldinger?
 - Responstid: Tid fra brukeren trykker 'OK' til systemet svarer



Brukskvalitet (usability)

- Hvordan gjennomføre en studie av brukskvalitet:
<https://www.usability.gov/how-to-and-tools/methods/running-usability-tests.html>

Når kriteriene for evaluering er klare, hvordan gjennomføre evalueringen?

- Hvilke forsknings- eller undersøkelsesmetoder egner seg?

Plan

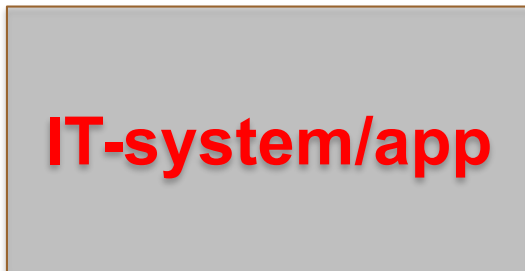
- IN2000 Case
- Behov for metodekunnskap
- Evaluering av IT-systemer/apper
- Typer forskningsmetoder 
 - Eksperiment 
 - Case-studie
 - Etnografi
 - Spørreskjema-undersøkelse (survey)

(Kontrollert) eksperiment

- Et eksperiment undersøker årsak-virkning – hva fører til hva?
- Manipulerer (endrer) det fenomenet man studerer, og så måler man virkningen
- Enkeltindivider eller team (deltakere) utfører like oppgaver der hensikten er å sammenligne ulike “treatments”

Eksperiment

Uavhengige variable
(treatment)



Effekt



Avhengige variable
(resultat)



Moderator
variable
(kontekst), f.eks. kompetanse



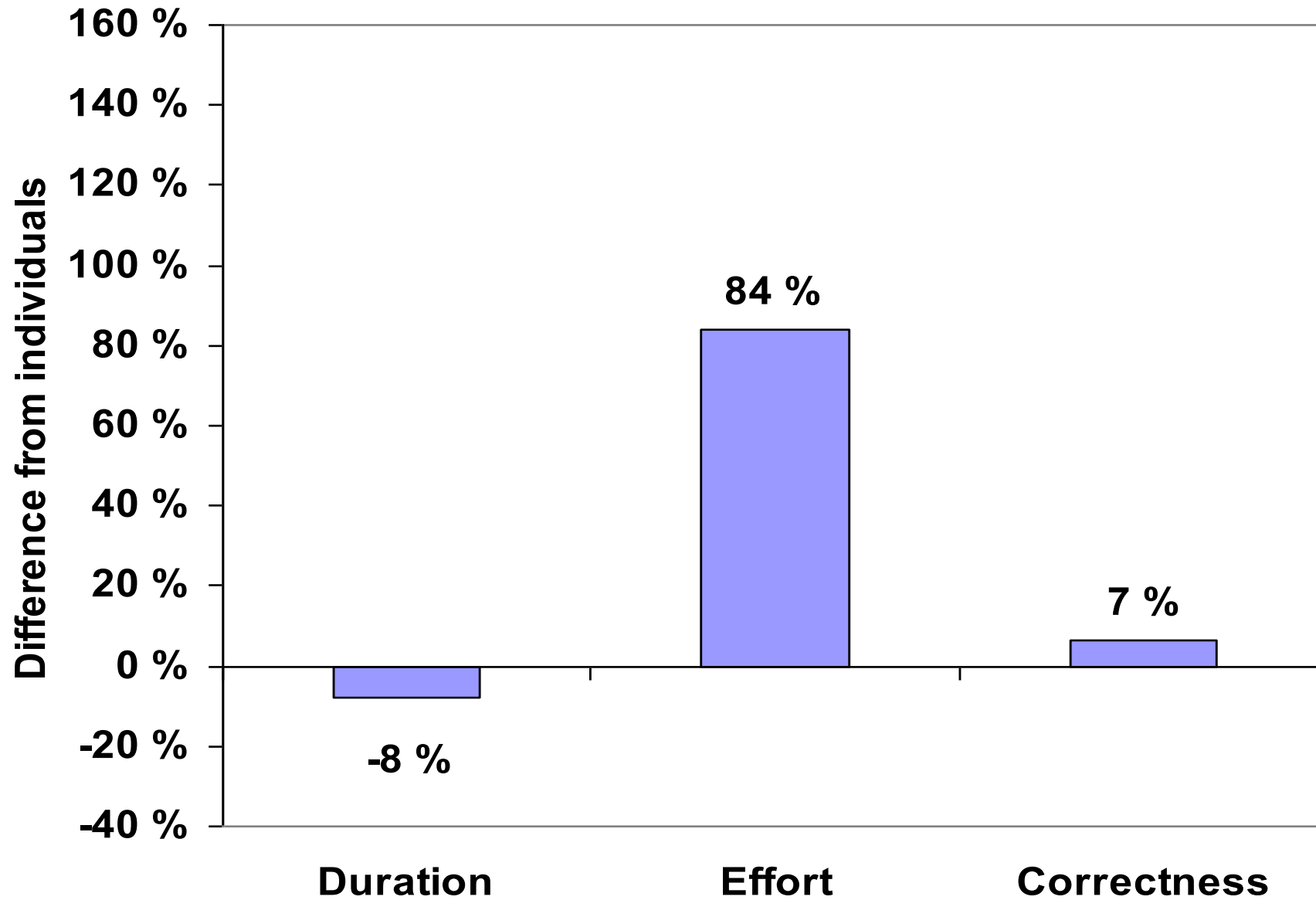
Eksperiment omtalt på IN1030:

Hva er kost-nyttten ved parprogramming?

Verdens største eksperiment i software engineering med 300 profesjonelle utviklere som deltakere

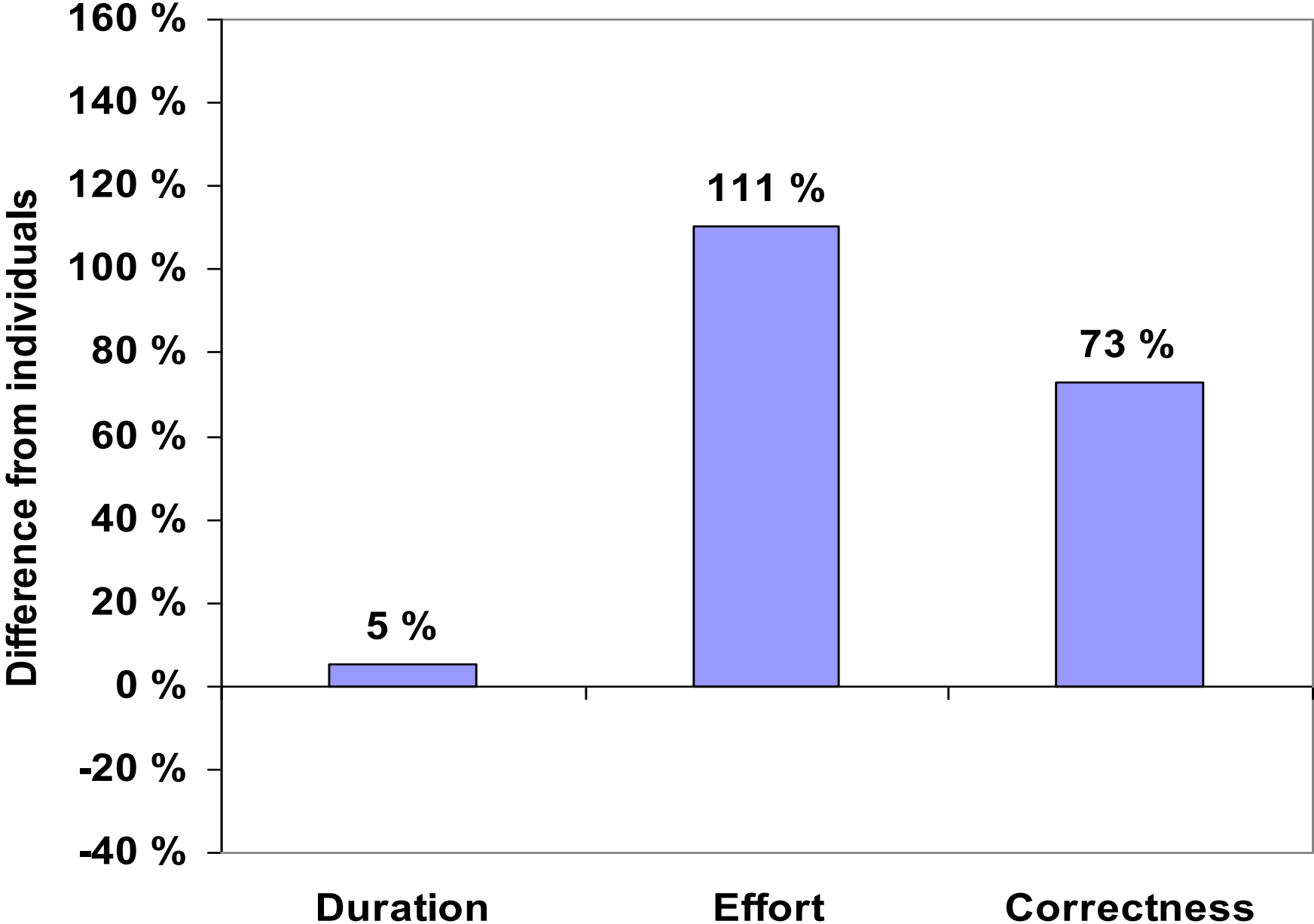
*E. Arisholm, H.E. Gallis, T. Dybå and D.I.K. Sjøberg. Evaluating Pair Programming with Respect to System Complexity and Programmer Expertise, *IEEE Transactions on Software Engineering* 33(2):65-86, 2007

Total Effect of PP

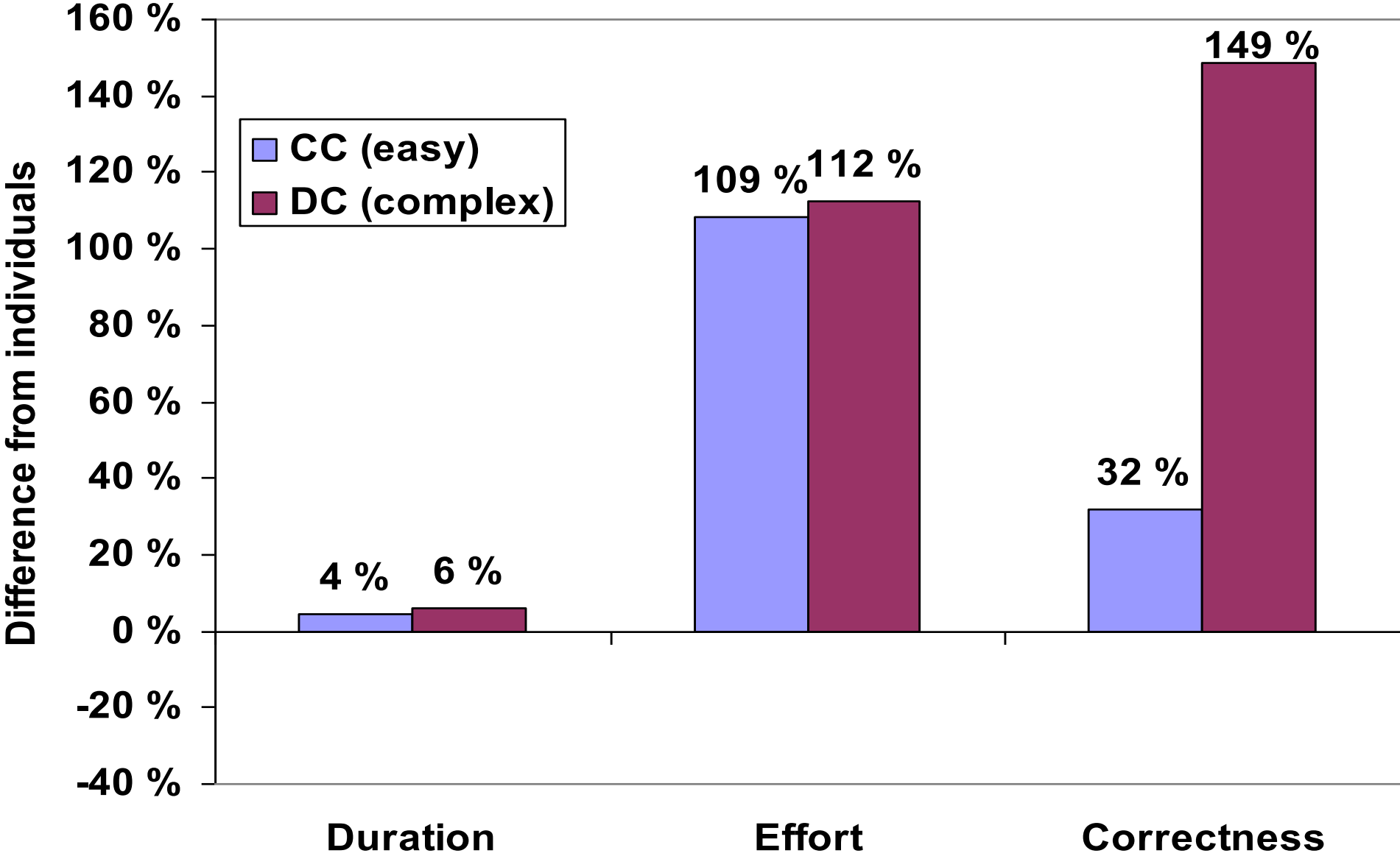


Correctness = Prosent flere riktige løsninger enn hos individuelle programmerere

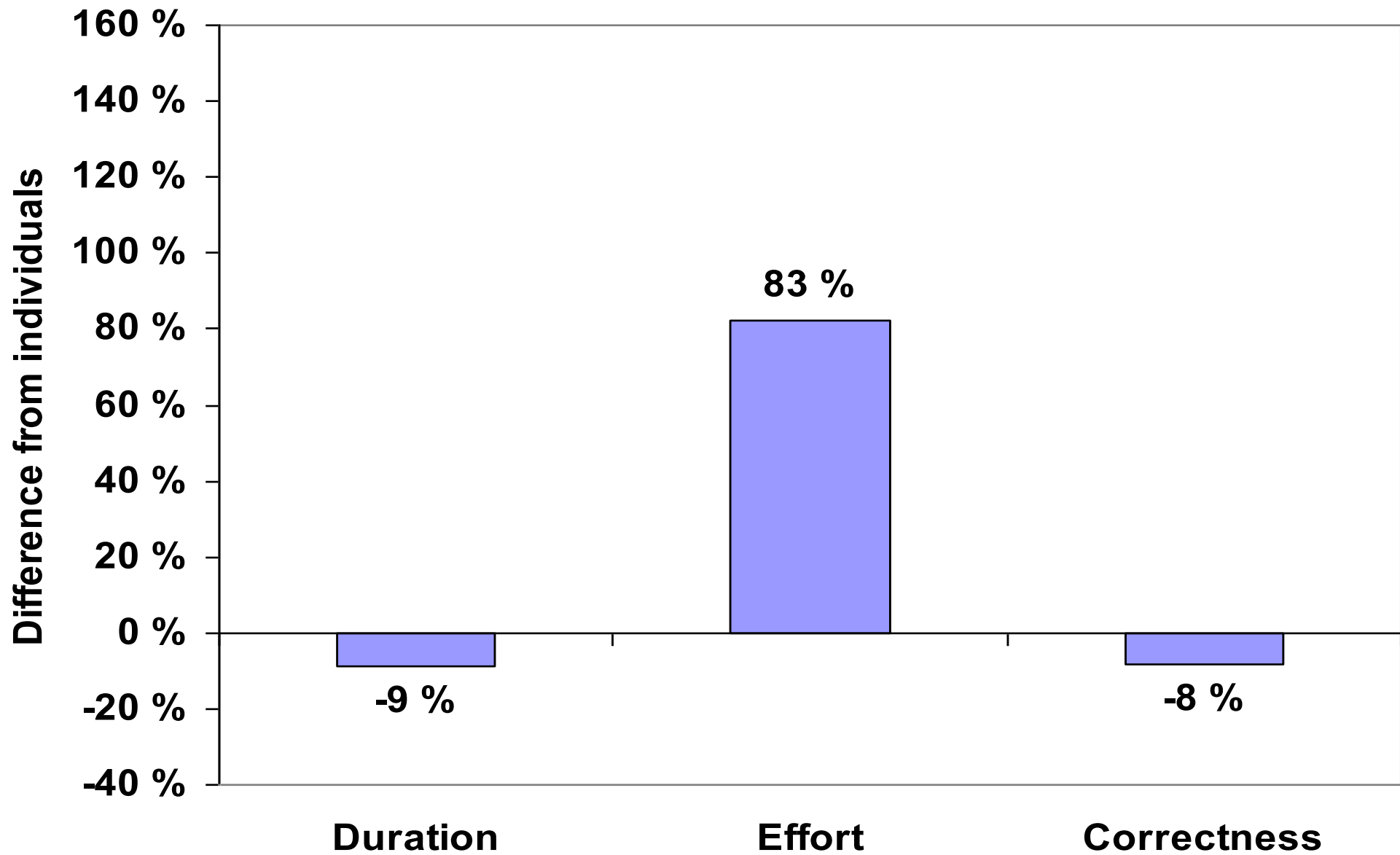
Effect of PP for Juniors



Moderating Effect of System Complexity for Juniors



Effect of PP for Seniors




Effekten av å bruke parprogrammering “kommer an på”

Programmerings- ekspertise	Oppgave- kompleksitet	Bruke PP?	Kommentar
Junior	Lett	Ja	Gitt at hovedmålet er god kvalitet
	Vanskelig	Ja	Gitt at hovedmålet er god kvalitet
Mellomnivå	Lett	Nei	
	Vanskelig	Ja	Gitt at hovedmålet er god kvalitet
Senior	Lett	Nei	
	Vanskelig	Nei *	

* Med mindre oppgaven er svært for vanskelig, selv for en senior

Plan

- IN2000 Case
- Behov for metodekunnskap
- Evaluering av IT-systemer/apper
- Typer forskningsmetoder
 - Eksperiment
 - Case-studie 
 - Etnografi
 - Spørreskjema-undersøkelse (survey)

Case-studier

- Eksperimenter svarer på “hva” effekten er, case-studier mer på “hvordan” og “hvorfor”
- Case-studier legger vekt på å studere fenomener i sine naturlige omgivelser
- Case-studier har få datapunkter og mange variable. Derfor generaliserer man ved bruk av teori, ikke statistikk som i eksperimenter

Intervjuer – ofte brukt i case-studier

- **Strukturerte intervjuer**

- Spørsmålene definert på forhånd, veldefinerte svaralternativer. Kan kvantifisere (telle opp) hvor mange som svarer hva på hvert spørsmål


- **Semistrukturerte intervjuer**

- Intervjuerne baserer seg på stikkord og spørsmål som evt. kan droppes underveis, og nye spørsmål kan stilles avhengig av hvordan intervjuet forløper

- **Åpne (ustrukturerte) intervjuer**

- Forløper seg mer som en samtale mellom intervjuer og intervjuobjekt

Plan

- IN2000 Case
 - Behov for metodekunnskap
 - Evaluering av IT-systemer/apper
 - Typer forskningsmetoder
 - Eksperiment
 - Case-studie
 - Etnografi
 - Spørreskjema-undersøkelse (survey)
- 

Etnografi/observasjon

- Dyp forståelse av folk, organisasjon og konteksten for arbeidet
- Forskeren observerer over lengre tid hva folk gjør og er mer involvert i gruppen som studeres enn i case-studier
- Personene som studeres trenger ikke å forklare hva de gjør

Plan

- IN2000 Case
- Behov for metodekunnskap
- Evaluering av IT-systemer/apper
- Typer forskningsmetoder
 - Eksperiment
 - Case-studie
 - Etnografi
 - Spørreskjema-undersøkelse (survey)



Spørreskjemaundersøkelser (Surveys)



Spørreskjemaer – type spørsmål

- Numeriske verdier, for eksempel alder
- Svarkategorier, for eksempel stillingstype
- Ja/Nei-svar
- Ordinalskala (rangering) som vanligvis er bedre for holdninger og preferanser. Tre typer
 - Enighetsskalaer, f.eks. 5-nivå Likert-skala med kategoriene: sterkt uenig, uenig, verken uenig eller enig, enig, sterkt enig
 - Frekvensskala, for eksempel aldri, sjelden, av og til, ofte, alltid
 - Evalueringsskalaer: svært dårlig, dårlig, passe, god, veldig god
- Åpne spørsmål

Oppsummering

- Evaluering av et produkt (og prosess) kan gjøres mer eller mindre vitenskapelig
 - Må være pragmatisk i IN2000-prosjektet
- God forskningsmetode gir pålitelig kunnskap
- Bruk av forskningsmetode krever kompetanse og tilstrekkelig med ressurser