

Know when to hold 'em, know when to fold 'em...

Combining estimates with planning poker

Kjetil Moløkken-Østvold

Simula Research Laboratory

kjetilmo@simula.no

Nils Christian Haugen

Objectnet

nch@objectnet.no





Agenda

- Combining estimates
- Industrial studies
 - Planning poker vs. unstructured combination (UK)
 - Planning poker vs. individual estimates (Norway)
- Discussion
- Q&A



Background

- Most professionals are subject to group processes when estimating a project
- Warning!
 - Much of the "traditional" software engineering literature misinterprets and simplifies psychological research on groups
 - Lack of empirical research
- Research has found that combination of expert estimates might reduce over-optimism
 - Which method is used to combine estimates?
 - How is the project climate (customer, priorities, management)?
 - Who are the experts?

Some methods for combining estimates

<i>Method</i>	<i>Structure</i>	<i>Anonymity</i>	<i>Interaction</i>	<i>Overhead</i>
<i>Delphi</i>	Heavy	Yes	No	Major
<i>Wideband Delphi</i>	Moderate	Limited	Limited	Moderate
<i>Planning Poker</i>	Light	No	Yes	Limited
<i>Unstructured group</i>	None	No	Yes	Limited



Central features of planning poker

- Group discussion helps define the tasks before estimating
- Combines knowledge from several sources
- Simultaneous revealing of estimates might reduce anchoring effect and social comparison concerns
- Minimal overhead
- Face-to-face interaction
- Developers take more ownership of estimates



Industrial studies

- Planning poker vs. unstructured combination (UK)
- Planning poker vs. individual estimation (Norway)



Planning poker vs. unstructured combination

- Goal: Compare estimation performance of unstructured combination and planning poker
- Estimates derived in release planning
 - New release every 2-3 months
 - Estimates used throughout project
- Data from 4 subsequent releases
 - Order: Unstructured, Unstructured, Planning poker, Planning Poker
 - Most likely estimates in pair days



Team and methodology

- Mix of employees, consultants and independent contractors
 - 8-12 developers
 - 15-20 people total
- eXtreme Programming
 - User stories
 - Pair programming
 - Storytest-driven development
 - Story cards on wall
 - Daily stand-ups



Estimation process

- Unstructured group
 - Business analyst presents story
 - Story is discussed
 - Estimate volunteered
 - Consensus sought
- Planning Poker
 - Business analyst presents story
 - Story is discussed
 - Individual estimates derived
 - Estimates revealed simultaneously
 - Lowest/highest estimate justified
 - Team decides on collective estimate



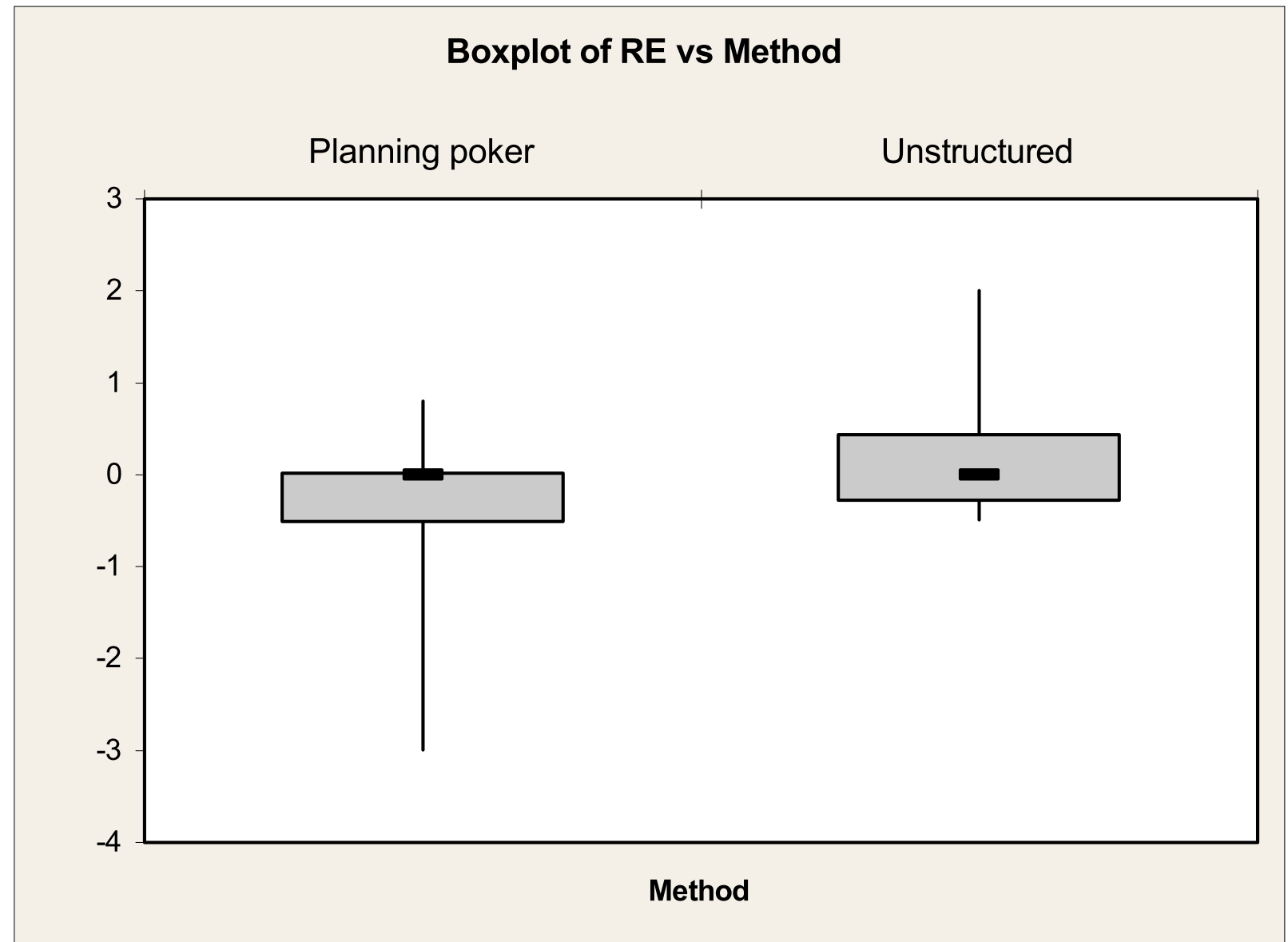
The effect of planning poker

- Better group discussions
 - Got everybody involved
 - Revealed more information about the tasks
- Team preferred planning poker to unstructured combination
 - Made estimation process more effective
 - Fun!
- Tendency for tasks to be overestimated with planning poker
 - Or, is it a tendency to finish under estimate?

Results PP vs Unstructured

Relative Error =
 $(\text{actual} - \text{estimate}) / \text{actual}$

Not balanced
 $(2-1) / 2 = 0.5$
 $(1-2) / 1 = -1.0$



	<i>PP</i>	<i>UC</i>	<i>Comment</i>
<i>Median</i>	0.00	0.00	Typical case on target for both groups
<i>Mean</i>	-0.26	-0.08	Some tasks where overestimated with PP



Planning poker vs. individual estimates

- Twofold purpose of study
 - Compare PP tasks with a set of control tasks, estimated by individuals
 - Investigate if there is a reduction of optimism after discussion in the PP tasks
- Planning poker estimates performed in sprint planning
 - 14-day sprints
 - Individual estimates performed on task creation
- Data from 4 sprints
 - 50% of tasks re-estimated using planning poker
 - Most likely estimates in hours



Team and methodology

- Consultants from same company
 - 4-6 people total
 - All developers
- Scrum
 - Solo programming
 - Tasks tested and QAed by another person on the team
 - Tasks kept in issue tracking system (Jira)
 - Daily scrum



Estimation process

- For the planning poker tasks:
 - Task presented by task creator
 - Task is discussed
 - Individual estimates derived
 - Estimates revealed simultaneously (not predefined units)
 - High/low estimator justifies
 - Consensus sought
- Control task were estimated by an individual expert



The tasks estimated with planning poker

- Compared with the mechanical combination of individual estimates, the consensus estimates were
 - less optimistic after discussion
 - more accurate after discussion
- This is opposite of what is found in most studies from areas like psychology (usually optimism is found to increase after discussions)
- Can be caused by different perspectives on a task and/or identification of sub-problems

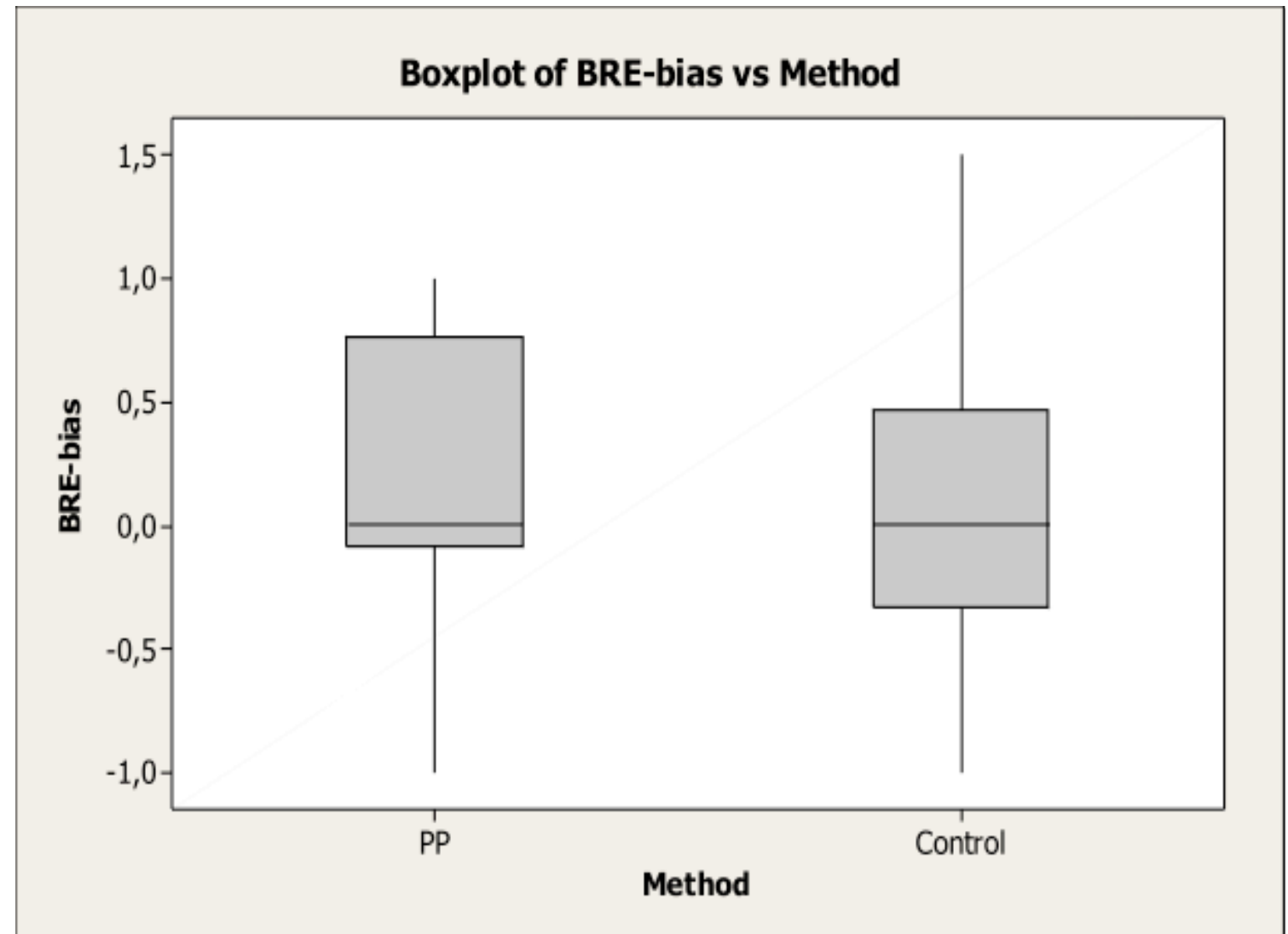
Results PP vs. Individual

BRE-bias =
 $(\text{actual} - \text{estimate}) / \min(\text{actual}, \text{estimate})$

Balanced

$$(2-1) / 1 = 1.0$$

$$(1-2) / 1 = -1.0$$



	<i>PP</i>	<i>Ind.</i>	<i>Comment</i>
<i>Median</i>	0,00	0,00	Typical case on target for both groups
<i>Mean</i>	0,33	-0,04	Some PP tasks were underestimated

Results from analysis of code

- Planning poker tasks had on average:
 - Twice as many deleted control statements (indicates that effort was spent to reduce complexity)
 - Twice as many out-of-class references deleted (indicates that effort was spent to reduce coupling)
- Extra time spent on restructuring and simplifying code?
 - Can explain why there were some overruns in the planning poker tasks
- Question: Did planning poker (group discussion) lead to increased focus on quality?



Possible benefits of group estimating

- Participants takes more ownership of estimates
 - Estimates are not forced
 - More motivation to work towards estimates
 - Easier to estimate ones own work
- Uncertainty related to the implementation can be discussed and handled at an early stage (depending on combination method)
- Reduced need for discussion during project execution



Hazards of groups

- Lack of decentralisation and independence may make the group decision vulnerable to peer-pressure (depending on technique)
- The "anchor-effect" can have an impact
- An unstructured discussion might have side-effects, such as increasing number of sub-tasks

Summary

- Combination of estimates may increase accuracy, but can have certain side-effects
 - Does planning poker lead to more focus on quality?
 - Does planning poker lead to more focus on time?
- Planning poker was popular among developers
 - Both the UK and Norway
 - Rated as fun and easy to implement
- Choice of combination method should depend on project characteristics
 - Political pressure
 - Project practises



Questions?

- Diverging results
 - Differences in study setup
 - Differences in project properties
 - Cultural differences?
- How will maturation in Planning Poker proficiency affect results?
- Does combining expert opinions mainly impact the estimation process, or are other effects more important?



References

- **Group Processes**
Rupert Brown, 2001
- **The Wisdom of Crowds**
James Surowiecki, 2004
- **Agile Estimating and Planning**
Mike Cohn, 2006
- **An Empirical Study of Using Planning Poker for User Story Estimation**
N.C. Haugen, Agile Conference, 2006
- **Group Processes in Software Effort Estimation**
K.J. Moløyken-Østfold and M. Jørgensen, Empirical Software Engineering 9(4):315-334, 2004
- **Combining Estimates with Planning Poker - An Empirical Study**
K.J. Moløyken-Østfold and N.C. Haugen, submitted to 18th Australian Conference on Software Engineering, 2007