*Are two heads better than one?*

# On the Effectiveness of Pair Programming

Tore Dybå
Erik Arisholm
Dag I.K. Sjøberg
Jo E. Hannay
Forrest Shull

# Beware: Very Short Paper!

5 Authors

3 Text pages

2 Figures

# About the Authors [1/5]

**Linked** in ®

**Erik Arisholm**

Daglig leder, Testify AS

Oslo Area, Norway | Information Technology and Services

| | |
|---|---|
| Current | • **Daglig leder** at **Testify AS**<br>• **Professor** at **Dept. of Informatics, Univ. of Oslo** |
| Past | • Research Scientist in Software Engineering at Simula Research Laboratory<br>• Associate Professor at Informatics, University of Oslo<br>• PhD-student at Informatics, University of Oslo<br><br>see all... |
| Education | • Univ. of Oslo<br>• University of Toronto<br>• Univ. of Oslo |
| Connections | **69** connections |
| Websites | • Company Website |

# About the Authors [2/5]

**Linked** in.



Jo Hannay
Research Scientist Simula Research Laboratory
Oslo Area, Norway | Research

| | |
|---|---|
| Current | • **Research Scientist** at **Simula Research Laboratory** |
| Past | • Adjunct Associate Professor at Universitetet i Oslo<br>• Software Developer at KLP |
| Education | • Diakonhjemmet Høgskole<br>• The University of Edinburgh<br>• Universitetet i Oslo |
| Connections | **88** connections |
| Websites | • Personal Website<br>• Company Website<br>• Company Website |
| Public Profile | http://no.linkedin.com/pub/jo-hannay/1/980/943 |

# About the Authors [3/5]

**Linked** in ®

## Dag Sjøberg

Professor at University of Oslo

Oslo Area, Norway | Research

| | |
|---|---|
| Current | • **Professor** at **University of Oslo** |
| Past | • Research Director / Head of Department at Simula Research Laboratory<br>• Associate Professor (1. amanuensis) at Universitetet i Oslo<br>• Post.Doc. at Universitet i Oslo<br>see all... |
| Education | • University of Glasgow<br>• Universitetet i Oslo<br>• Universitetet i Oslo<br>see all... |
| Connections | **89** connections |

# About the Authors [4/5]

**Linked** in.



**Tore Dybå**
Chief Scientist at SINTEF
Trondheim Area, Norway | Research

| | |
|---|---|
| Current | • **Adjunct Professor** at **University of Oslo**<br>• **Chief Scientist** at **SINTEF** |
| Past | • Visiting Scientist at Simula Research Laboratory<br>• Senior Consultant at Fjellanger Widerøe<br>• Consultant at Teleplan<br><div align="right">see all...</div> |
| Education | • Norges teknisk-naturvitenskapelige universitet<br>• Norges teknisk-naturvitenskapelige universitet<br>• Norwegian Military College (BSIV) |
| Connections | **222** connections |
| Websites | • Selected Publications (DBLP) |

# About the Authors [5/5]

**Linked** in.



**Forrest Shull**
Editor in Chief at IEEE Software
Washington D.C. Metro Area | Research

| | |
|---|---|
| Current | • **Editor in Chief** at **IEEE Software**<br>• **Senior Scientist and Division Director** at **Fraunhofer Center - Maryland** |
| Past | • Associate Editor in Chief at IEEE Software<br>• Member at International Software Engineering Research Network (ISERN)<br>• Faculty Research Associate at University of Maryland at College Park |
| Education | • University of Maryland College Park<br>• University of Maryland College Park<br>• Loyola College in Maryland |
| Connections | **121** connections |
| Public Profile | http://www.linkedin.com/pub/forrest-shull/0/191/4a1 |

# #define PairProgramming

- 2 Programmers
- 1 Keyboard
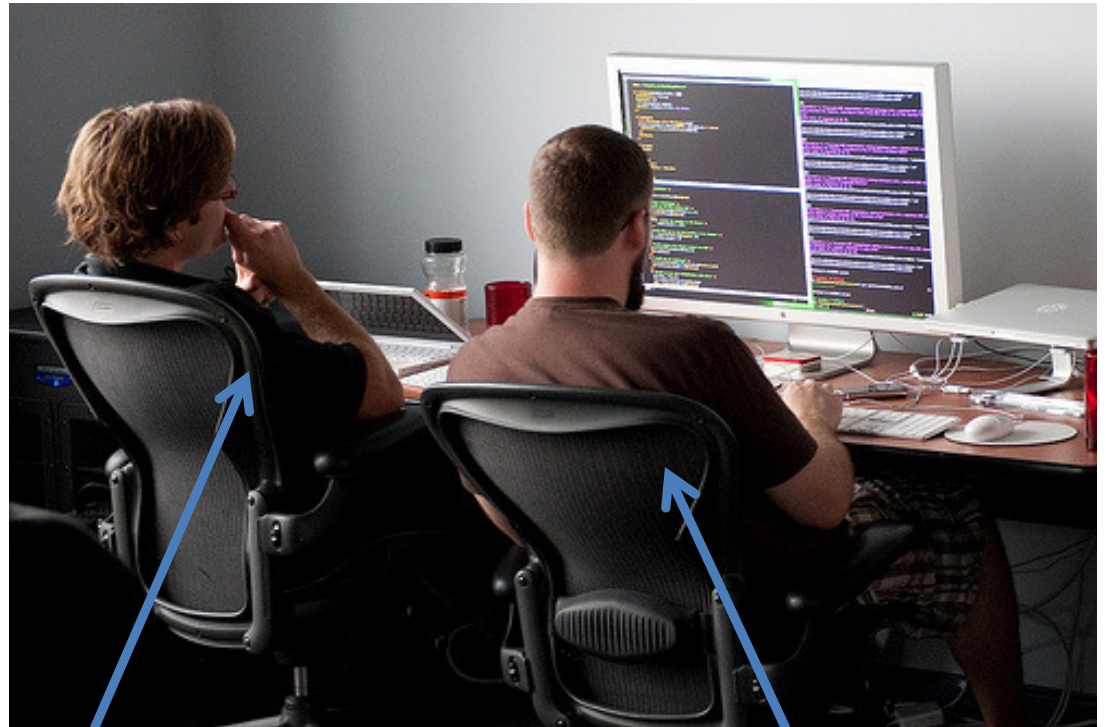


Observer/Navigator

Driver

# Overview

- Paper's Goal:

  Empirically decide whether PP is "effective"

- Research Method:

  Combine results from 15 different studies

- Conclusion: IT DEPENDS!

# 15 Studies → 15 Results Formats

- Studies focused on different issues:
  - Number of subject varies: 12 to 295
  - US vs Europe
  - Teams vs. Individuals
  - Students vs. Pros.
  - **Different grading method, scales**
- Solution:
  - Normalize: Subtract mean & Divide by std dev.

# Measuring "Effectiveness"

- Duration == Calendar time to complete task

- Effort == Person-hours consumed
  - 1 Person  x 1 Hour  == 1 Person-Hour
  - 2 Persons x 1 Hour  == 2 Person-Hours
  - 2 Persons x 2 Hours == 4 Person-Hours

- Quality == How good the final product is

# Not Measuring

- Programmer's Happiness
- Teamwork Improvement
- Knowledge Transfer
- Learning

# Duration-Effort-Quality Relation



1 Year



~1/2 Year

Same Effort,
Shorter Duration

1 Year

Same Duration,
Twice the Effort,
Better quality

# Duration-Effort-Quality Relation
## **FAIL**
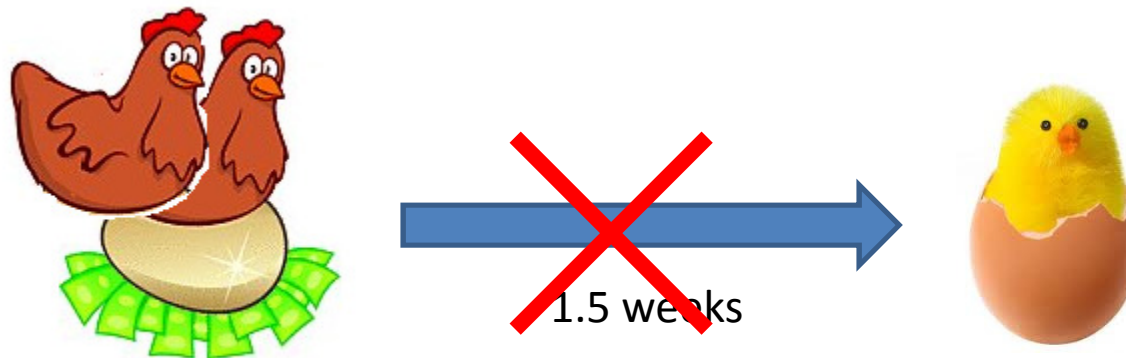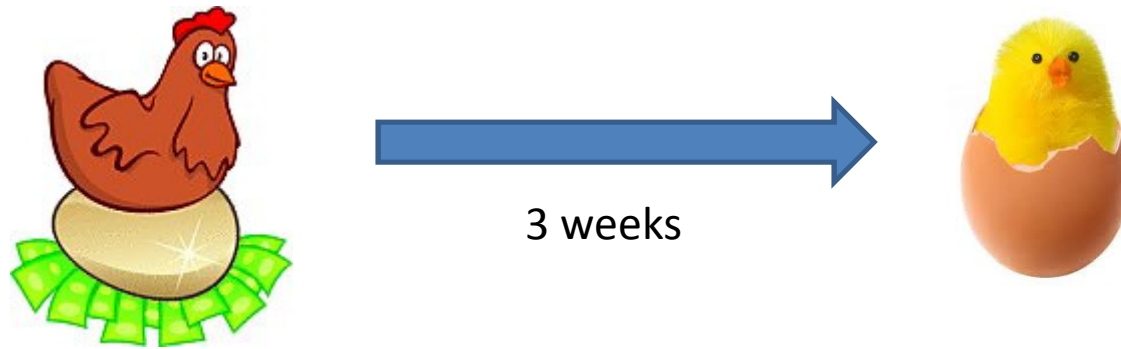
# Which is it for Pair Programming?



OR

# The Result (DON'T PANIC)



| Study | Effect size | Lower limit | Upper limit | | Relative weight |
|---|---|---|---|---|---|
| P07a | 0.11 | −0.24 | 0.46 | | 23.24 |
| S06a | 0.08 | −0.28 | 0.43 | | 22.85 |
| S00 | 1.04 | 0.65 | 1.43 | | 18.64 |
| S03 | 0.10 | −0.44 | 0.64 | | 9.80 |
| S05b | 0.28 | −0.32 | 0.88 | | 7.95 |
| P07b | 0.69 | −0.09 | 1.46 | | 4.73 |
| S02 | 0.30 | −0.50 | 1.09 | | 4.53 |
| S06c | 0.32 | −0.69 | 1.32 | | 2.81 |
| S06b | 0.51 | −0.59 | 1.62 | | 2.33 |
| P98 | 0.91 | −0.28 | 2.10 | | 2.02 |
| S06d | 2.20 | 0.58 | 3.82 | | 1.09 |
| **(a) Overall effect** | **0.38** | **0.21** | **0.55** | | |
| | | | | | |
| P07a | 0.21 | −0.13 | 0.54 | | 33.33 |
| S05a | 0.57 | 0.07 | 1.07 | | 15.24 |
| S01 | 0.16 | −0.40 | 0.73 | | 11.83 |
| S03 | 0.55 | −0.03 | 1.14 | | 10.97 |
| S05b | 1.30 | 0.64 | 1.95 | | 8.66 |
| P07b | −0.59 | −1.37 | 0.20 | | 6.09 |
| S02 | −0.05 | −0.84 | 0.74 | | 6.01 |
| P02 | 0.06 | −1.08 | 1.21 | | 2.87 |
| S06b | 0.98 | −0.24 | 2.20 | | 2.51 |
| S06d | 1.85 | 0.34 | 3.35 | | 1.65 |
| P98 | 4.09 | 1.98 | 6.20 | | 0.84 |
| **(b) Overall effect** | **0.40** | **0.21** | **0.59** | | |
| | | | | | |
| P07a | −0.68 | −1.03 | −0.34 | | 47.66 |
| S05a | −1.09 | −1.61 | −0.58 | | 21.17 |
| S05b | −0.25 | −0.85 | 0.35 | | 15.73 |
| S05c | −0.49 | −1.62 | 0.63 | | 4.46 |
| S06b | −0.64 | −1.83 | 0.54 | | 4.04 |
| S06d | −0.11 | −1.31 | 1.10 | | 3.88 |
| S06c | 2.52 | 1.17 | 3.88 | | 3.07 |
| **(c) Overall effect** | **−0.57** | **−0.81** | **−0.33** | | |

Favors solo programming | Favors pair programming

Figure 1. Meta-analyses of pair programming's effects on (a) quality, (b) duration, and (c) effort.

# The Result (DON'T PANIC)



Quality

| Study | Effect size | Lower limit | Upper limit | Effect size and 95% confidence interval | Relative weight |
|---|---|---|---|---|---|
| P07a | 0.11 | −0.24 | 0.46 | | 23.24 |
| S06a | 0.08 | −0.28 | 0.43 | | 22.85 |
| S00 | 1.04 | 0.65 | 1.43 | | 18.64 |
| S03 | 0.10 | −0.44 | 0.64 | | 9.80 |
| S05b | 0.28 | −0.32 | 0.88 | | 7.95 |
| P07b | 0.69 | −0.09 | 1.46 | | 4.73 |
| S02 | 0.30 | −0.50 | 1.09 | | 4.53 |
| S06c | 0.32 | −0.69 | 1.32 | | 2.81 |
| S06b | 0.51 | −0.59 | 1.62 | | 2.33 |
| P98 | 0.91 | −0.28 | 2.10 | | 2.02 |
| S06d | 2.20 | 0.58 | 3.82 | | 1.09 |
| (a) Overall effect | 0.38 | 0.21 | 0.55 | | |
| P07a | 0.21 | −0.13 | 0.54 | | 33.33 |
| S05a | 0.57 | 0.07 | 1.07 | | 15.24 |
| S01 | 0.16 | −0.40 | 0.73 | | 11.83 |
| S03 | 0.55 | −0.03 | 1.14 | | 10.97 |
| S05b | 1.30 | 0.64 | 1.95 | | 8.66 |
| P07b | −0.59 | −1.37 | 0.20 | | 6.09 |
| S02 | −0.05 | −0.84 | 0.74 | | 6.01 |
| P02 | 0.06 | −1.08 | 1.21 | | 2.87 |
| S06b | 0.98 | −0.24 | 2.20 | | 2.51 |
| S06d | 1.85 | 0.34 | 3.35 | | 1.65 |
| P98 | 4.09 | 1.98 | 6.20 | | 0.84 |
| (b) Overall effect | 0.40 | 0.21 | 0.59 | | |
| P07a | −0.68 | −1.03 | −0.34 | | 47.66 |
| S05a | −1.09 | −1.61 | −0.58 | | 21.17 |
| S05b | −0.25 | −0.85 | 0.35 | | 15.73 |
| S05c | −0.49 | −1.62 | 0.63 | | 4.46 |
| S06b | −0.64 | −1.83 | 0.54 | | 4.04 |
| S06d | −0.11 | −1.31 | 1.10 | | 3.88 |
| S06c | 2.52 | 1.17 | 3.88 | | 3.07 |
| (c) Overall effect | −0.57 | −0.81 | −0.33 | | |

−2.00   −1.00   0.00   1.00   2.00

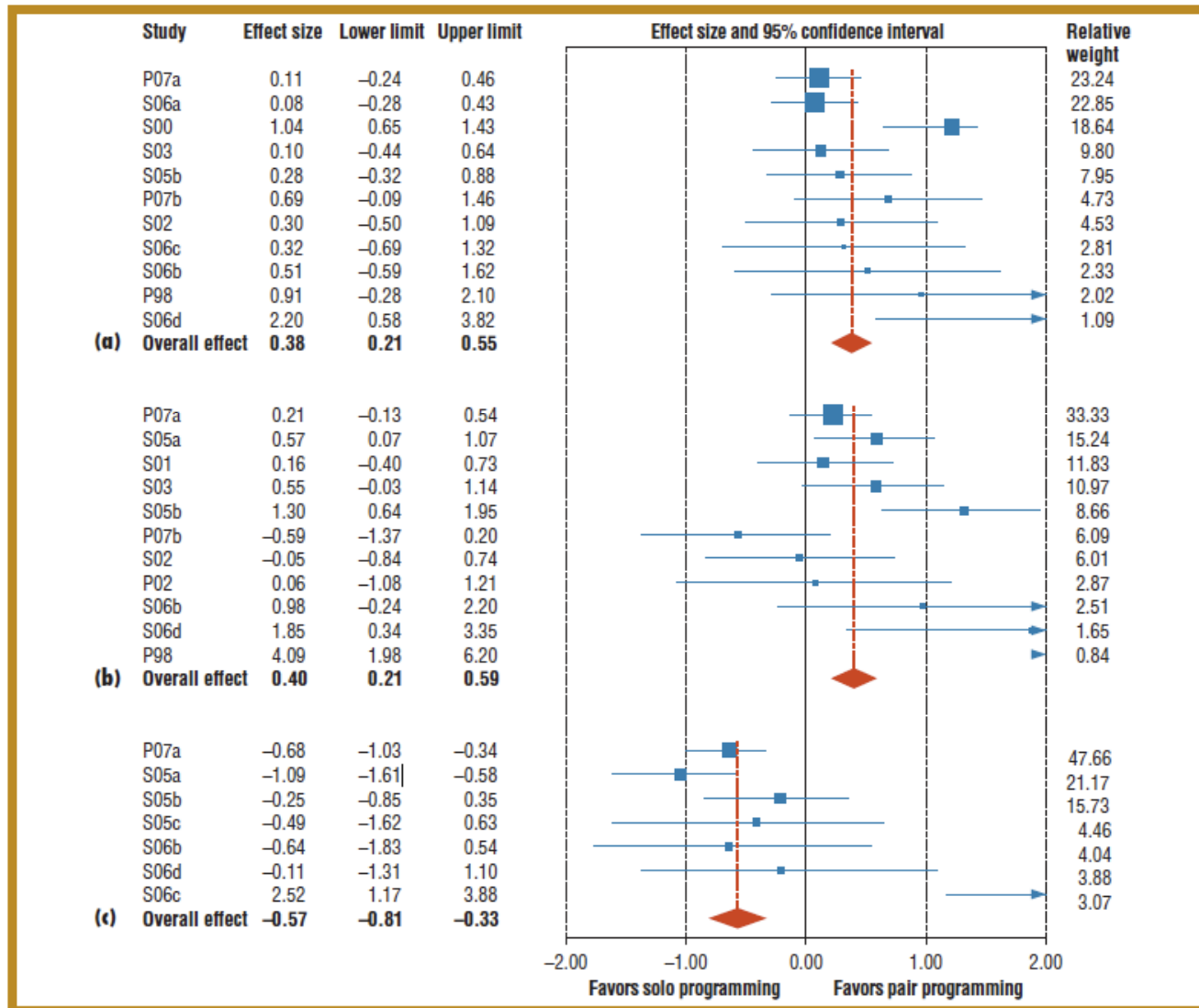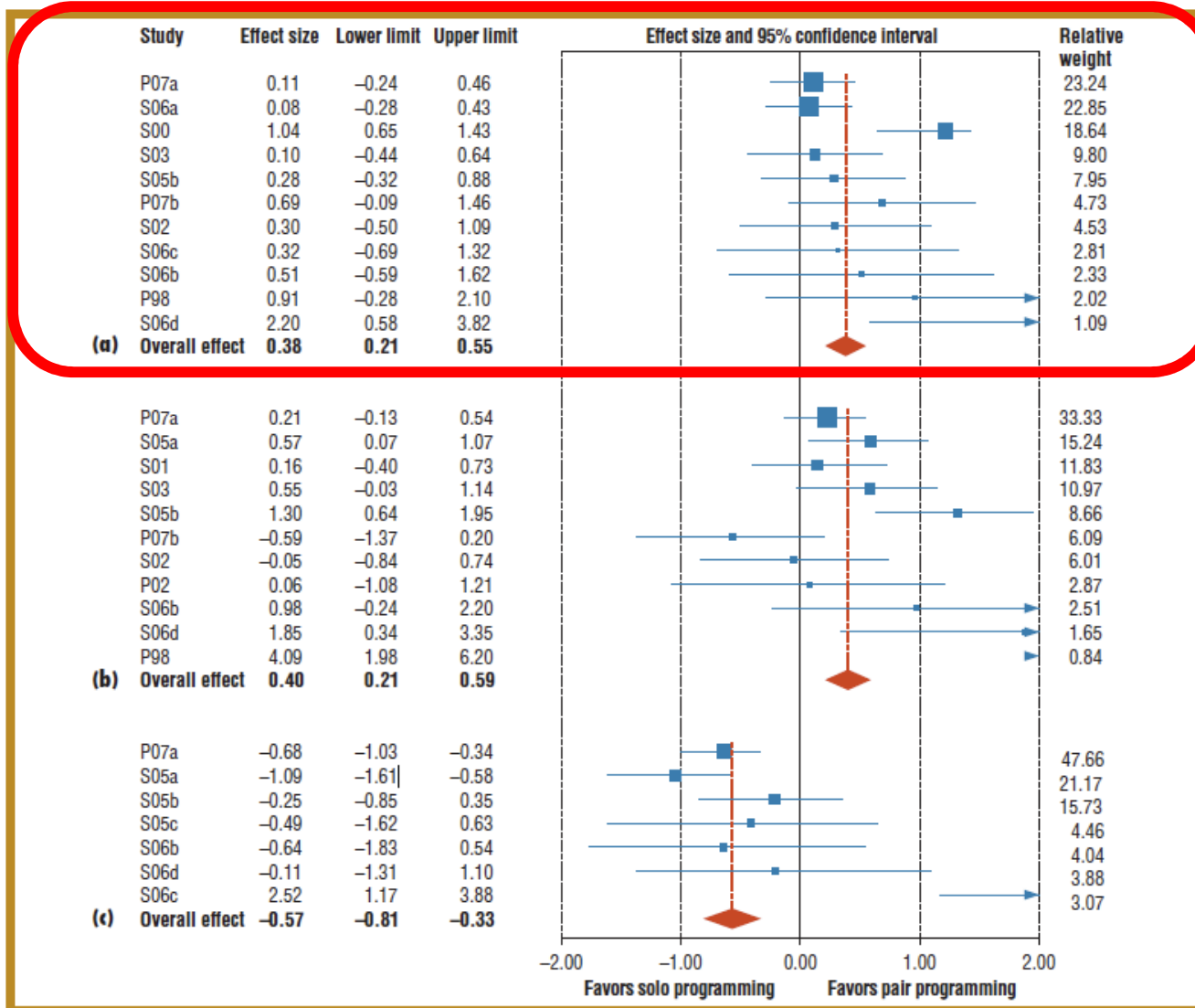Favors solo programming | Favors pair programming

Figure 1. Meta-analyses of pair programming's effects on (a) quality, (b) duration, and (c) effort.

# The Result (DON'T PANIC)

Quality



Figure 1. Meta-analyses of pair programming's effects on (a) quality, (b) duration, and (c) effort.
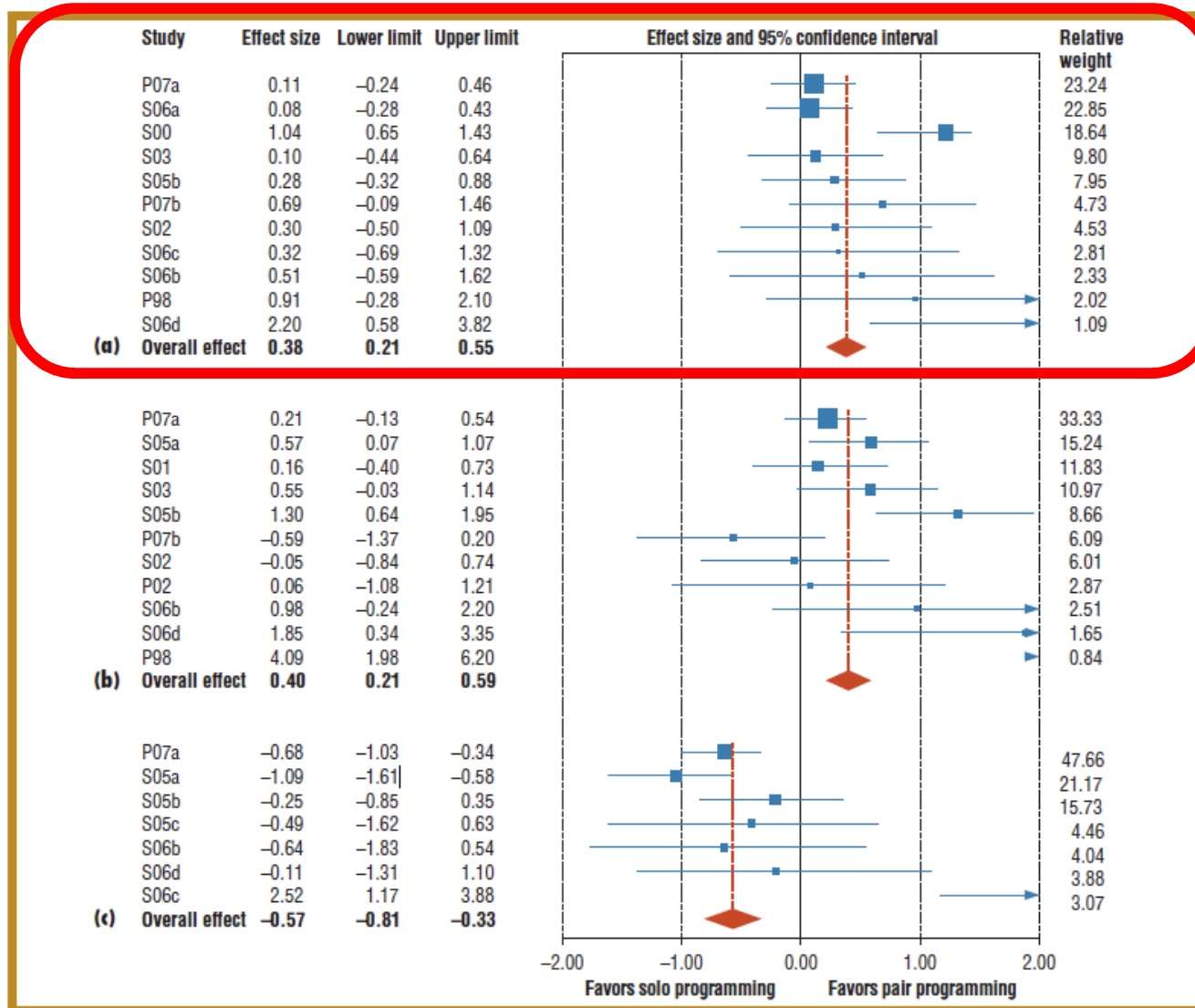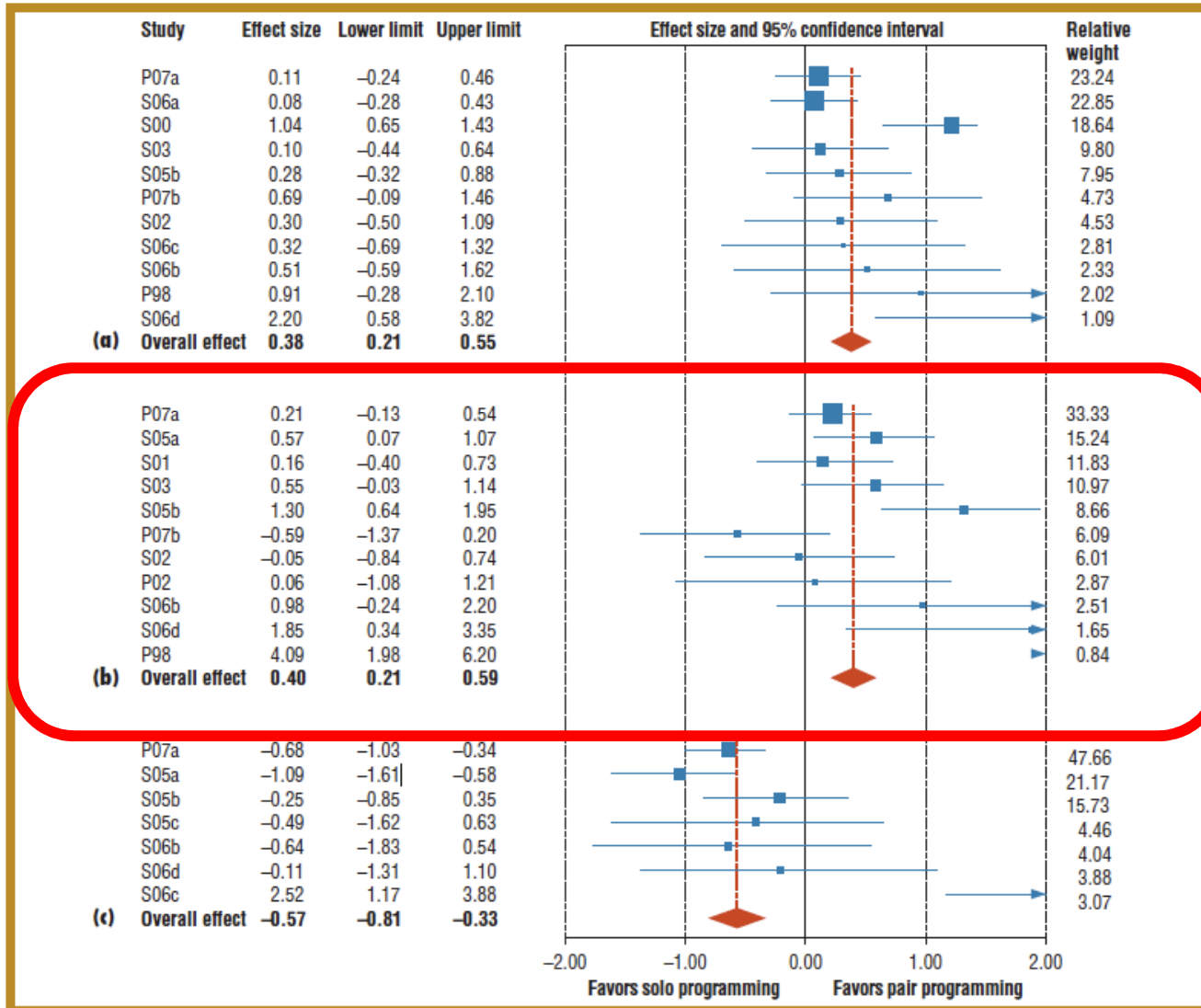
# The Result (DON'T PANIC)



Figure 1. Meta-analyses of pair programming's effects on (a) quality, (b) duration, and (c) effort.

# The Result (DON'T PANIC)



Effort

| Study | Effect size | Lower limit | Upper limit | | Relative weight |
|-------|-------------|-------------|-------------|---|-----------------|
| P07a | 0.11 | −0.24 | 0.46 | | 23.24 |
| S06a | 0.08 | −0.28 | 0.43 | | 22.85 |
| S00 | 1.04 | 0.65 | 1.43 | | 18.64 |
| S03 | 0.10 | −0.44 | 0.64 | | 9.80 |
| S05b | 0.28 | −0.32 | 0.88 | | 7.95 |
| P07b | 0.69 | −0.09 | 1.46 | | 4.73 |
| S02 | 0.30 | −0.50 | 1.09 | | 4.53 |
| S06c | 0.32 | −0.69 | 1.32 | | 2.81 |
| S06b | 0.51 | −0.59 | 1.62 | | 2.33 |
| P98 | 0.91 | −0.28 | 2.10 | | 2.02 |
| S06d | 2.20 | 0.58 | 3.82 | | 1.09 |
| **(a) Overall effect** | **0.38** | **0.21** | **0.55** | | |
| P07a | 0.21 | −0.13 | 0.54 | | 33.33 |
| S05a | 0.57 | 0.07 | 1.07 | | 15.24 |
| S01 | 0.16 | −0.40 | 0.73 | | 11.83 |
| S03 | 0.55 | −0.03 | 1.14 | | 10.97 |
| S05b | 1.30 | 0.64 | 1.95 | | 8.66 |
| P07b | −0.59 | −1.37 | 0.20 | | 6.09 |
| S02 | −0.05 | −0.84 | 0.74 | | 6.01 |
| P02 | 0.06 | −1.08 | 1.21 | | 2.87 |
| S06b | 0.98 | −0.24 | 2.20 | | 2.51 |
| S06d | 1.85 | 0.34 | 3.35 | | 1.65 |
| P98 | 4.09 | 1.98 | 6.20 | | 0.84 |
| **(b) Overall effect** | **0.40** | **0.21** | **0.59** | | |
| P07a | −0.68 | −1.03 | −0.34 | | 47.66 |
| S05a | −1.09 | −1.61 | −0.58 | | 21.17 |
| S05b | −0.25 | −0.85 | 0.35 | | 15.73 |
| S05c | −0.49 | −1.62 | 0.63 | | 4.46 |
| S06b | −0.64 | −1.83 | 0.54 | | 4.04 |
| S06d | −0.11 | −1.31 | 1.10 | | 3.88 |
| S06c | 2.52 | 1.17 | 3.88 | | 3.07 |
| **(c) Overall effect** | **−0.57** | **−0.81** | **−0.33** | | |

Favors solo programming          Favors pair programming

**Figure I. Meta-analyses of pair programming's effects on (a) quality, (b) duration, and (c) effort.**

# The Result (English)

- PP moderately **contributes to QUALITY**
- PP moderately **reduces the DURATION**
  - Note, there are exceptions!
- PP moderately **increases EFFORT** ($$$)

- Exploring beyond the avg case reveals more..

# The Result (DON'T PANIC)



**Quality**

**Duration**

**Effort**

| Study | Effect size | Lower limit | Upper limit | Effect size and 95% confidence interval | Relative weight |
|-------|-------------|-------------|-------------|------------------------------------------|-----------------|
| P07a | 0.11 | −0.24 | 0.46 | | 23.24 |
| S06a | 0.08 | −0.28 | 0.43 | | 22.85 |
| S00 | 1.04 | 0.65 | 1.43 | | 18.64 |
| S03 | 0.10 | −0.44 | 0.64 | | 9.80 |
| S05b | 0.28 | −0.32 | 0.88 | | 7.95 |
| P07b | 0.69 | −0.09 | 1.46 | | 4.73 |
| S02 | 0.30 | −0.50 | 1.09 | | 4.53 |
| S06c | 0.32 | −0.69 | 1.32 | | 2.81 |
| S06b | 0.51 | −0.59 | 1.62 | | 2.33 |
| P98 | 0.91 | −0.28 | 2.10 | | 2.02 |
| S06d | 2.20 | 0.58 | 3.82 | | 1.09 |
| **(a) Overall effect** | **0.38** | **0.21** | **0.55** | | |
| P07a | 0.21 | −0.13 | 0.54 | | 33.33 |
| S05a | 0.57 | 0.07 | 1.07 | | 15.24 |
| S01 | 0.16 | −0.40 | 0.73 | | 11.83 |
| S03 | 0.55 | −0.03 | 1.14 | | 10.97 |
| S05b | 1.30 | 0.64 | 1.95 | | 8.66 |
| P07b | −0.59 | −1.37 | 0.20 | | 6.09 |
| S02 | −0.05 | −0.84 | 0.74 | | 6.01 |
| P02 | 0.06 | −1.08 | 1.21 | | 2.87 |
| S06b | 0.98 | −0.24 | 2.20 | | 2.51 |
| S06d | 1.85 | 0.34 | 3.35 | | 1.65 |
| P98 | 4.09 | 1.98 | 6.20 | | 0.84 |
| **(b) Overall effect** | **0.40** | **0.21** | **0.59** | | |
| P07a | −0.68 | −1.03 | −0.34 | | 47.66 |
| S05a | −1.09 | −1.61 | −0.58 | | 21.17 |
| S05b | −0.25 | −0.85 | 0.35 | | 15.73 |
| S05c | −0.49 | −1.62 | 0.63 | | 4.46 |
| S06b | −0.64 | −1.83 | 0.54 | | 4.04 |
| S06d | −0.11 | −1.31 | 1.10 | | 3.88 |
| S06c | 2.52 | 1.17 | 3.88 | | 3.07 |
| **(c) Overall effect** | **−0.57** | **−0.81** | **−0.33** | | |

−2.00   −1.00   0.00   1.00   2.00
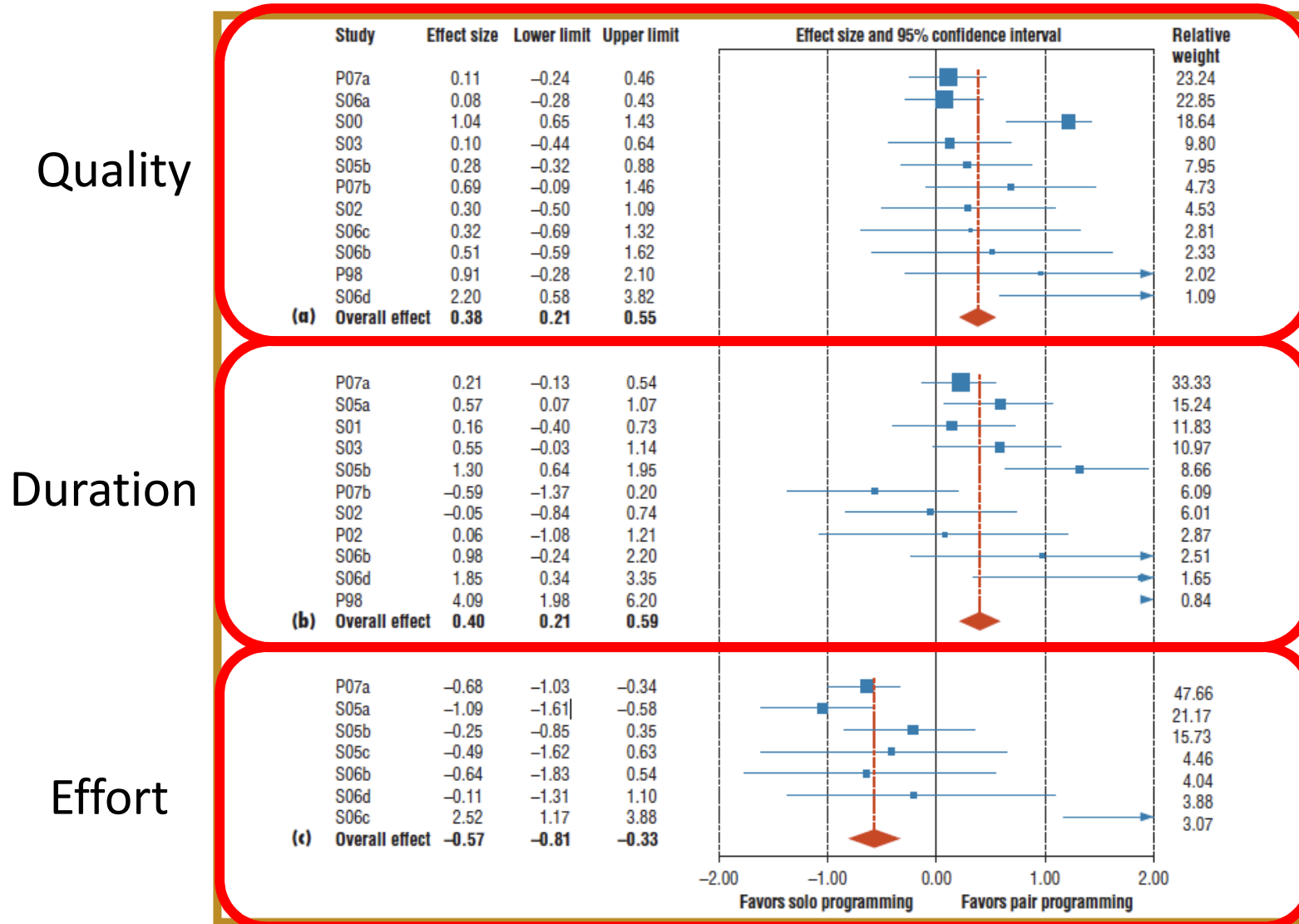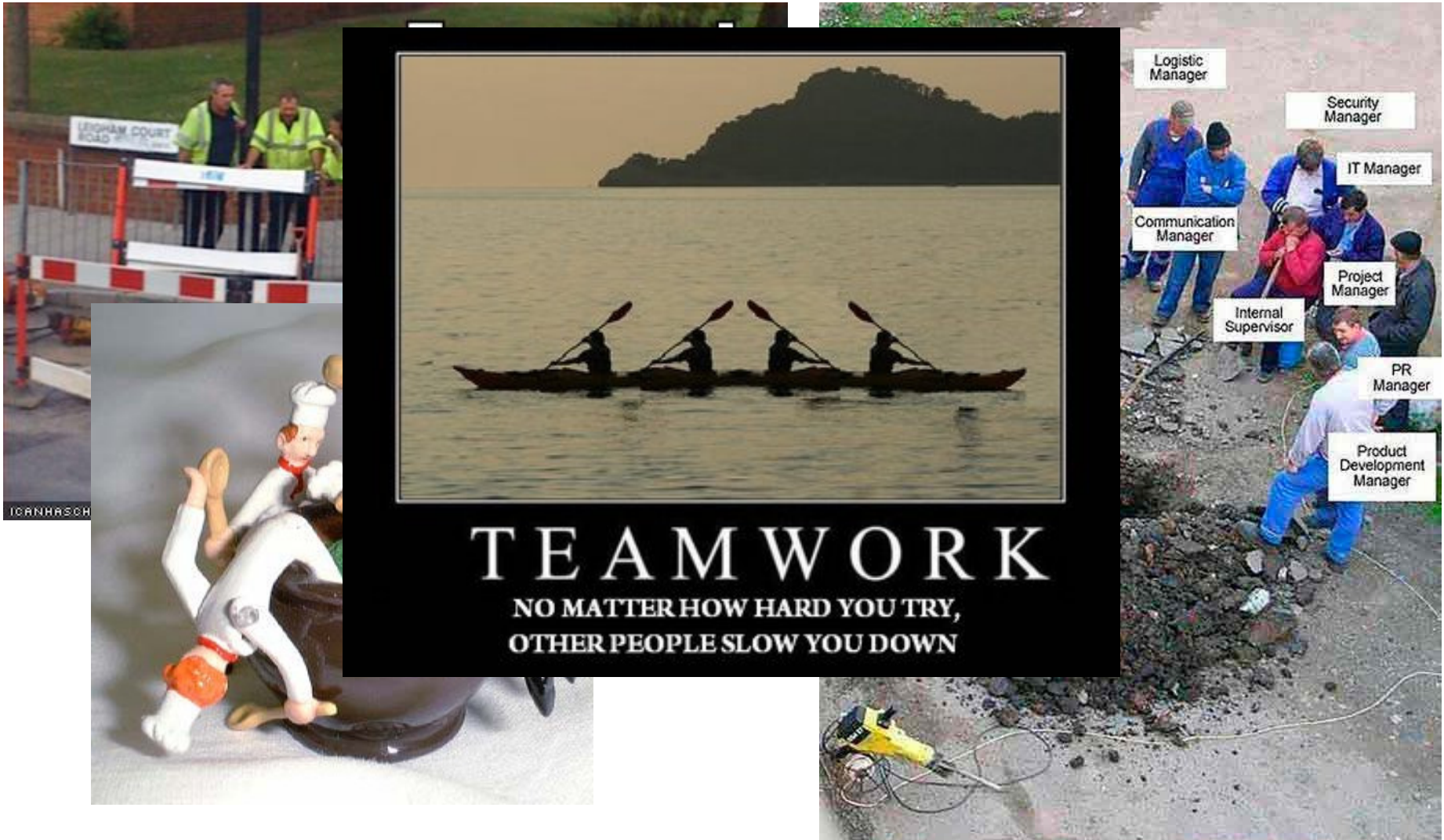Favors solo programming    Favors pair programming

**Figure 1. Meta-analyses of pair programming's effects on (a) quality, (b) duration, and (c) effort.**

"...The extent to which group **performance**
exceeds that of individuals depends on
the group's **composition** and the **tasks' characteristics**."



TEAMWORK

NO MATTER HOW HARD YOU TRY,
OTHER PEOPLE SLOW YOU DOWN

Logistic Manager
Security Manager
IT Manager
Communication Manager
Project Manager
Internal Supervisor
PR Manager
Product Development Manager

# Conclusion: When to use PP

| Programmer expertise | Task complexity | Use PP? |
| --- | --- | --- |
| Junior | Easy | Yes, provided that increased quality is the main goal |
| | Complex | Yes, provided that increased quality is the main goal |
| Intermediate | Easy | No |
| | Complex | Yes, provided that increased quality is the main goal |
| Senior | Easy | No |
| | Complex | No, unless you're sure that the task is too complex to be solved satisfactorily by an individual senior programmer |

?