Measurement in Higher Maturity Organizations: What's Different and What's Not?

Dennis R. Goldenson

27 July 2004

Today's Talk

Why does measurement matter?

What characterizes measurement in high maturity organizations?

How can we expedite things?

CarnegieMellon
Software Engineering Institute

CMMI ®

Why Does Measurement Matter?

What gets measured gets done!

- Or so we believe…

But what do we know that's convincing to the skeptics?

- Know thy users, for they are not you!

Well, more capable measurement <u>can</u> pay off

What Do We Mean By Success?

More than longevity and persistence over time!

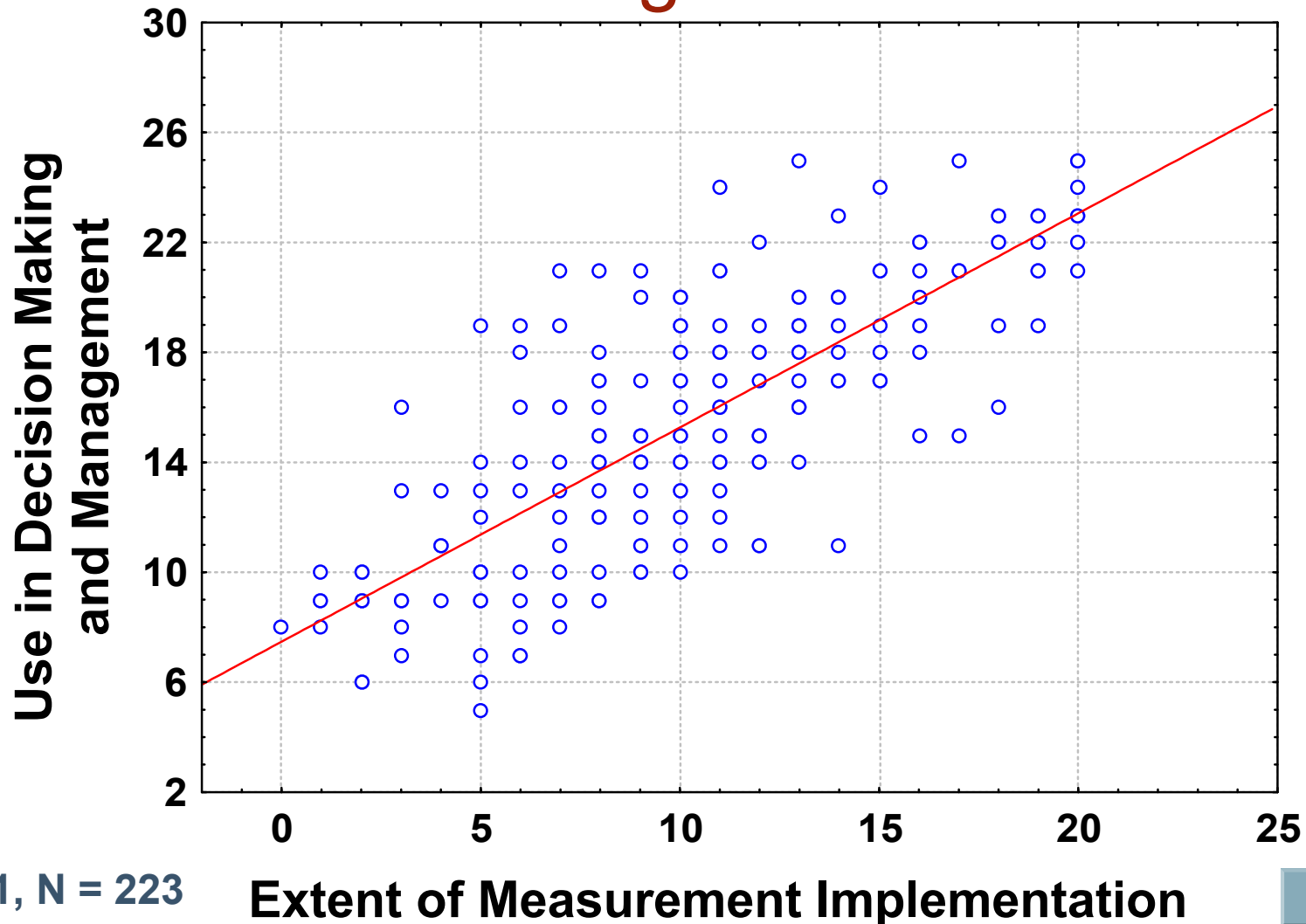- Technically defensible shelfware is not enough...

Regular use in decision making

Improvements in organizational performance

- Demonstrable impact on business value needed to justify continued investment
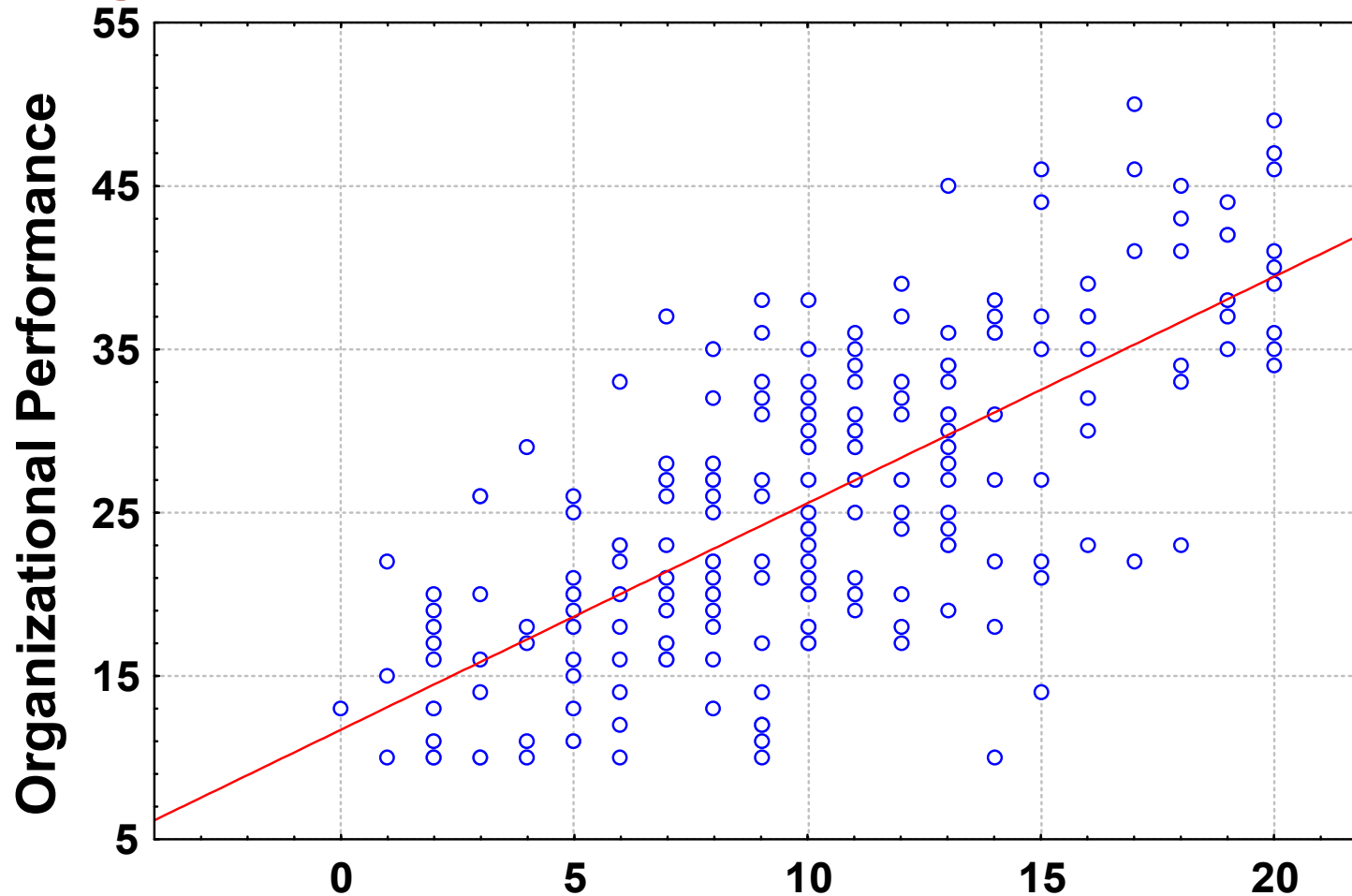
Measurement Implementation & Use in Decision Making

$R^2 = .61$, N = 223

Extent of Measurement Implementation

Measurement Implementation & Organizational Performance

R² = .49, N = 223   **Extent of Measurement Implementation**

How Can We Account for Success?

Alignment with business goals

Organizational commitment and resource sufficiency

Technical characteristics of the measurement program

Alignment with Business Goals

Predictors of Use	r^2
Aligned with intended users	.42
Aligned with measurement providers	.21
Conflict among stakeholders	-.01

Organizational Commitment and Resource Sufficiency

Predictors of Use r^2

Management commitment	.47
Technical commitment	.10
Sufficient funding	.20
Measurement training quality	.18
Qualified measurement personnel	.20
Existence of a measurement "guru"	.07

Technical Characteristics of the Measurement Program

Predictors of Use r^2

Use of analytic methods .48

Availability of automated support .21

Well defined data gathering procedures .33

Data quality .28

A Simple Multivariate Accounting

Based on bivariate results & preliminary multivariate analyses:

- One simple MANOVA:

- Model includes only 3 predictor variables … one from each of the three sets initially considered

- Main effect: about two thirds of observed variance in criterion index

Some multicolinearity

But variance explained noticeably higher than any of single bivariate relationships

Found no significant interaction effects

$R^2 = .66$, N = 223

Today's Talk

Why does measurement matter?

What characterizes measurement in high maturity organizations?

How can we expedite things?

Measurement in High Maturity Organizations

By definition…

- Attention to organizational issues
- Bringing processes under management control
- Attention to models
- Causal analysis & proactive piloting

At ML 3

- Focus on organizational definitions & a common repository

At ML 4

- Improve process adherence

(Especially at) ML 5

- Enhance & improve the processes themselves

How Well Do They Do It?

Well, it depends

Classes (if not nuances) of problems persist

- Even as organizational maturity increases

E.g., what about enterprise measures?

- How do you roll up measures from projects to enterprise relevance?
  - Asked by sponsor at a (deservedly) ML 5 organization
- Remains a pertinent, and difficult, issue for us as measurement experts today
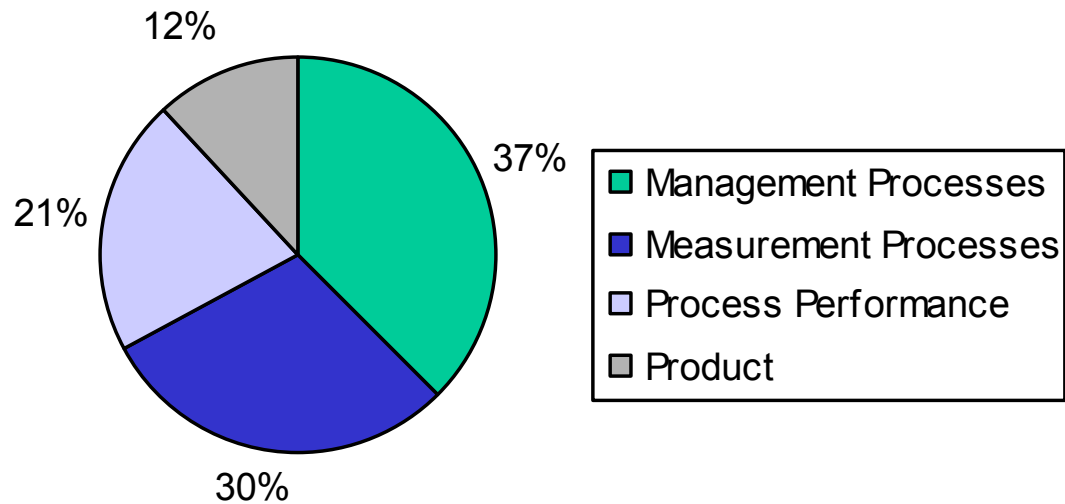
From SW-CMM® Appraisal Findings

663 appraisals

- 19 February 1987 through 28 June 2003
- 1350 weaknesses and opportunities for improvement that included the root word "measure"

Typical measurement related findings

- Lack of a consistent approach for capturing quality and productivity measurement data and comparing actuals with forecasts.
- There is no common understanding, definition and measurement of Quality Assurance.
- Test coverage data is inconsistently measured and recorded.
- Measurements of the effectiveness and efficiency of project management activities are seldom made.

Grouped Measurement Findings

12%

37%

21%

30%

☐ Management Processes
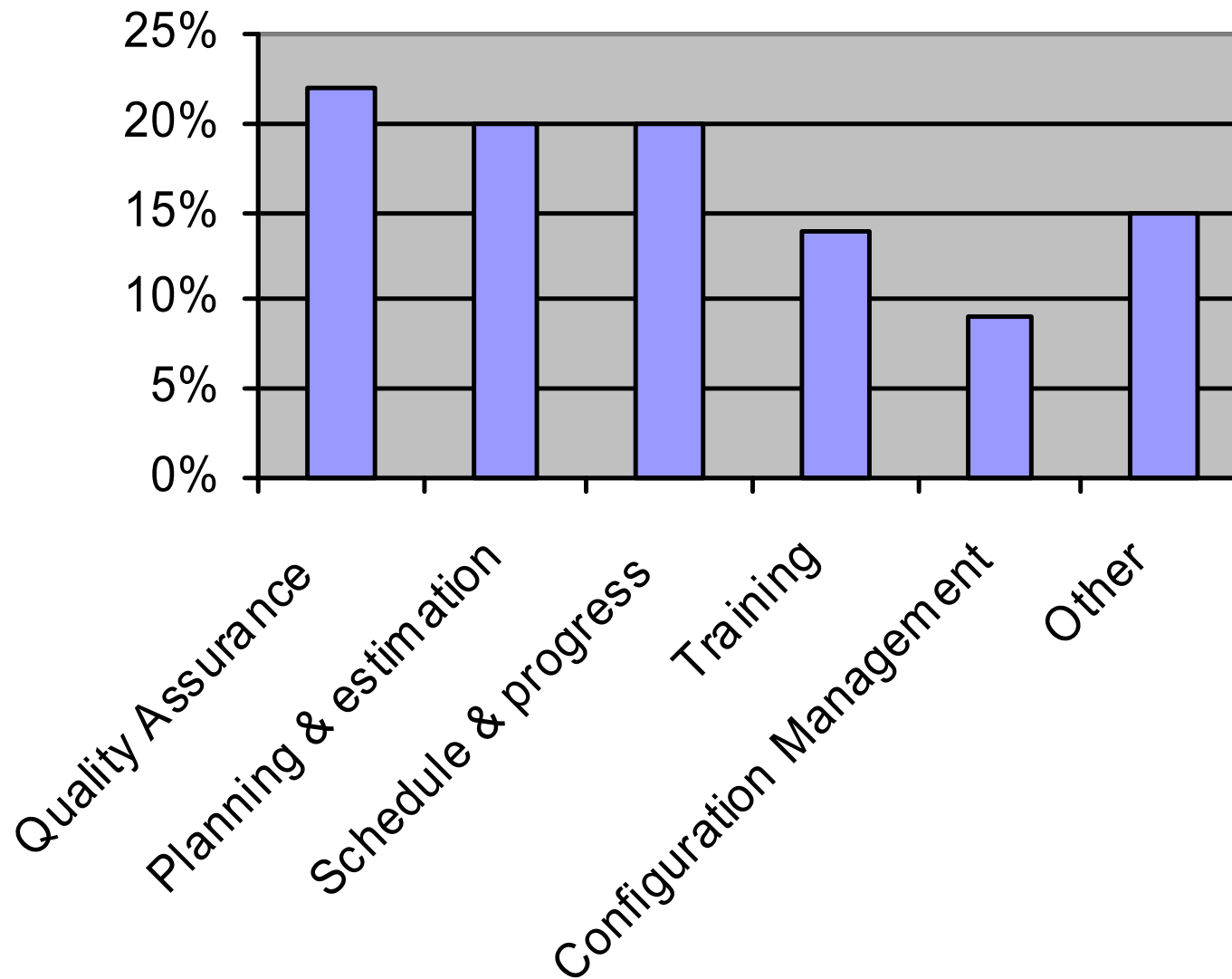■ Measurement Processes
☐ Process Performance
☐ Product

Appraisal findings typically arranged by KPA or other CMM model content

Not surprisingly: Largest of four groups addresses management

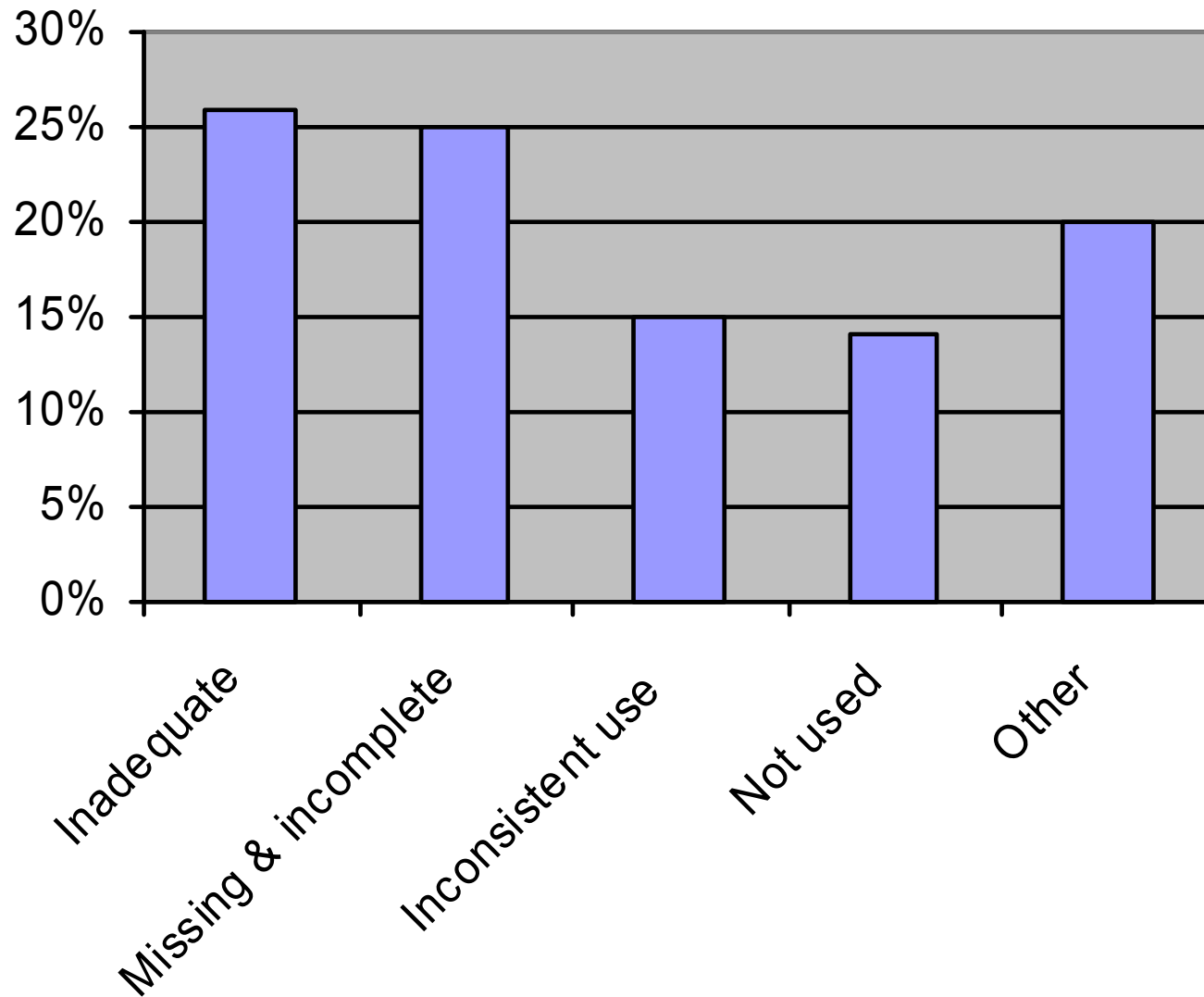- Difficulties with, or lack of use, of measurement for management purposes

Measurement of Management Processes

Measurement Processes Themselves

A Little More Detail

Measurement findings particularly noteworthy

- Appraisers tend to focus on model structure & content
- Measurement related content in SW-CMM considerably less explicit & complete than CMMI®

**26%: Existing measures inadequate for intended purposes**

- Findings are terse, but…
- Many or most seem to say measurement is poorly aligned with business & technical needs

"Other" category includes:

- Improvement of measurement processes (43 instances)
- Inter group activities related to measurement (34)
- Measurements misunderstood / not understood (12)
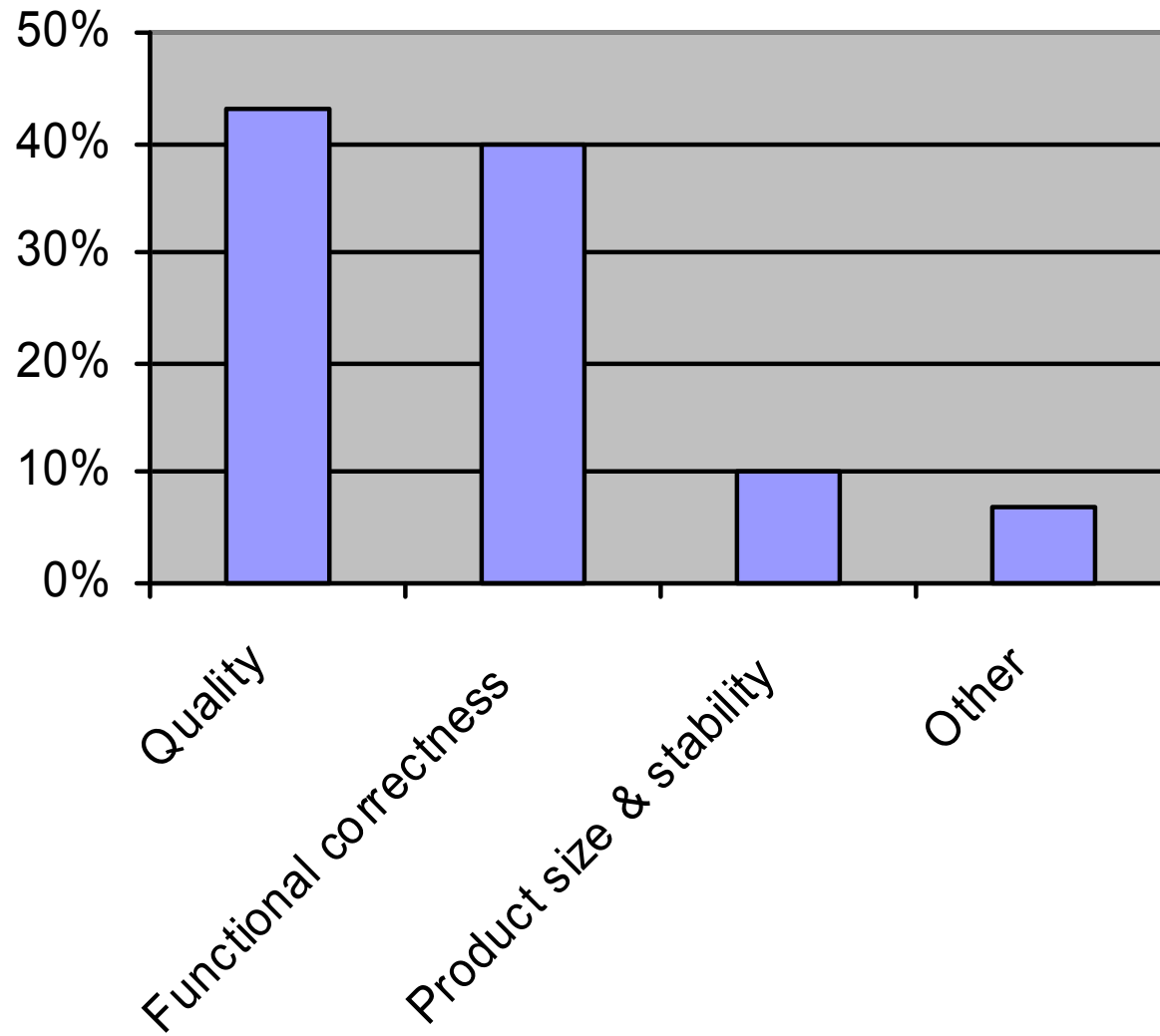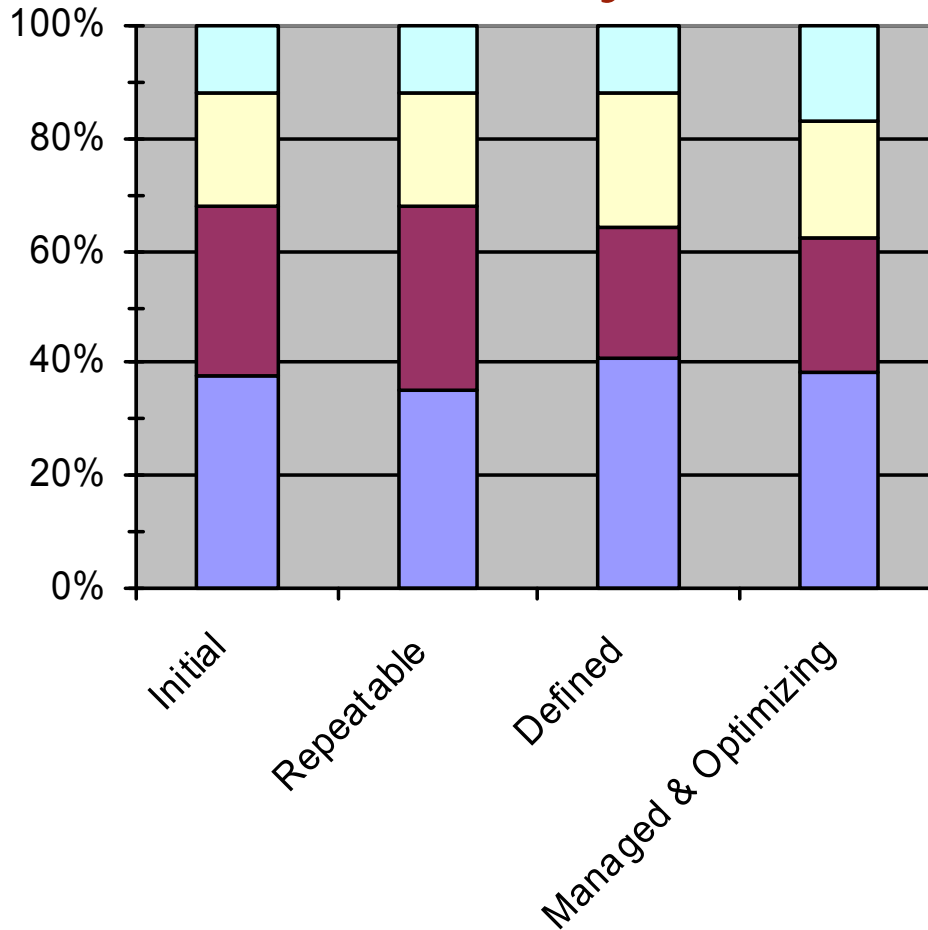- Leadership in the organization (3)

Process Performance
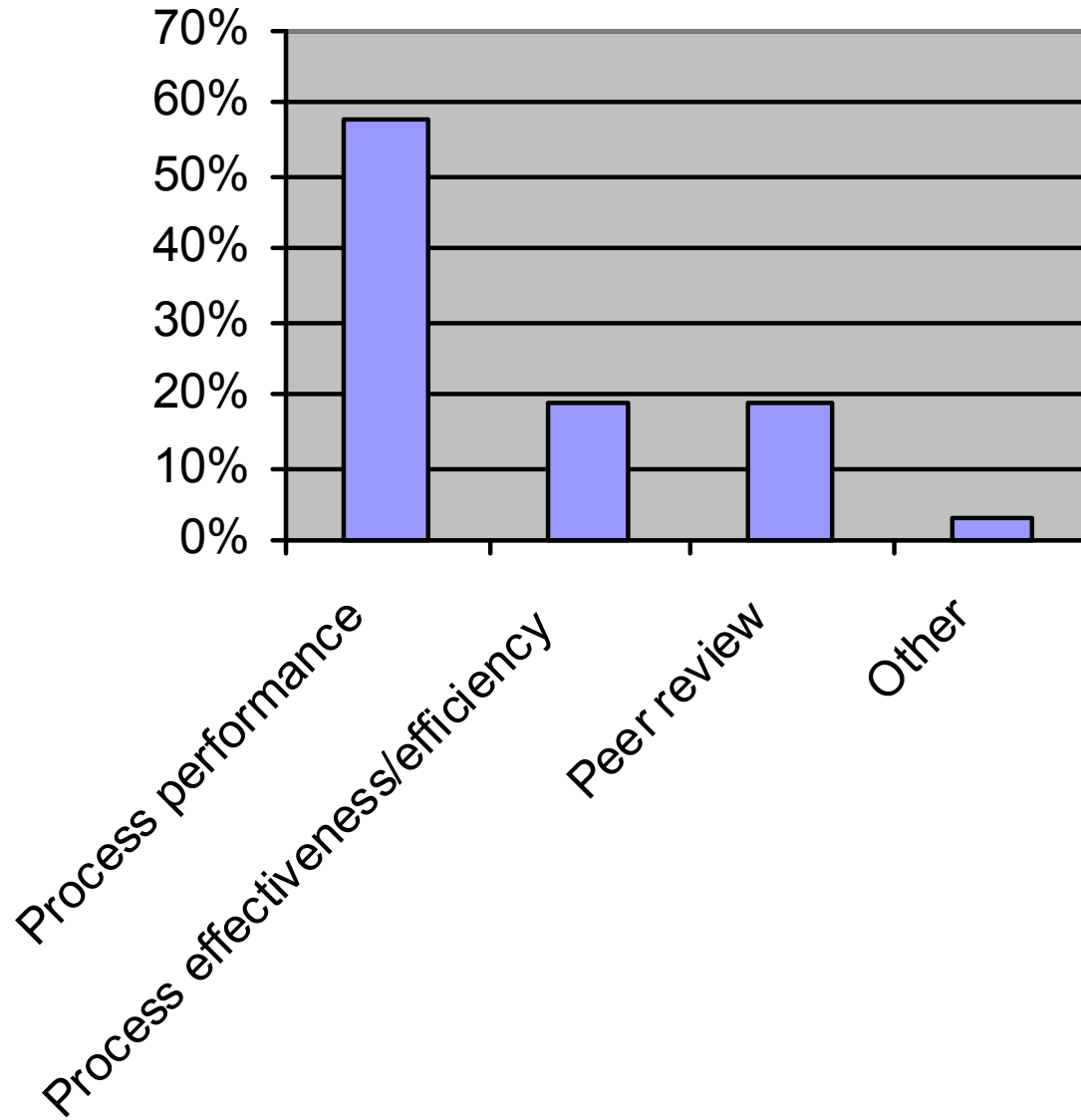
Product Quality & Technical Effectiveness

Differences by Maturity Level?

All four groups remain problematic throughout

- Including the measurement process itself
  - Nature of difficulties may differ
  - But proper enactment & institutionalization remains a problem for higher maturity organizations
- Similar pattern for process performance
  - Particularly pertinent at maturity levels 4 and 5
  - But noticeable proportions also address similar issues in lower maturity organizations

Chart legend:
- ☐ Management processes
- ☐ Measurement processes
- ☐ Process performance
- ☐ Product

Chart x-axis labels: Initial, Repeatable, Defined, Managed & Optimizing

What Typically Gets Measured?

Heavily influenced by SW-CMM

CMM models focus first on project planning & management

- Estimation (not always so well done)
- Monitoring & controlling schedule & budget

Followed by engineering

- Of course, some do focus on defects early …

What Changes as Organizations Mature?

Measurement definitions & procedures improve

- Measures get more finely grained, e.g., defect classification, insertion, find, fix and repair costs
- Project performance & quality measures are coupled explicitly with separate measures of process adherence & performance

Processes become better defined

- Sometimes influenced by being measured
- Routine reliance on quantitative management, causal analysis & piloting enhance process discipline

But…

- Serious attention to measurement often is delayed, if ever considered

There's Still Room for Improvement

Quantitative Process Management still emphasizes statistical process control (SPC)

- That's a good thing after all!
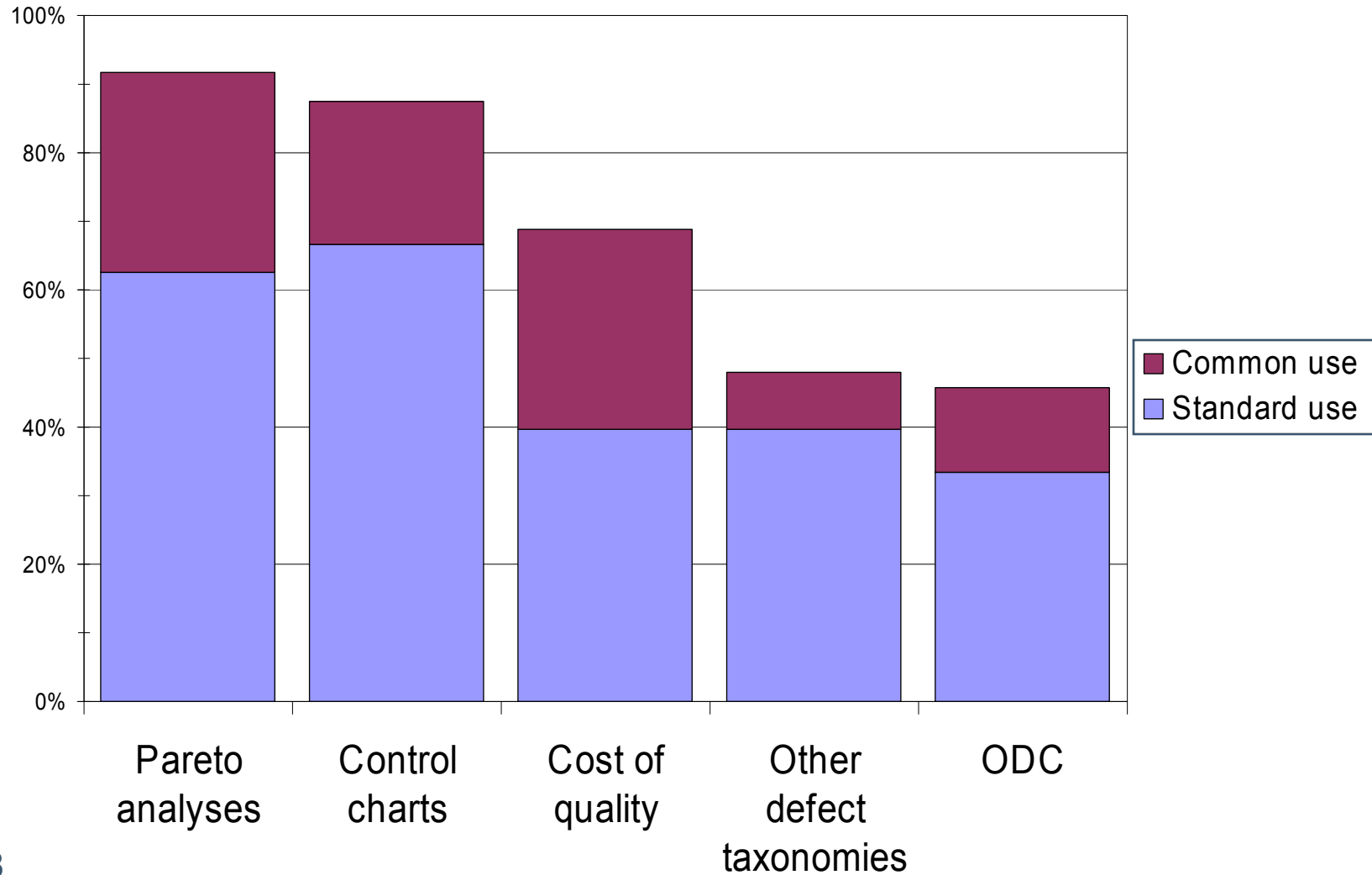- But there's a lot more out there too

Non SPC techniques <u>are</u> used

- Six Sigma
- Orthogonal Defect Classification
- Regression
- ANOVA

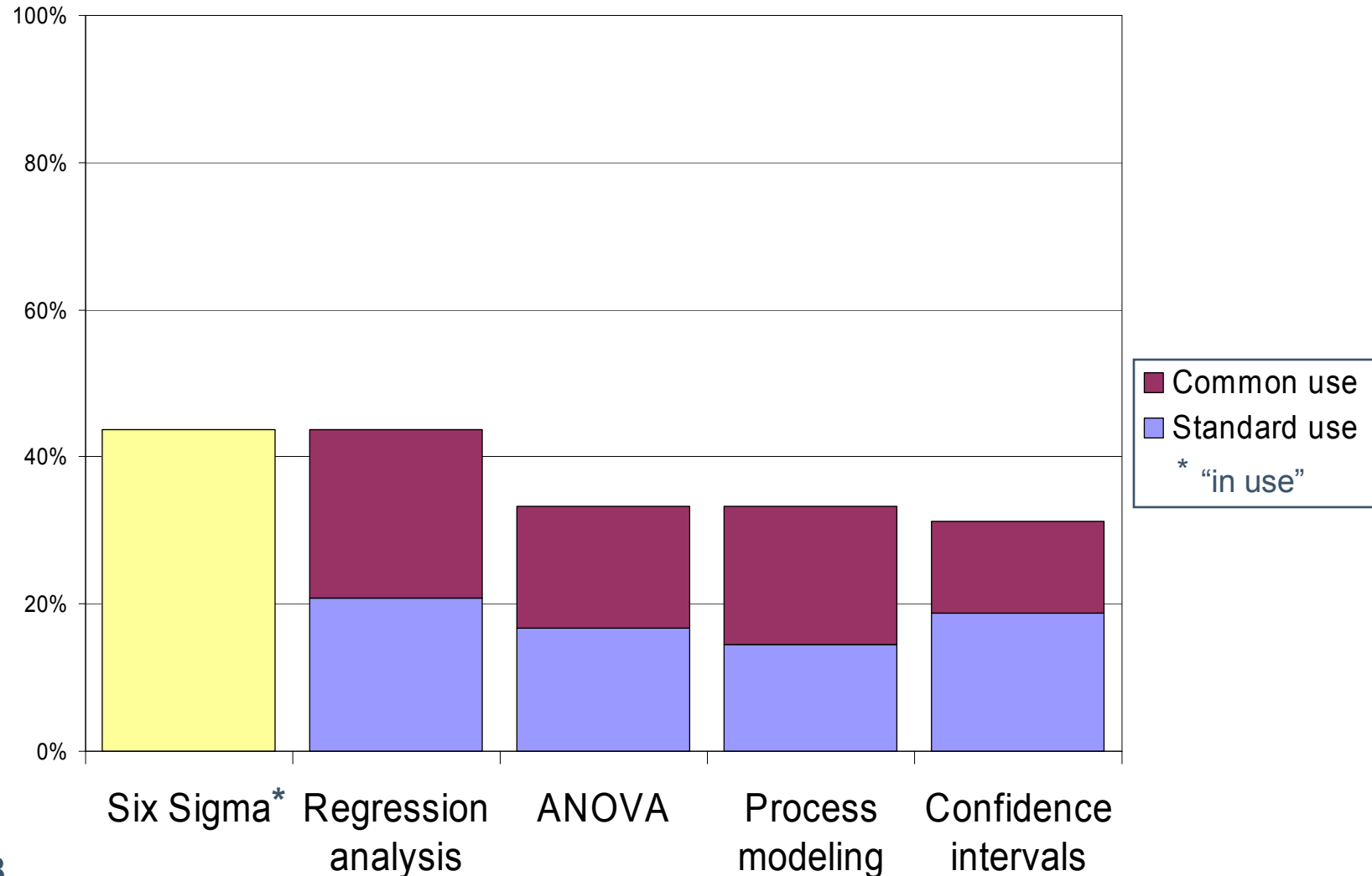Yet higher maturity organizations often don't have a particularly broad analytic tool kit
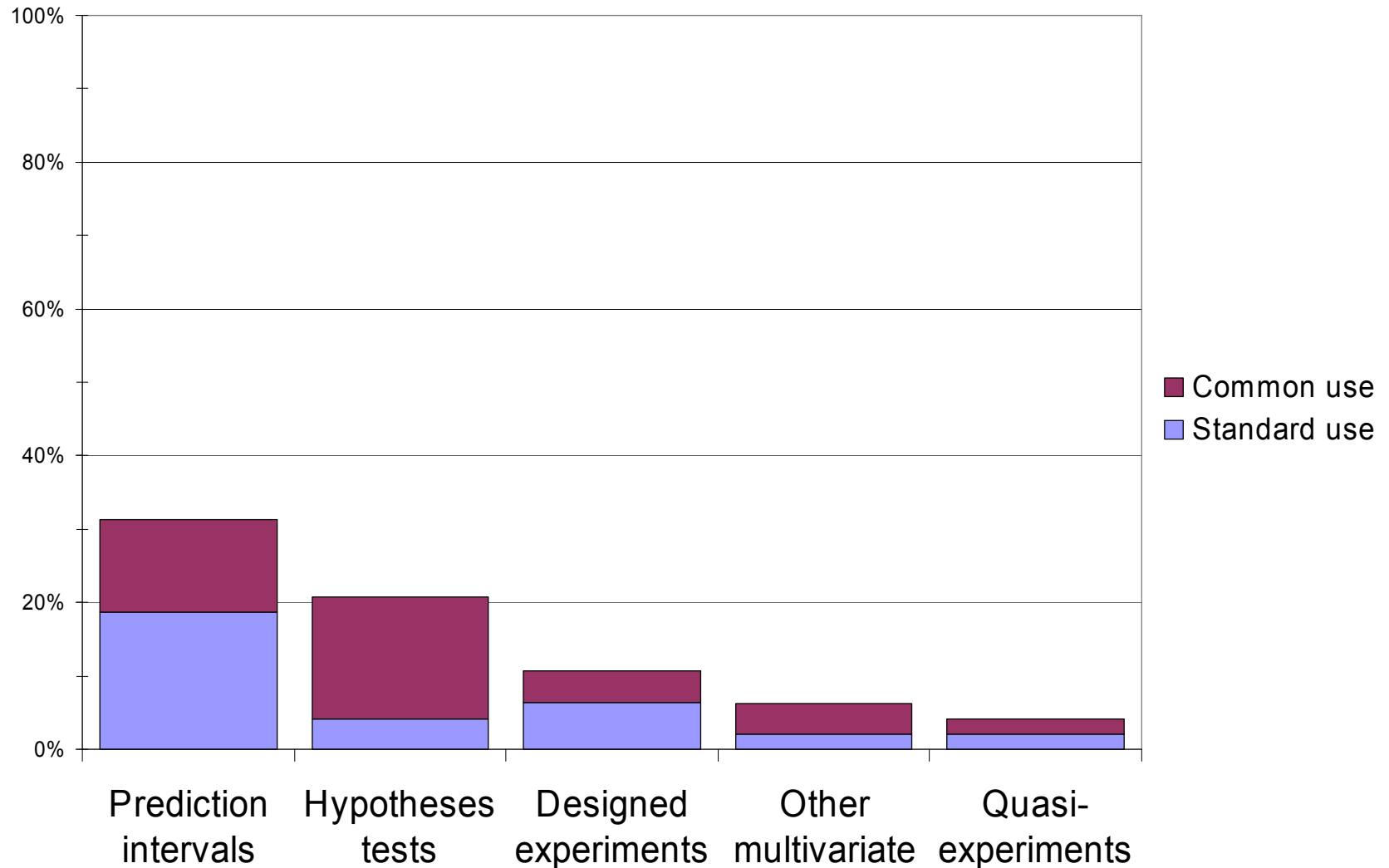
High Maturity Use of Analytics –1

N = 48

High Maturity Use of Analytics –2

High Maturity Use of Analytics –3

Legend:
- Common use
- Standard use

Categories: Prediction intervals, Hypotheses tests, Designed experiments, Other multivariate, Quasi-experiments

N = 48

Today's Talk

Why does measurement matter?

What characterizes measurement in high maturity organizations?

How can we expedite things?

How Can We Do Better?

Measurement and Analysis is at CMMI Maturity Level 2

- Put there to get it right from the start
- Lots of favorable anecdotes, but…
  - Intent not yet well understood by process champions
  - And we still need better (measurement based) evidence

The bulk of the measurement content is at Maturity Level 3 & above … mostly at levels 4 & 5

Why wait?

- Causal thinking is (or should be) the essence of statistics 101
- The problem is keeping the management commitment in an *ad hoc*, reactive environment
- But, it can be done…

Measurement Done Early and Well

Two examples (reported under non disclosure)

Level 1 organization used Measurement and Analysis:

- Significantly reduced the cost of quality in one year
- Realized an 11 percent increase in productivity, corresponding to $4.4M in additional value
- 2.5:1 ROI over 1st year, with benefits amortized over less than 6 months

Level 2 organization used Causal Analysis and Resolution:

- 44 percent defect reduction following one causal analysis cycle
- Reduced schedule variance over 20 percent
- $2.1 Million in savings in hardware engineering processes
- 95 percent on time delivery

Aligning Measurement & Information Needs

CMM based measurement always got done

- However much was required by appraisers…
- But less likely to be used if divorced from the real improvement effort

Organizations still struggle, even at higher Maturity Levels

- Need a marriage of domain, technical & measurement knowledge
- Yet, measurement often assigned to new hires with little deep understanding or background in domain or measurement

How can we do better?

- GQ(I)M when the resources & commitment are there
- Prototype when they aren't … or maybe always
- May be easier in small settings because of close communications & working relationships

Performance Models

Called out explicitly in CMM and CMMI

- Especially at Maturity Levels 4 & 5
- But, what do they (usually) mean?
  - Often poorly understood
  - Little more than informal causal thinking

We (the measurement mafia) can do better

- In fact, some have done better…
- By applying modeling & simulation models to process improvement
  - Not common, but it has been & is being done
  - 10 years ago, as an integral part of one organization's process definition, implementation & institutionalization
  - The organization is gone now, but that's another (measurement) story

Modeling & Simulation

Analytic method can be applied in many domains

- Estimate when experimentation, trial & error are impractical
- By being explicit about variables & relationships, process definitions, business & technical goals & objectives

Use it to:

- Proactively inform decisions to begin, modify or discontinue a particular improvement or intervention
- By comparing alternatives & alternative scenarios

Of course, there's still a need for measurement…!

- To estimate model parameters based on fact
- To validate and improve the models

What's Next? (Or, what do I think should be next…?)

Can early attention to measurement really expedite organizational maturation?

- That's part of the rationale for Six Sigma too
- But it's not well, or at least widely, understood
    - How can we demonstrate the relationship?
    - What data & research designs do we need?

Cause and effect?

- Do the analyses early and well

Pay more attention to performance measures

- Including enterprise measures
- And including quality attributes beyond defects
  (See ISO/IEC Working Group 6, ISO 25000)

And don't ignore (or wait to do) modeling and simulation

Contact

Dennis R. Goldenson

- Software Engineering Institute
- Pittsburgh, PA 15213-3890
- 1.412.268.8506
- dg@sei.cmu.edu

Back Pocket

Slides follow…

Carnegie Mellon
Software Engineering Institute

From Symposium 2000

The Prescribed Order: Items in Presumptive Maturity Level 2

- Schedule e.g., actual versus planned completion, cycle time (85%)

- Cost/budget e.g., estimate over-runs, earned value (77%)

- Effort e.g., actual versus planned staffing profiles (73%)

- Field defect reports (68%)

- Product size e.g., in lines of code or function points (60%)

The Prescribed Order: Items in Presumptive Maturity Level 3

- Test results or other trouble reports (81%)

- Data, documentation, and reports are saved for future access (76%)

- Organization has common suite of software measurements collected and/or customized for all projects or similar work efforts (67%)

- Results of inspections and reviews (58%)

- Customer or user satisfaction (56%)

The Prescribed Order: Items in Presumptive Maturity Level 4

- **Quality assurance and audit results** (54%)
- **Comparisons regularly made between current project performance and previously established performance baselines and goals** (44%)
- **Requirements stability** e.g., number of customer change requests or clarifications (43%)
- **"Other" quality measures** e.g., maintainability, interoperability, portability, usability, reliability, complexity, reusability, product performance, durability (31%)
- **Process stability** (31%)
- **Sophisticated methods of analyses are used on a regular basis** e.g., statistical process control, simulations, latent defect prediction, or multivariate statistical analysis (14%)
- **Statistical analyses are done to understand the reasons for variations in performance** e.g., variations in cycle time, defect removal efficiency, software reliability, or usability as a function of differences in coverage and efficiency of code reviews, product line, application domain, product size, or complexity (14%)

The Prescribed Order: Items in Presumptive Maturity Level 5

- Experiments and/or pilot studies are done prior to widespread deployment of major additions or changes to development processes and technologies (38%)

- Evaluations are done during and after full-scale deployments of major new or changed development processes and technologies (e.g., in terms of product quality, business value, or return on investment) (27%)

- Changes are made to technologies, business or development processes as a result of our software measurement efforts (20%)

◄   ◄

Use   Impact

Exceptions

Exceptions
- Level 5
  - Experiments and/or pilot studies (38%)
- Level 4
  - Sophisticated analyses (14%)
  - Statistical analyses of variations (14%)
- Level 3
  - Test results or other trouble reports (81%)
  - Data, documentation, and reports saved (76%)
- Level 2
  - Product size (60%)

May be due to

- Measurement error in this study

- Differences among organizational contexts

- Subtleties in "natural" order

Where Do the Exceptions Occur?

Of the possible comparisons with presumptively lower level items …

Level 3
- 14% fail level 2 items

Level 4
- 6% fail level 3 items
- 4% fail level 2 items

Level 5
- 14% fail level 4 items
- 6% fail level 3 items
- 6% fail level 2 items

Use in Making Management and Development decisions

Cronbach's
alpha = .79

- Monitoring and managing individual projects or similar work efforts
- Used by individual engineers, programmers and other practitioners
- "Software measurement and data analysis are an integral part of the way we normally do business"
- "The need for objective evidence about quality and performance is highly valued in our organization"
- "There is resistance to doing measurement around here e.g., people think of it as unnecessary, extra work, unfair, or an imposition on the way they do their work" (reverse coded)

Use in Decision Making

Percent of Respondents
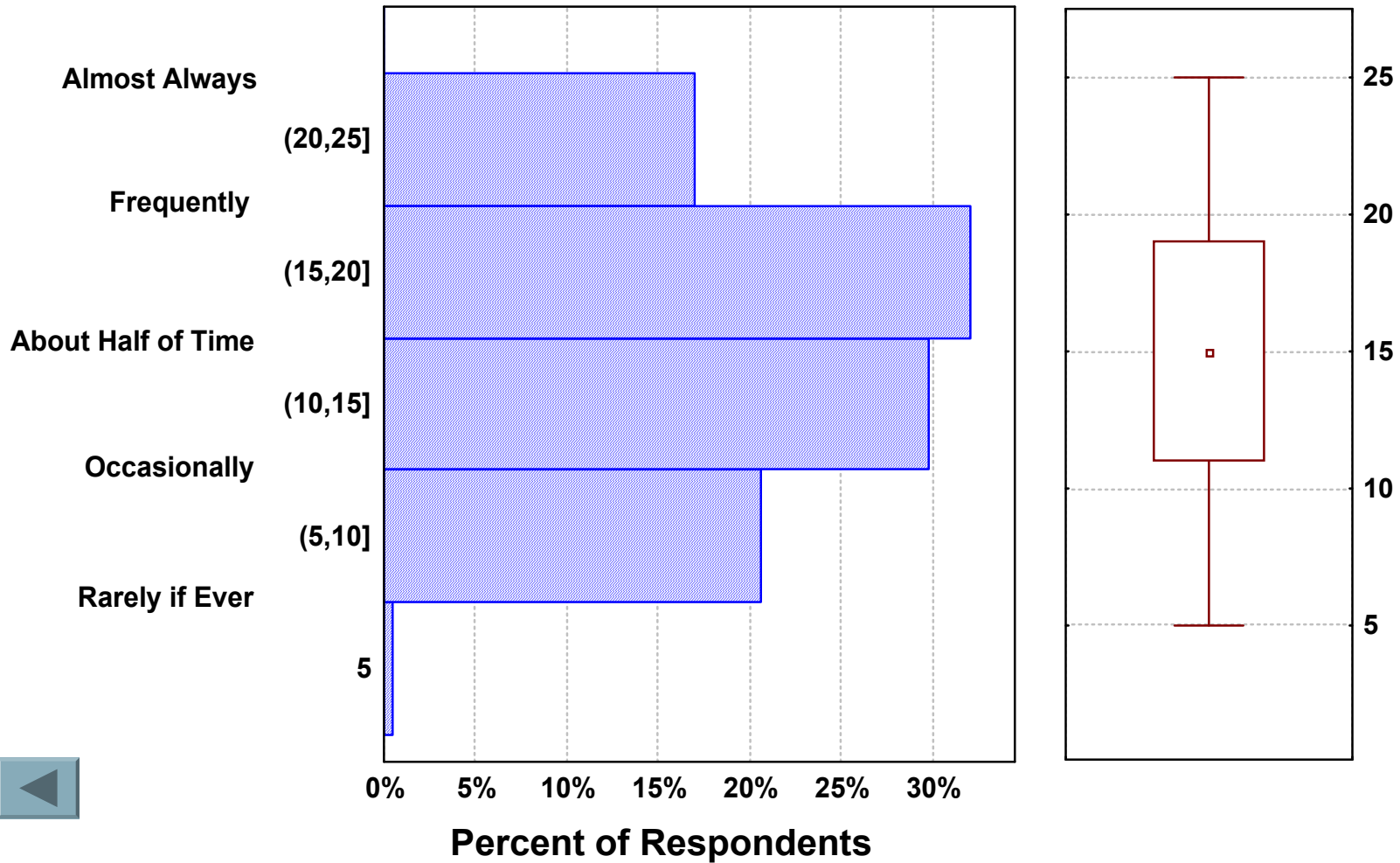
"In your judgment, how much has the use of software measurement improved your organization's performance?

Cronbach's alpha = .94

- More accurate budget estimates or ability to reduce costs
- More accurate schedule estimates or ability to reduce cycle time
- Better adherence to customer or user requirements or improved customer satisfaction
- Fewer software defects, faults or failures

- **Better functionality or user interface**
- **Better over-all quality of products and services**
- **Improved staff productivity or reduced rework**
- **More informed judgments about the adoption or improvement of work processes and technologies**
- **Better work processes**
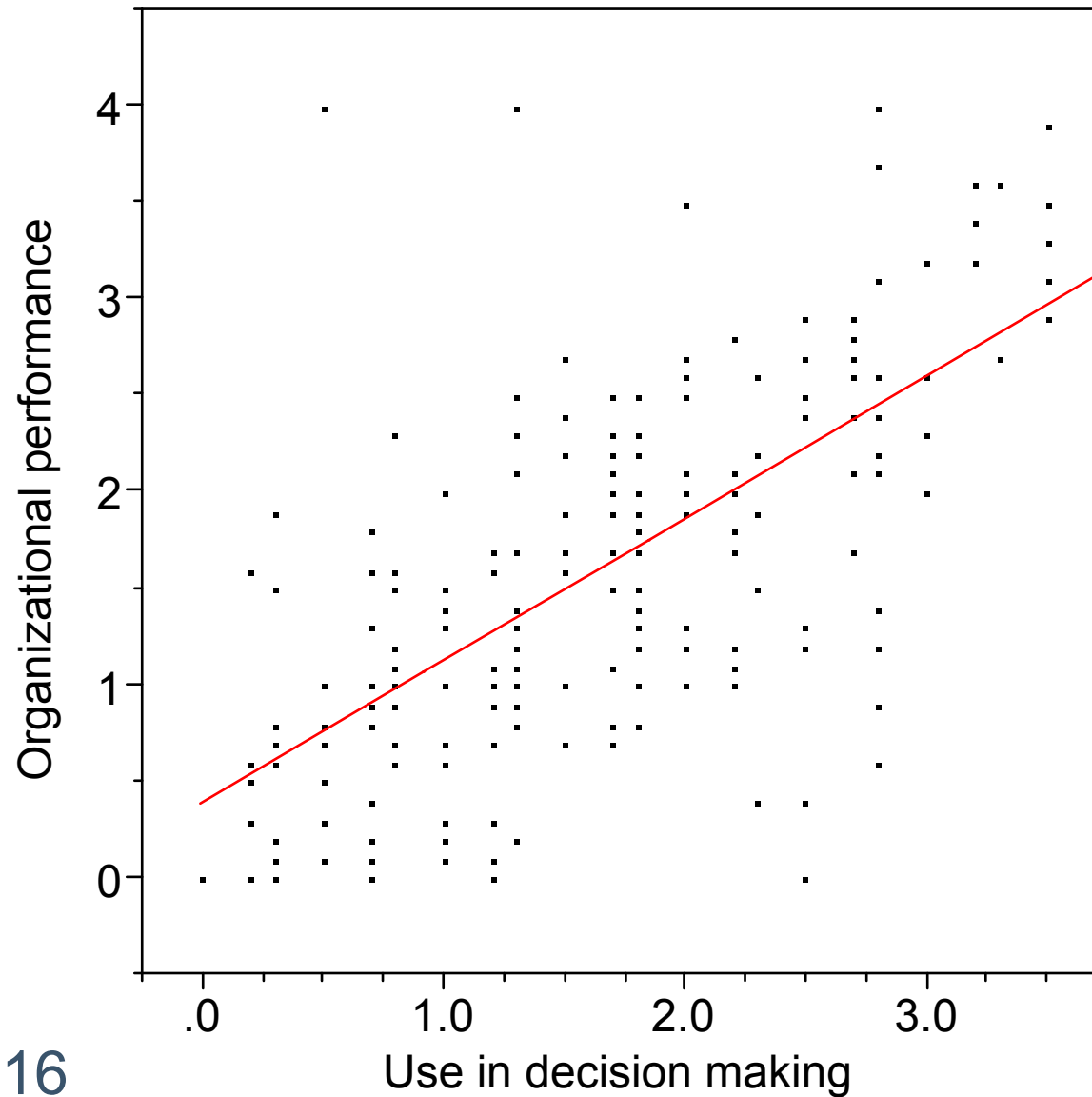- **Better strategic decision-making**

Organizational Performance

**Percent of Respondents**
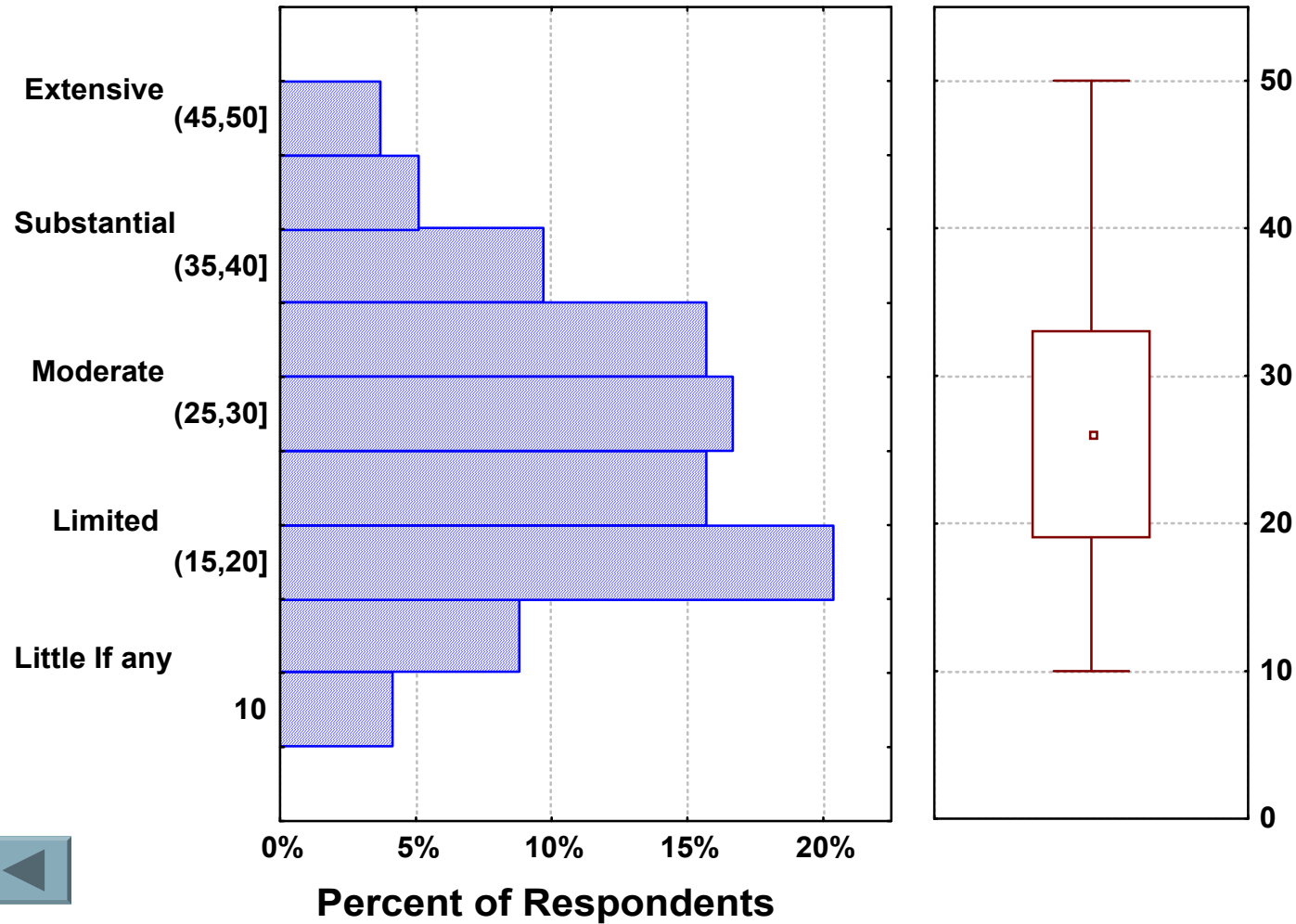
From Metrics 1999

N = 216

Aligned with Intended Users

"How would you characterize the involvement of various potential stakeholders in setting goals and deciding on plans of action for measurement in your organization?"

- Senior enterprise and organization level managers
- Project level managers
- Individual engineers, programmers or other practitioners
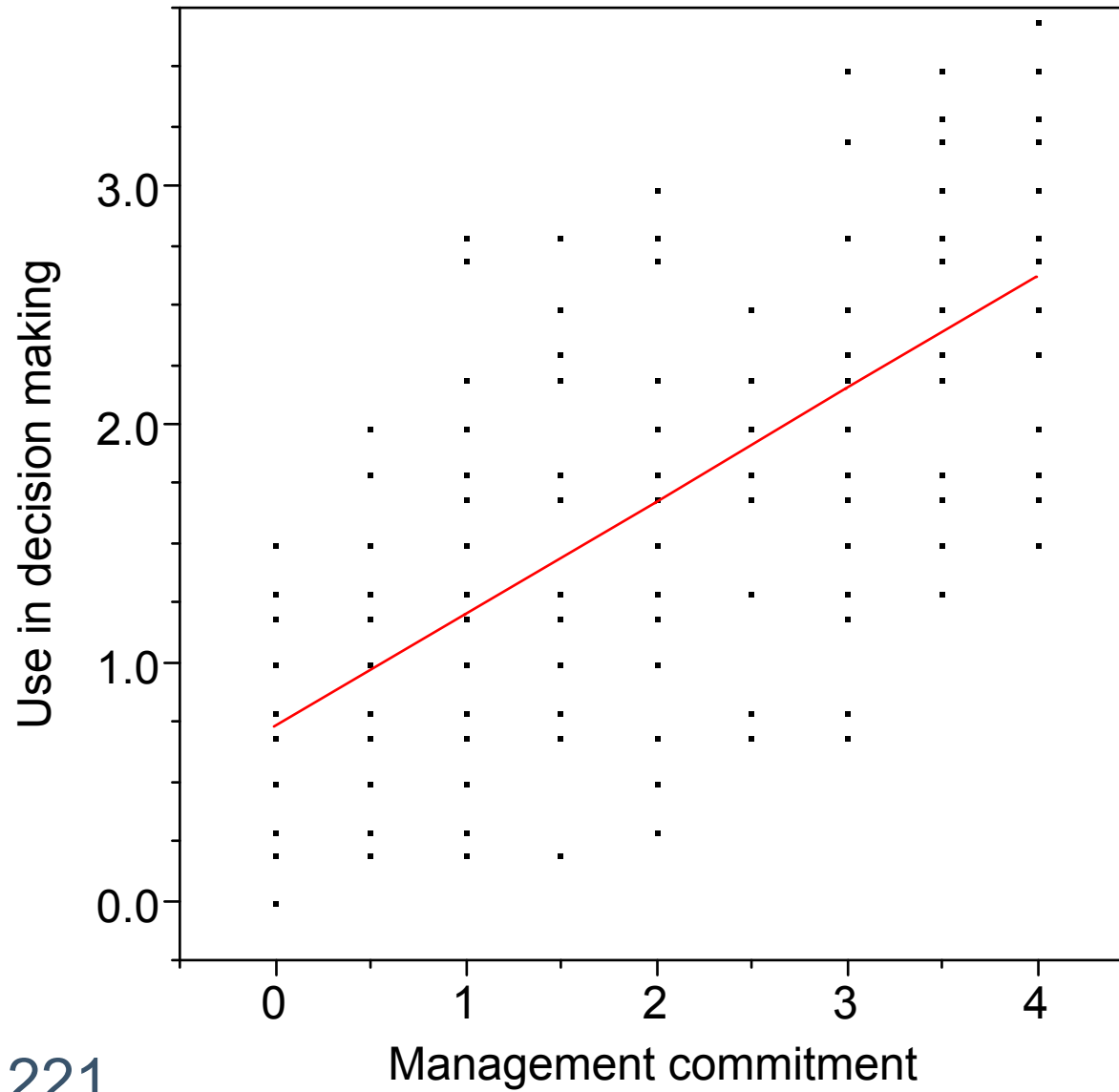- Business support units, e.g. finance, marketing
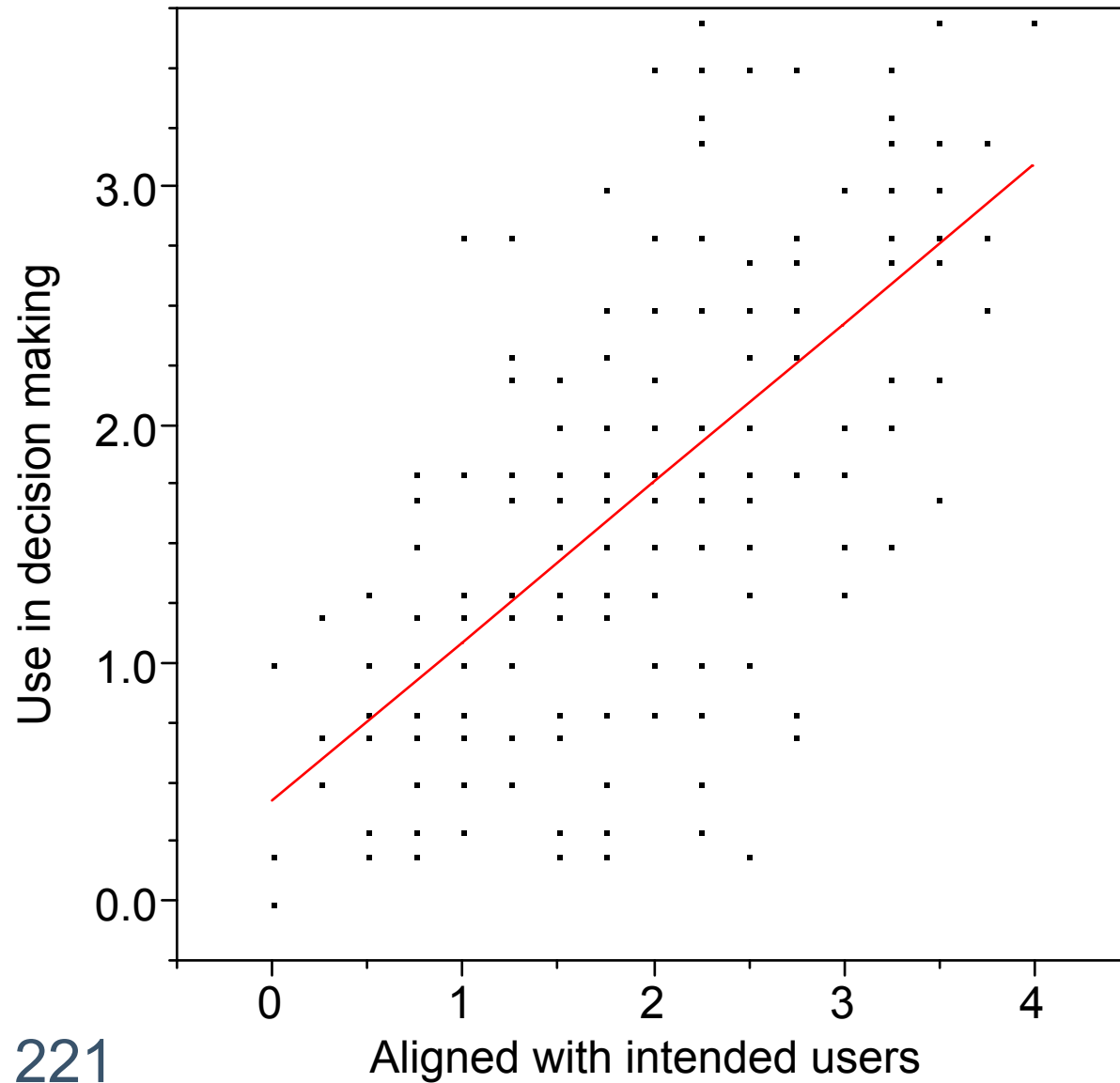
N = 221

Management Commitment

"Management regularly monitors the progress of software measurement activities"

"Management clearly demonstrates commitment to measurement"

N = 221

Use of Analytic Methods -1

"Comparisons are regularly made between current project performance and previously established performance baselines and goals"

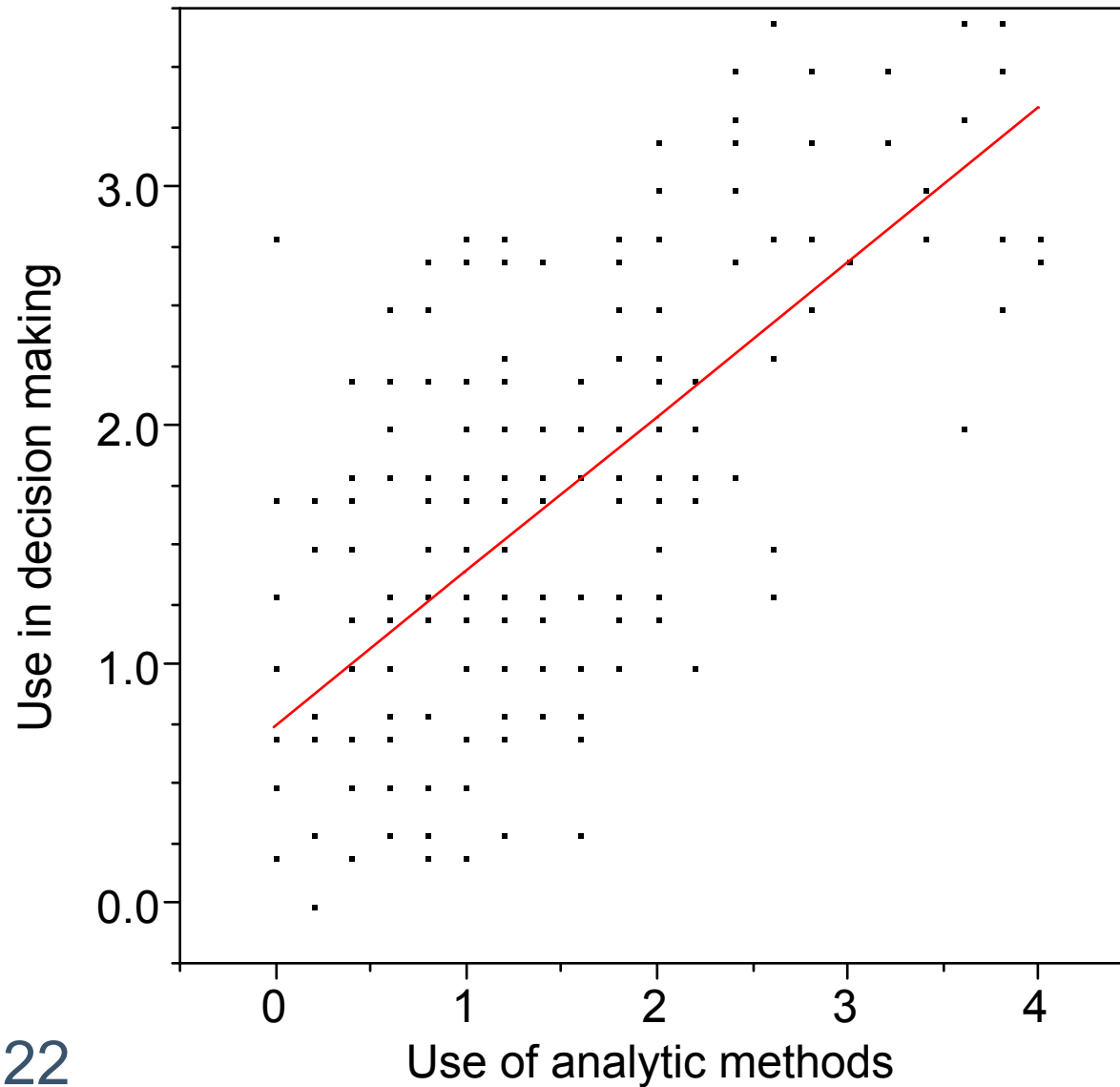**"Sophisticated methods of analyses are used on a regular basis"**

"Statistical analyses are done to understand the reasons for variations in performance"
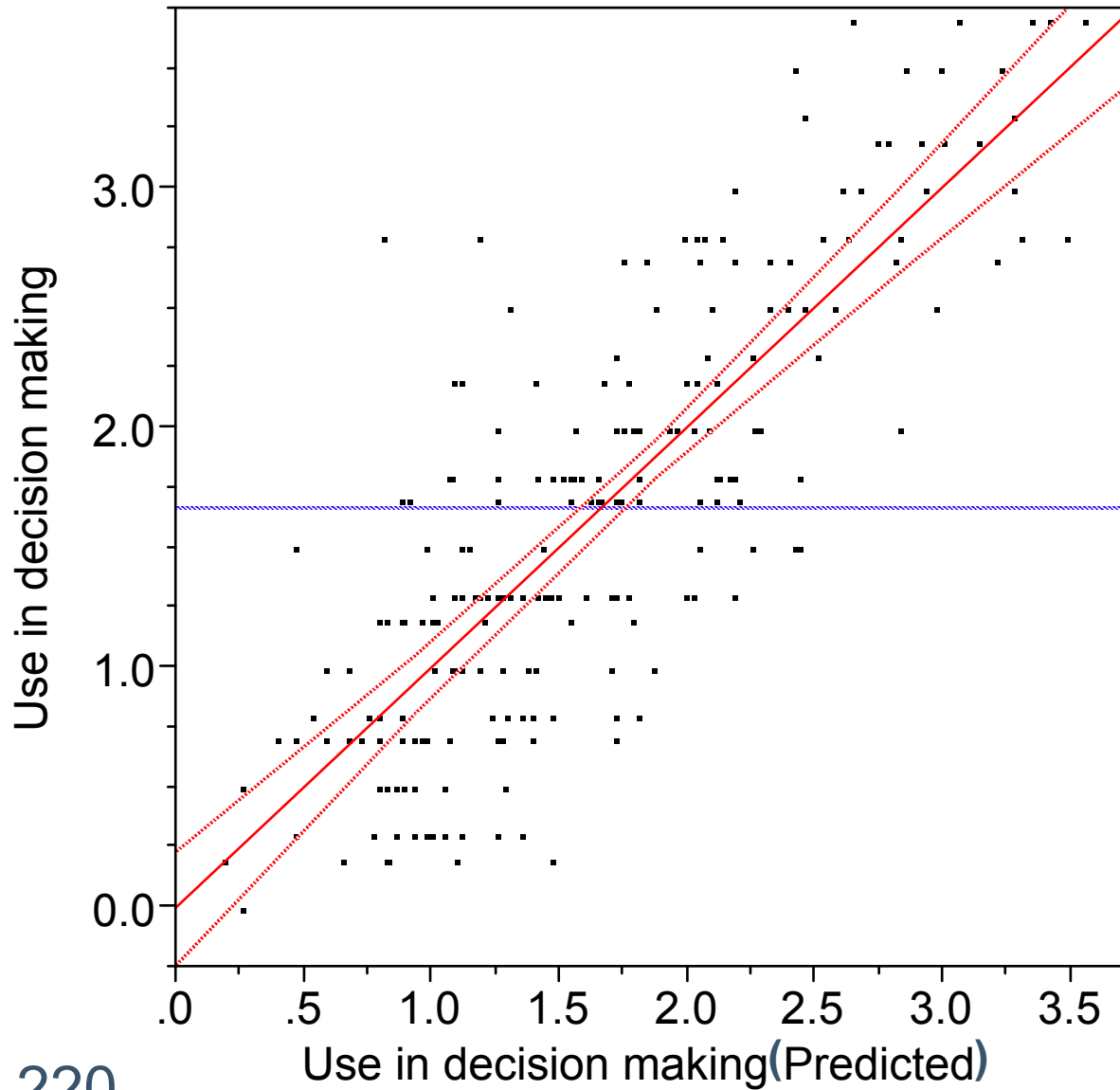
Use of Analytic Methods -2

"Experiments and/or pilot studies are done to prior to widespread deployment of major additions or changes to development processes and technologies"

"Evaluations are done during and after full-scale deployments of major new or changed development processes and technologies"

N = 222

N = 220