Understanding CMMI Measurement Capabilities & Impact on Performance: Results from the 2007 SEI State of the Measurement Practice Survey

Dennis R. Goldenson
Software Engineering Institute

CMMI Technology Conference
14 November 2007

Software Engineering Institute | **Carnegie Mellon**

Purpose & scope of the survey

Results

- É The respondents & their organizations

- É Measurement resources & infrastructure

- É Value added by measurement

- É Software measures used

- É Data quality & integrity

- É Organizational perspectives on software measurement

Summary, lessons learned & next steps

the State of Measurement Practice

Careful & well executed use of measurement & analysis

É Is a well accepted tenet in many fields of endeavor

É Including of course CMMI

Basic aims

É To inform management & technical decisions based on empirical evidence

É & to judge the results of those decisions once made

But, how well, and how frequently, are measurement practices put into effect in our own field?

chmarking

Benchmarking: The current state

- É Some professional & consulting organizations maintain repositories they use for establishing benchmarks & facilitating benchmarking activities
- É However, their measures & measurement definitions differ in many ways
- É In that sense, one cannot speak confidently about "industry standards+
- É Which is why the SEI has launched the Performance Benchmarking Consortium {as described at last years CMMI Technology Conference}

The state of the practice surveys

- É Aim to provide data that's not yet widely available
  - ô Updates of trends in typical use of measurement in software & systems engineering
  - ô To help projects & organizations judge their progress relative to others
- É But there *also* will be a continuing need to track qualitative as well as quantitative descriptions about the quality & frequency use of measurement in our field

Measurement Practice Survey

New this year

- Screening question to identify respondents whose organizations develop software but rarely if ever do measurement

- Questions about

  - Resources & infrastructure devoted to measurement

  - Practices to ensure data quality & integrity

  - Value added by doing measurement

  - The kinds of measures used by the responding organizations

Among other things, these questions allow us to make some useful comparisons by CMMI maturity level

me

1st survey described at last year's CMMI technology Conference

Similar results this year

- É Moderately strong relationships exist when comparing the replies of respondents based on:

    - ô Management versus staff roles

    - ô Industry *versus* government organizations

    - ô The United States *versus* other countries

    - ô Organization size

But that's a topic for another time

ment Capabilities & Performance Outcomes

Today's focus

- Provide evidence about the circumstances under which measurement capabilities and performance outcomes are likely to vary
- As a consequence of achieving higher levels of CMMI maturity

Most differences *are* consistent with expectations based on CMMI

- Which provides confidence in the validity of the model structure & content

However, the results also highlight areas where sometimes considerable room for improvement remains

- Even at maturity levels 4 and 5
- For example
    - A rather strong overall relationship between maturity level & use of measures about quality attributes
    - Little attention to quality attributes at the lower maturity levels
    - Yet, almost half of maturity level 4 & 5 respondents' organizations track quality attributes only occasionally at best

Random sample of SEI customers

- 944 valid email invitations to participate

Data collected 20 February through 10 April 2007

- Two reminders

Response rate

- 41% completed all or part of the questionnaire

- N = 384

- Individual questions answered by 75-97% of respondents

    - ~29 . 39% of the sample invitees

Purpose & scope of the survey

Results

- **The respondents & their organizations**
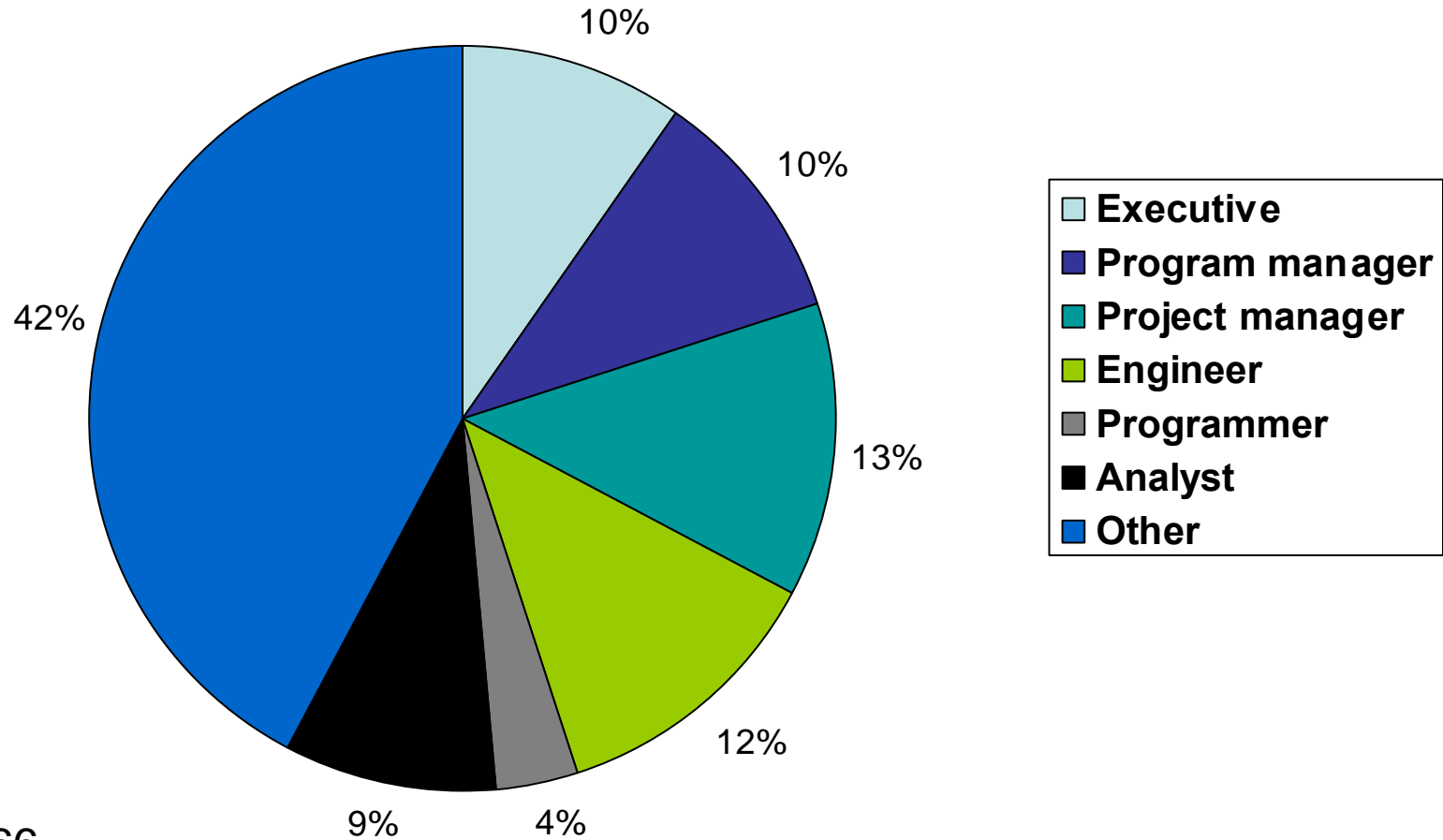- Measurement resources & infrastructure
- Value added by measurement
- Software measures used
- Data quality & integrity
- Organizational perspectives on software measurement

Summary, lessons learned & next steps

anization

10%

10%

Executive
Program manager
Project manager
Engineer
Programmer
Analyst
Other

42%

13%

12%

9%

4%

N = 366

hers?

= 8% of all those responding

26%

24%

9%

15%

6%

20%

Legend:
- ☐ **Quality**
- ■ **Process**
- ■ **Process + Quality**
- ■ **Consultant**
- ■ **Management**
- ■ **Other Others**
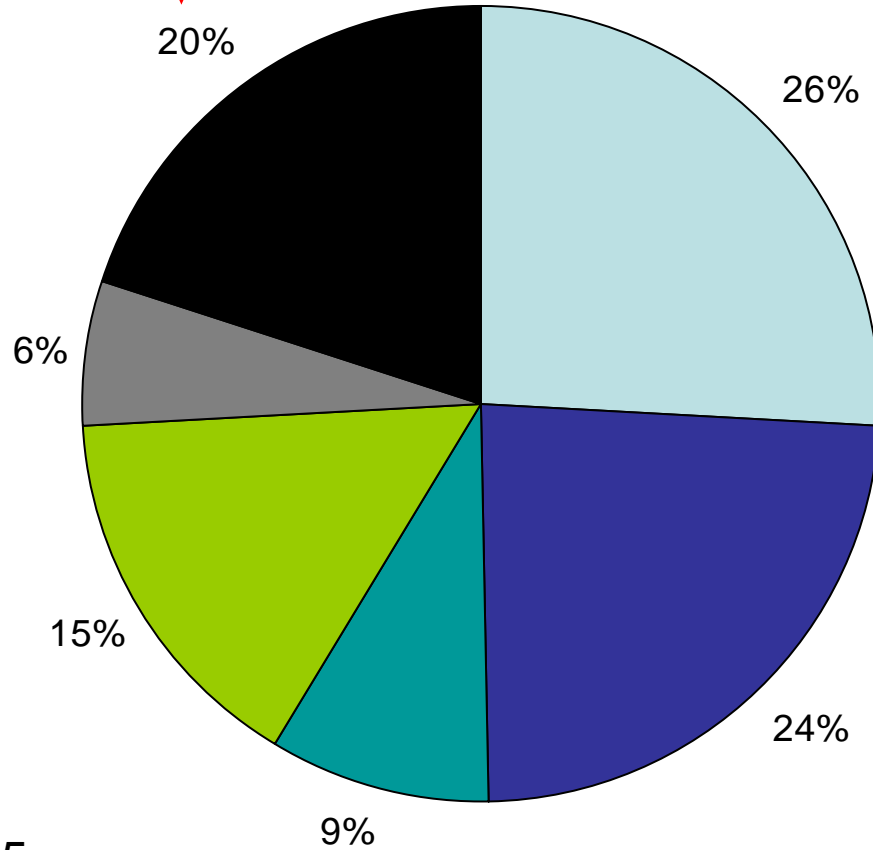
N = 155

e other others?

Process + Measurement	3
Measurement Specialist	1
Process + Quality + Measurement + Training	1
Quality + Process + Measurement	1
Training	6
Architect	4
Security	2
Testing	2

} 6

N = 31

One each:

- É Administrative support
- É Coach
- É Consultant + researcher
- É Engineering Manager + Process
- É Process + Project engineer
- É Program / team lead
- É Program manager + Quality + Process
- É Project manager + Quality
- É Project manager + Engineer
- É Not specified

N = 366

Legend:
- Commercial shrink-wrap
- Custom software development
- In-house or proprietary
- Defense contractor
- Other government contractor
- Defense or military organization
- Other government agency
- Consultancy
- Other

Values: 4%, 37%, 13%, 16%, 5%, 3%, 4%, 11%, 7%

Pie chart legend:
- United States
- India
- Japan
- France
- Germany
- United Kingdom
- Canada
- Netherlands
- All others

Pie chart percentages: 48%, 12%, 4%, 3%, 3%, 3%, 2%, 2%, 23%

N = 363

Percent

N = 364

Percent

N = 365

Maturity Level: Use of Measurement in the Organization

2%   3%   3%

8%   1% (Occasional)

30%   8%   22%   96%

34%   19%

          70%   75%

28%

ML1&DK   ML2   ML3   ML4&5
N = 151   N = 84   N = 59   N = 71

Gamma = .73   p < .0001

Legend:
- Don't know
- Rare or never
- Occasional
- Routine
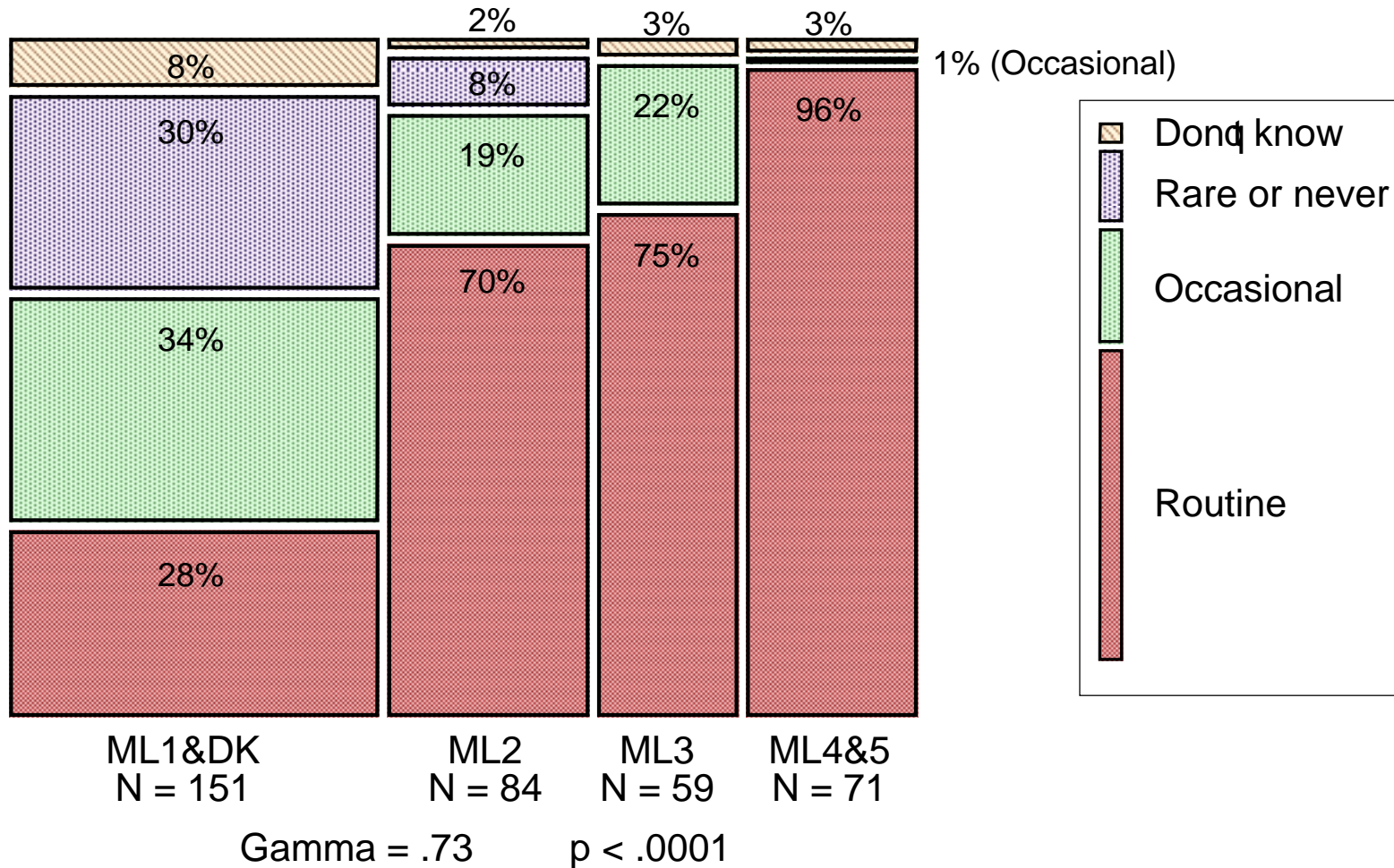
e results: The Respondents' Measurement Roles
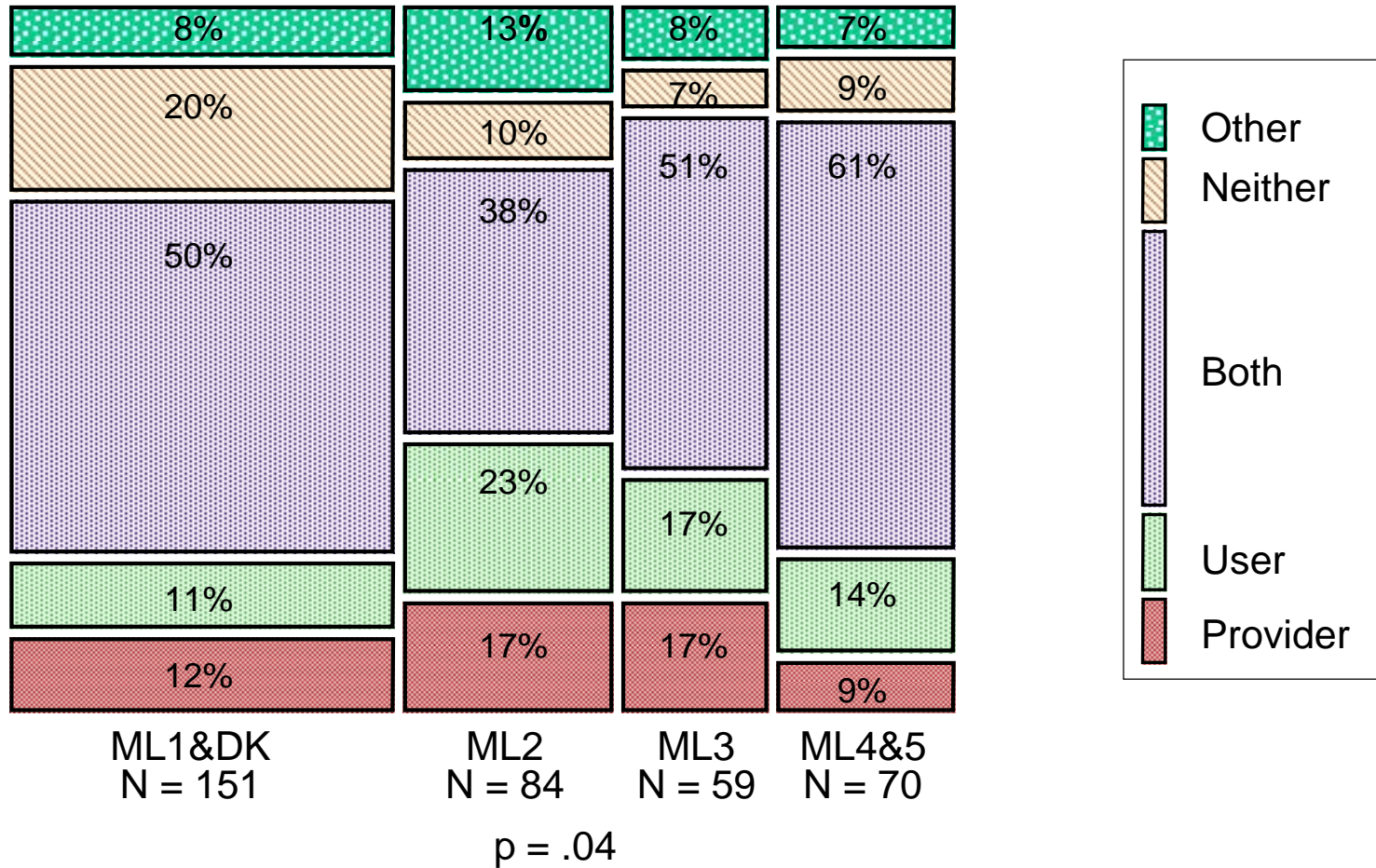
Purpose & scope of the survey

Results

- É The respondents & their organizations
- É **Measurement resources & infrastructure**
- É Value added by measurement
- É Software measures used
- É Data quality & integrity
- É Organizational perspectives on software measurement

Summary, lessons learned & next steps

...ent Work is Staffed

3%, 1%, 2% & 3% respectively

Legend:
- Other
- Don't know
- Organization wide group
- Project level
- A few key experts

	ML1&DK N = 78	ML2 N = 60	ML3 N = 58	ML4&5 N = 60
Other	9%	13%	12%	7%
Organization wide group	34%	34%	28%	20%
Project level	13%	19%	30%	50%
A few key experts	41%	33%	28%	20%

p < .006

dgets for Measurement

21%	18%	22%	28%
72%	65%	56%	34%
			38%
7%	18%	22%	

ML1&DK
N = 76

ML2
N = 68

ML3
N = 50

ML4&5
N = 61

Legend:
- Don't know
- No
- Yes

$p < .0001$
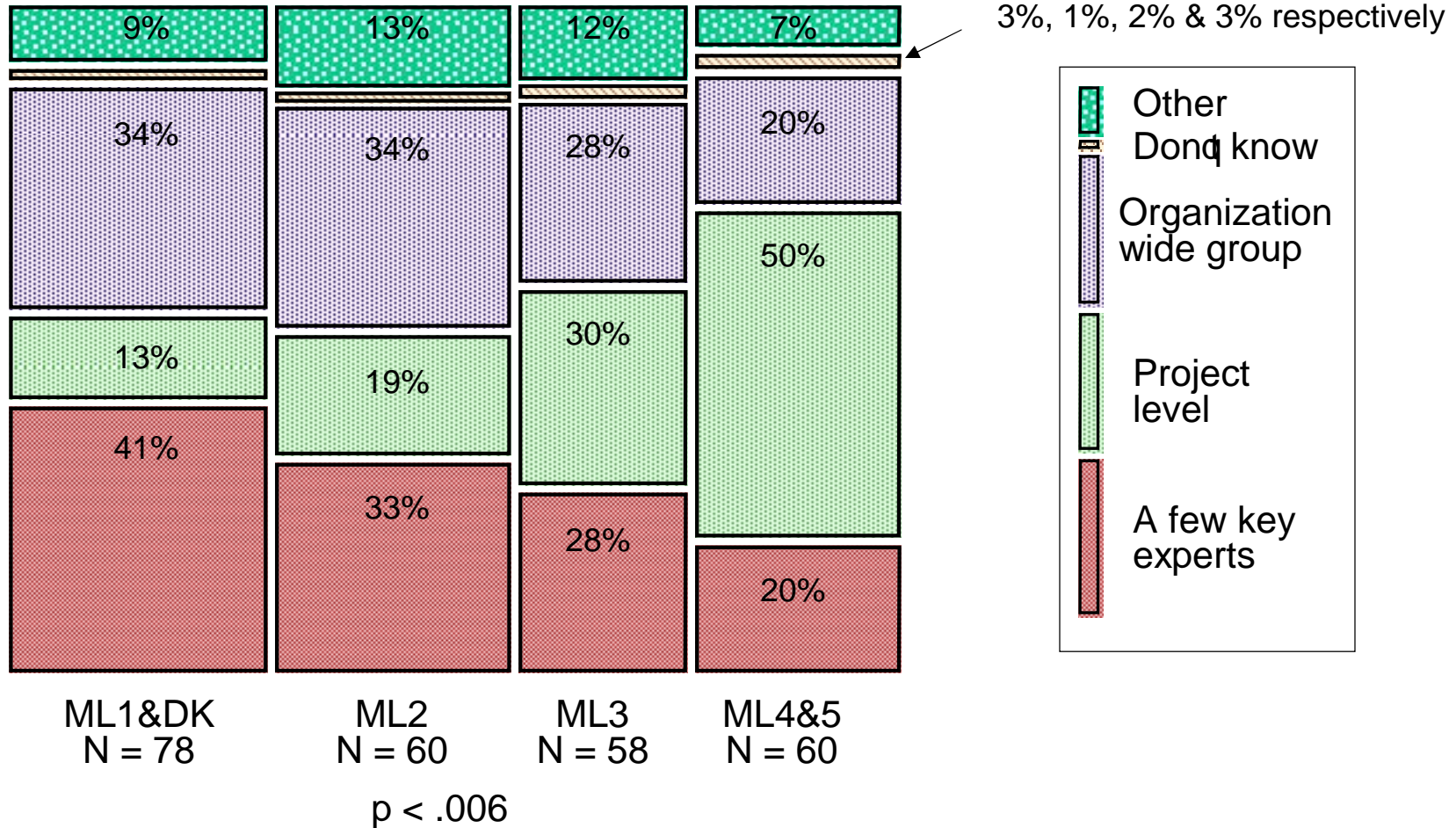
Qualified Measurement Staff

ML1&DK N = 76	ML2 N = 65	ML3 N = 50	ML4&5 N = 61	Legend
51%	26%	26%	11%	Rarely, never & don't know
30%	38%	34%	28%	Half the time & occasionally
18%	35%	40%	61%	Almost always & frequently

Gamma = .44   p < .0001

For:

- É Automated measurement support for data collection, data management, data analysis & reporting

- É Use of commercial measurement packages & tools

- É Existence of common, integrated organizational measurement repositories

- É Availability of measurement related training

Proportions sometimes vary across the distributions.

But there are consistent differences by maturity level.

Purpose & scope of the survey

Results

- É The respondents & their organizations
- É Measurement resources & infrastructure
- É **Value added by measurement**
- É Software measures used
- É Data quality & integrity
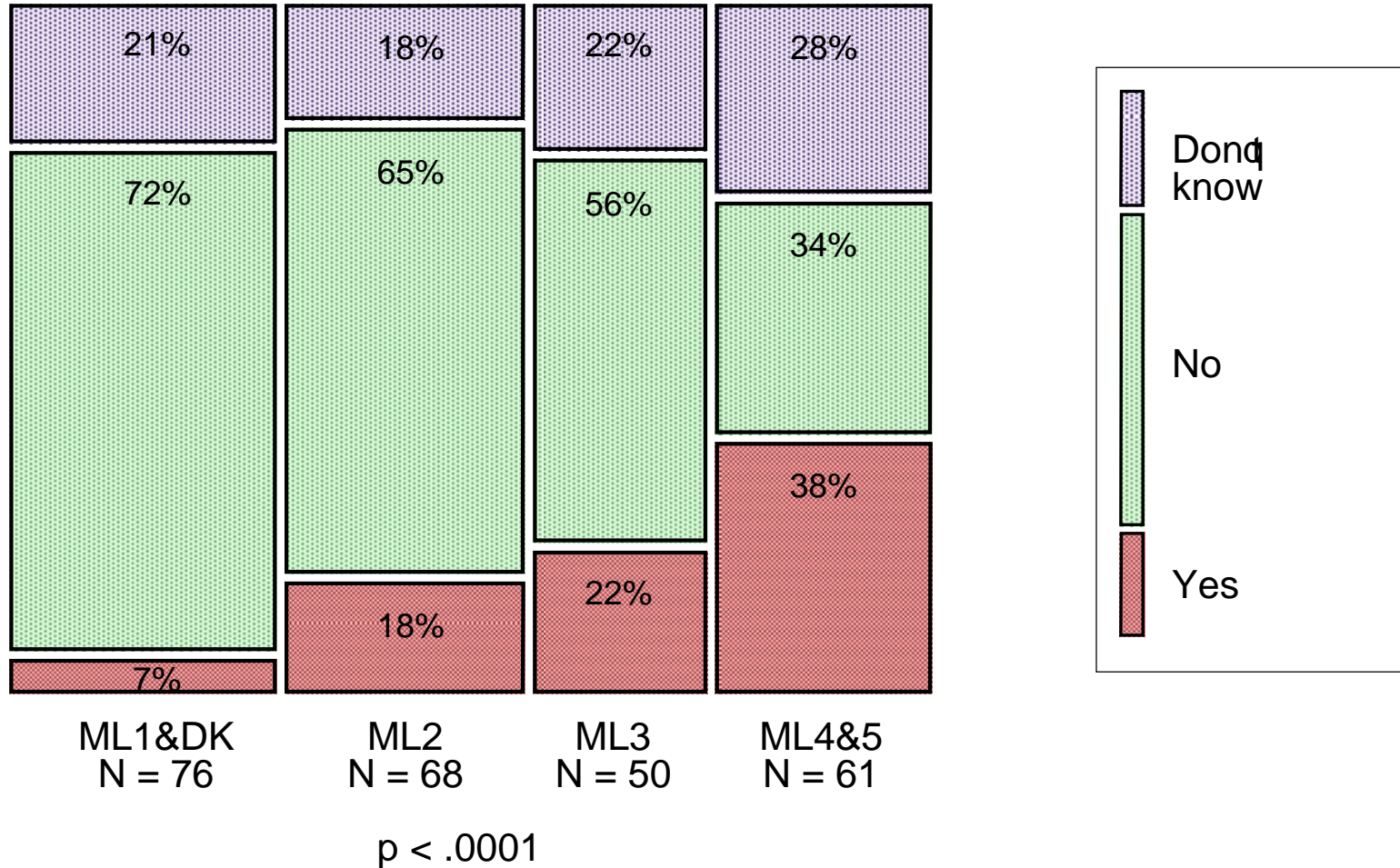- É Organizational perspectives on software measurement
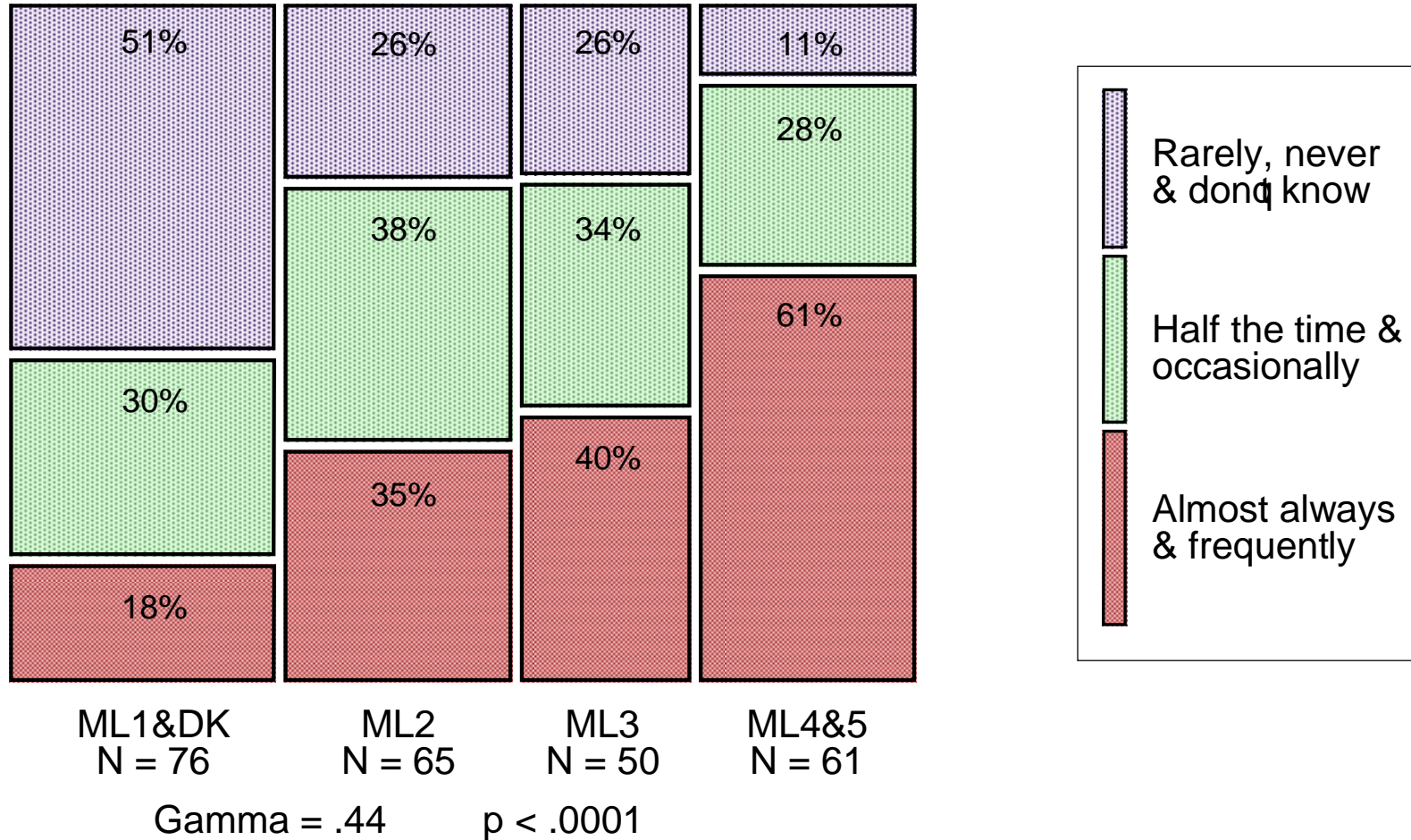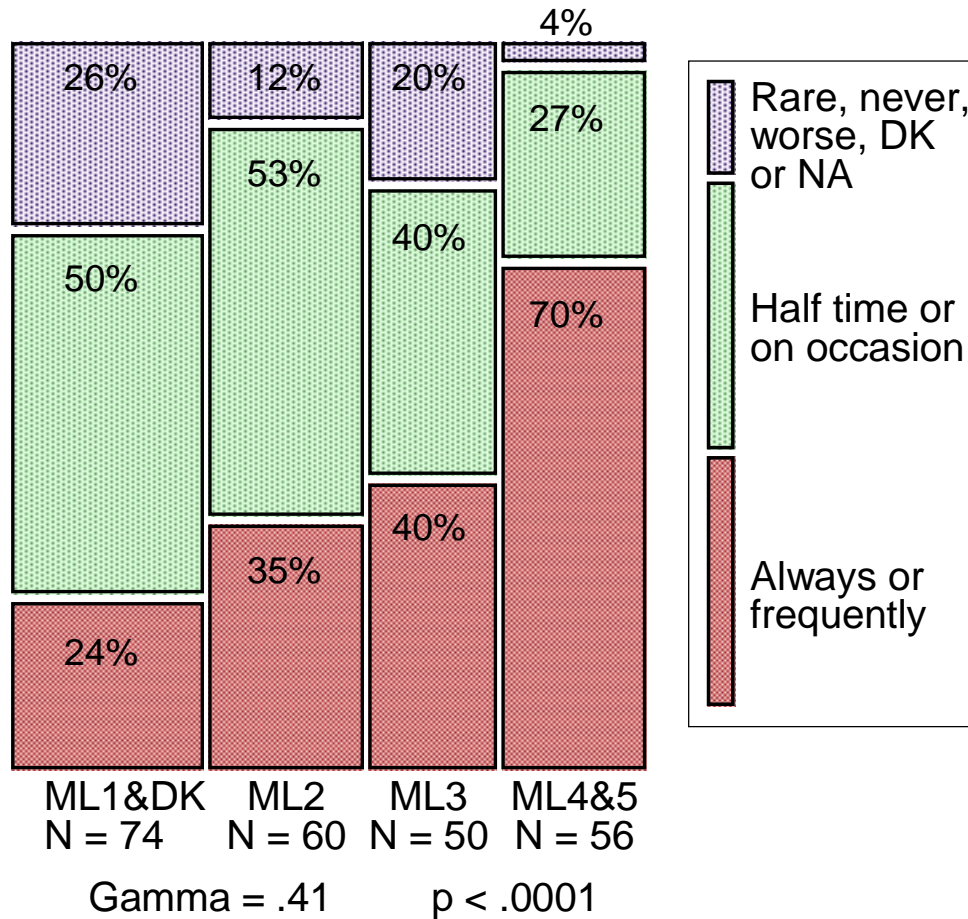
Summary, lessons learned & next steps

...surement on the Organizations[1]

Better Project Performance

Better Product Quality

Legend:
- Rare, never, worse, DK or NA
- Half time or on occasion
- Always or frequently

Better Project Performance chart values:
- ML1&DK, N = 74: 24%, 50%, 26%
- ML2, N = 60: 35%, 53%, 12%
- ML3, N = 50: 40%, 40%, 20%
- ML4&5, N = 56: 70%, 27%, 4%

Gamma = .41   p < .0001

Better Product Quality chart values:
- ML1&DK, N = 74: 26%, 49%, 26%
- ML2, N = 60: 38%, 48%, 13%
- ML3, N = 50: 44%, 34%, 22%
- ML4&5, N = 56: 63%, 30%, 7%

Gamma = .34   p < .0002

...surement on the Organizations₂

Better Tactical Decisions

ML1&DK N = 74	ML2 N = 59	ML3 N = 50	ML4&5 N = 56
27%	20%	26%	9%
57%	58%	36%	38%
16%	22%	38%	54%
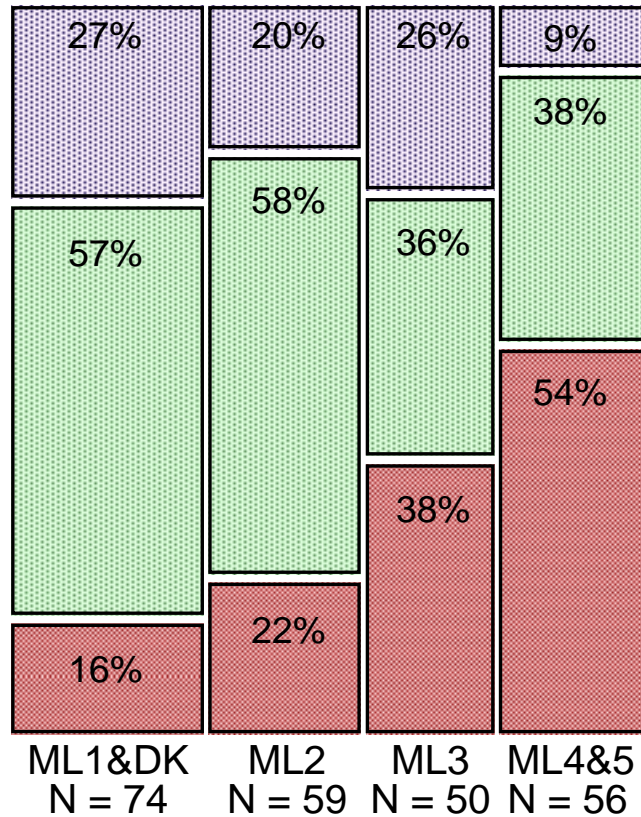
Gamma = .35   p = .0001

Better Strategic Decisions
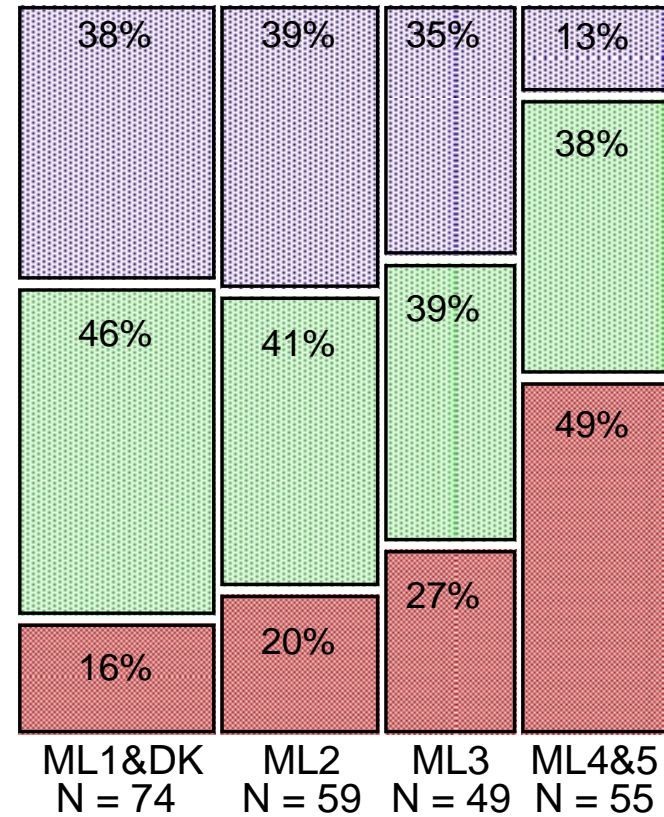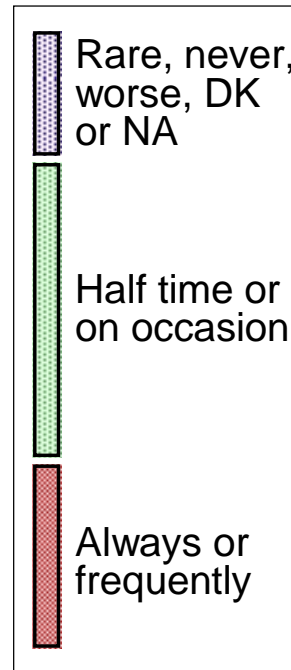
ML1&DK N = 74	ML2 N = 59	ML3 N = 49	ML4&5 N = 55
38%	39%	35%	13%
46%	41%	39%	38%
16%	20%	27%	49%

Gamma = .31   p = .0008

Legend:
- Rare, never, worse, DK or NA
- Half time or on occasion
- Always or frequently

Purpose & scope of the survey

Results

- É The respondents & their organizations
- É Measurement resources & infrastructure
- É Value added by measurement
- É **Software measures used**
- É Data quality & integrity
- É Organizational perspectives on software measurement
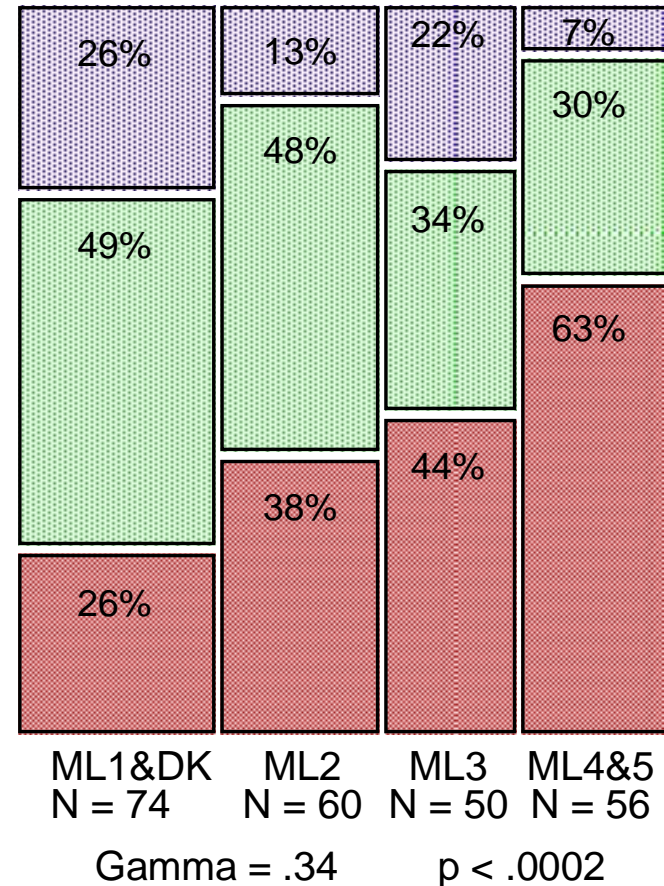
Summary, lessons learned & next steps

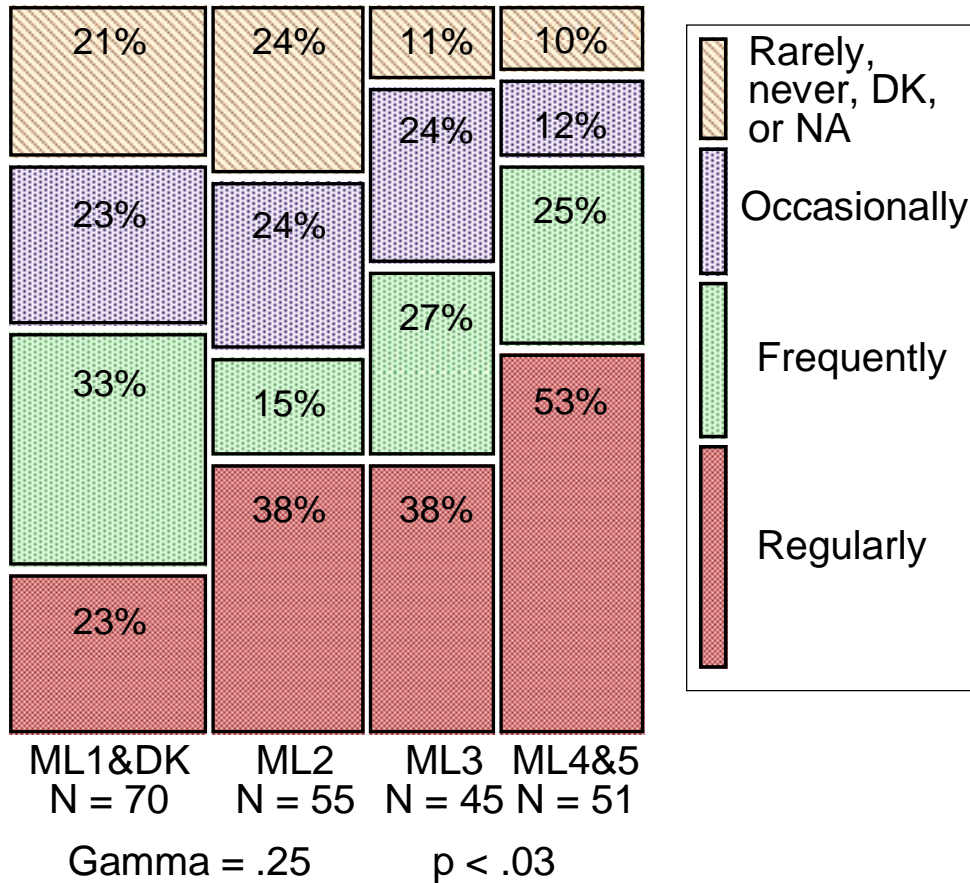# ...nizational Measurement Results Reported[1]

Cost Performance

Schedule Performance

Legend:
- Rarely, never, DK, or NA
- Occasionally
- Frequently
- Regularly

Cost Performance:

	ML1&DK N = 70	ML2 N = 55	ML3 N = 45	ML4&5 N = 51
Rarely/never	21%	24%	11%	10%
Occasionally	23%	24%	24%	12%
Frequently	33%	15%	27%	25%
Regularly	23%	38%	38%	53%

Gamma = .25   p < .03

Schedule Performance:

	ML1&DK N = 70	ML2 N = 56	ML3 N = 44	ML4&5 N = 51
Rarely/never	14%	7%	11%	4%
				2%
Occasionally	19%	11%	16%	33%
Frequently	34%	34%	73%	61%
Regularly	33%	48%		
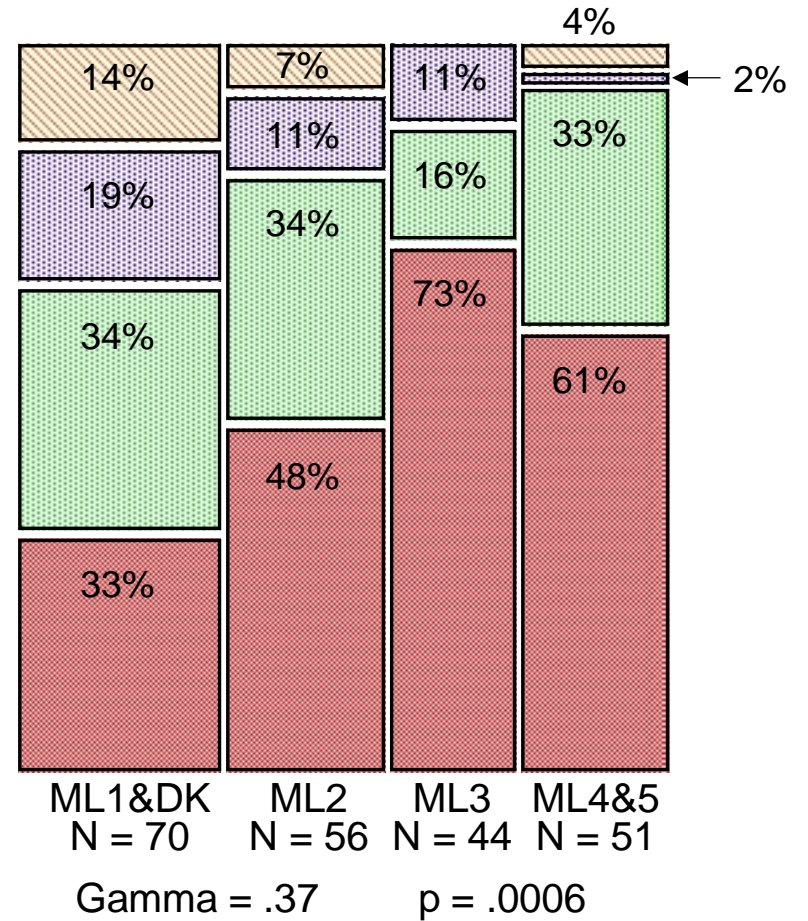
Gamma = .37   p = .0006

nizational Measurement Results Reported₂

Business Growth & Profitability

ML1&DK N = 70	ML2 N = 55	ML3 N = 45	ML4&5 N = 51
40%	33%	33%	20%
21%	31%	20%	20%
23%	15%	22%	29%
16%	22%	24%	31%

Legend:
- Rarely, never, DK, or NA
- Occasionally
- Frequently
- Regularly

Gamma = .20   p = .2244

Understanding CMMI Measurement Capabilities
Dennis R. Goldenson, 14 November 2007

29

© 2007 Carnegie Mellon University
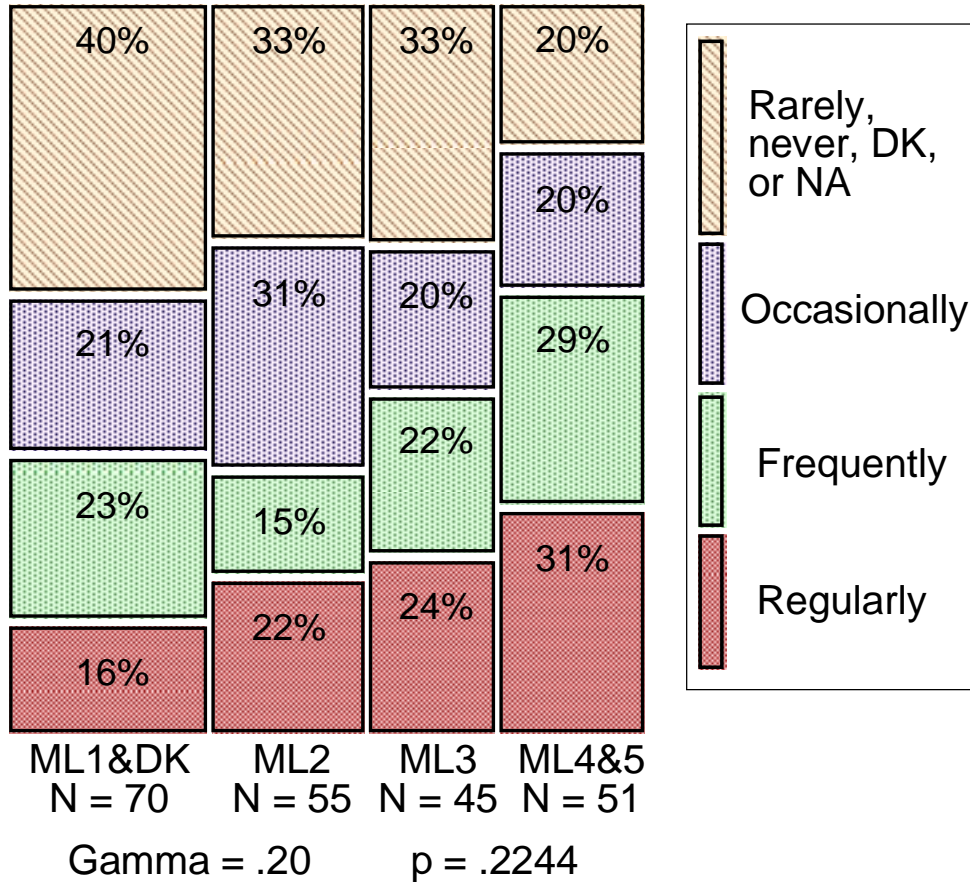
lity Measurement Results Reported[1]

Requirements / Architectures

	ML1&DK N = 70	ML2 N = 55	ML3 N = 45	ML4&5 N = 51
Rarely, never, DK, or NA	24%	18%	13%	10%
Occasionally	37%	15%	24%	8%
Frequently	21%	31%	18%	27%
Regularly	17%	36%	44%	55%
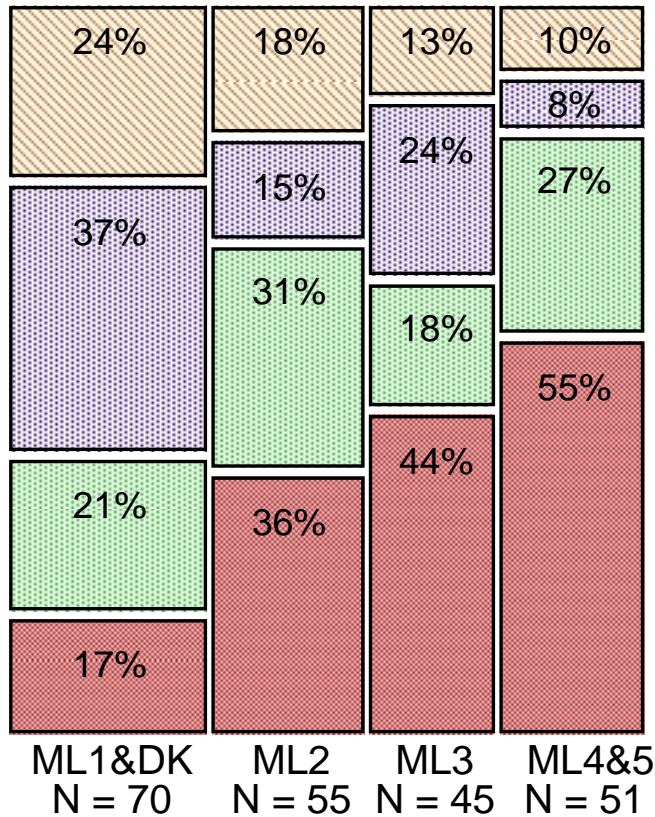
Gamma = .37   p = .0002

Quality Attributes

	ML1&DK N = 70	ML2 N = 55	ML3 N = 45	ML4&5 N = 52
Rarely, never, DK, or NA	57%	40%	42%	25%
Occasionally	21%	31%	24%	21%
Frequently	16%	11%	16%	23%
Regularly	6%	18%	18%	31%

Gamma = .32   p < .008

Legend:
- Rarely, never, DK, or NA
- Occasionally
- Frequently
- Regularly
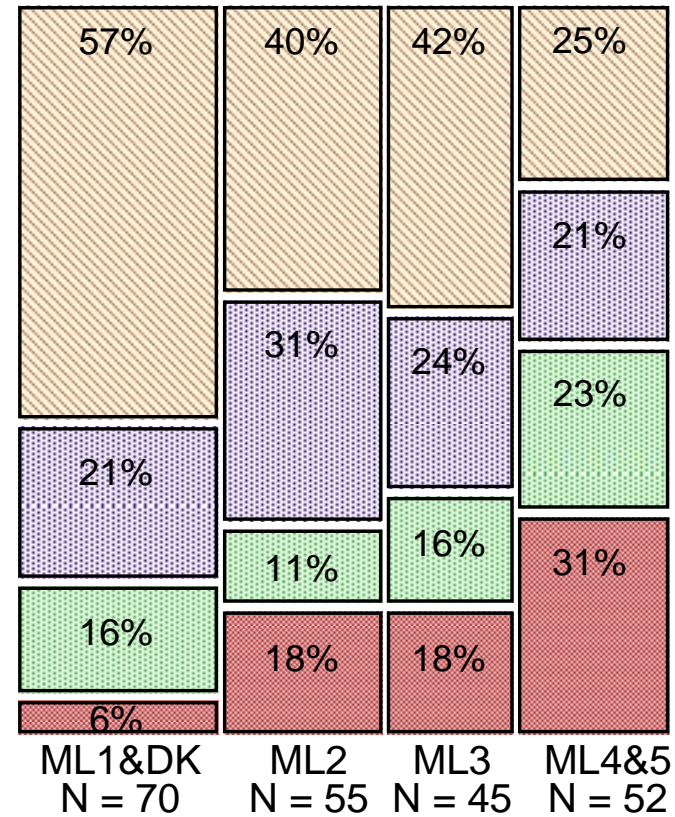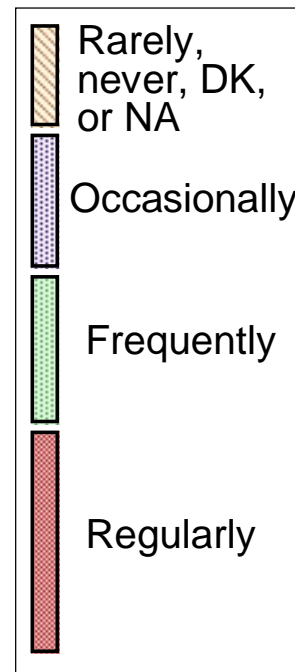
...lity Measurement Results Reported[2]

Defect Density

ML1&DK N = 70	ML2 N = 56	ML3 N = 45	ML4&5 N = 52
30%	13%	11%	4%
19%	31%	16%	6%
31%	23%	22%	33%
20%	34%	51%	58%

Gamma = .41   p < .0001
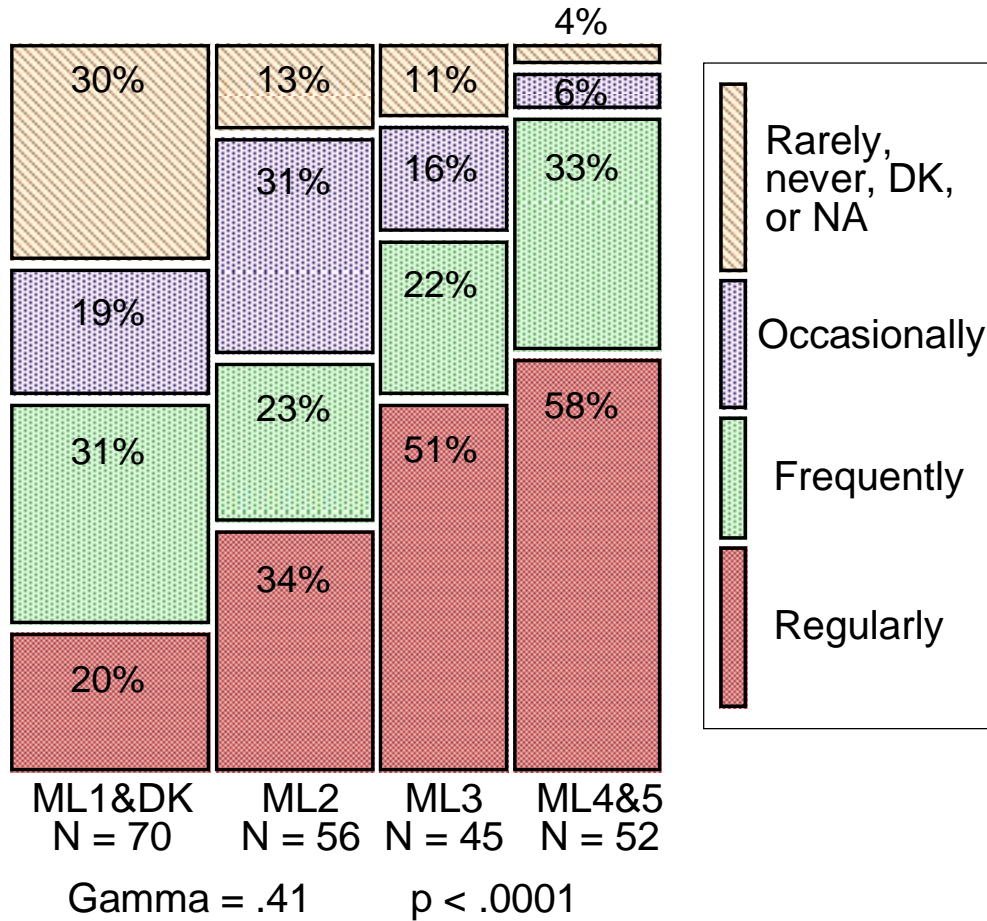
Defect Phase Containment

ML1&DK N = 70	ML2 N = 56	ML3 N = 45	ML4&5 N = 51
50%	30%	27%	8%
23%	29%	20%	14%
17%	13%	27%	29%
10%	29%	27%	49%

Gamma = .44   p < .0001

Legend:
- Rarely, never, DK, or NA
- Occasionally
- Frequently
- Regularly
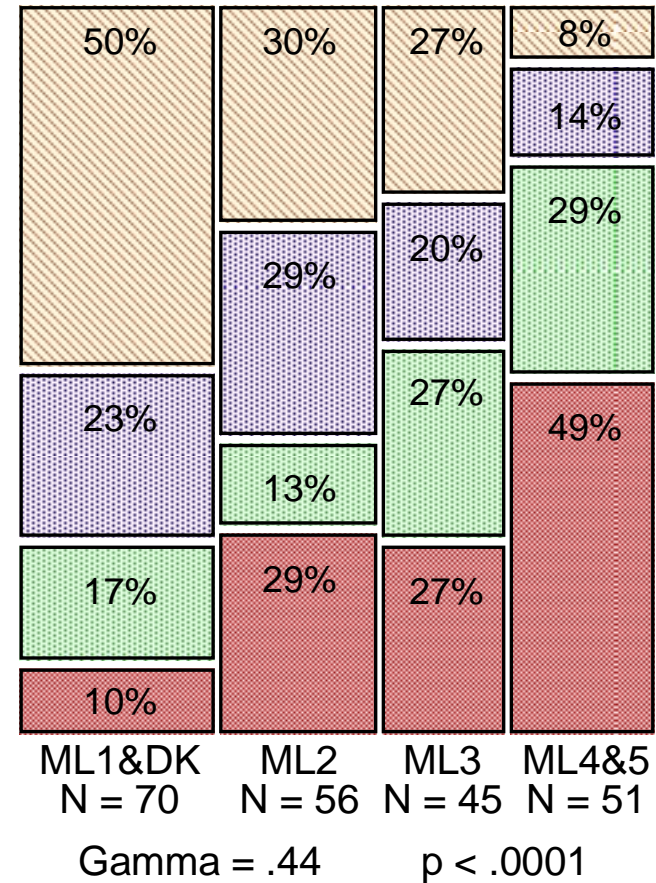
lity Measurement Results Reported[3]

Customer Satisfaction

| | Rarely, never, DK, or NA |
| Occasionally |
| Frequently |
| Regularly |

ML1&DK N = 70   ML2 N = 56   ML3 N = 45   ML4&5 N = 52

Gamma = .31 $p < .005$
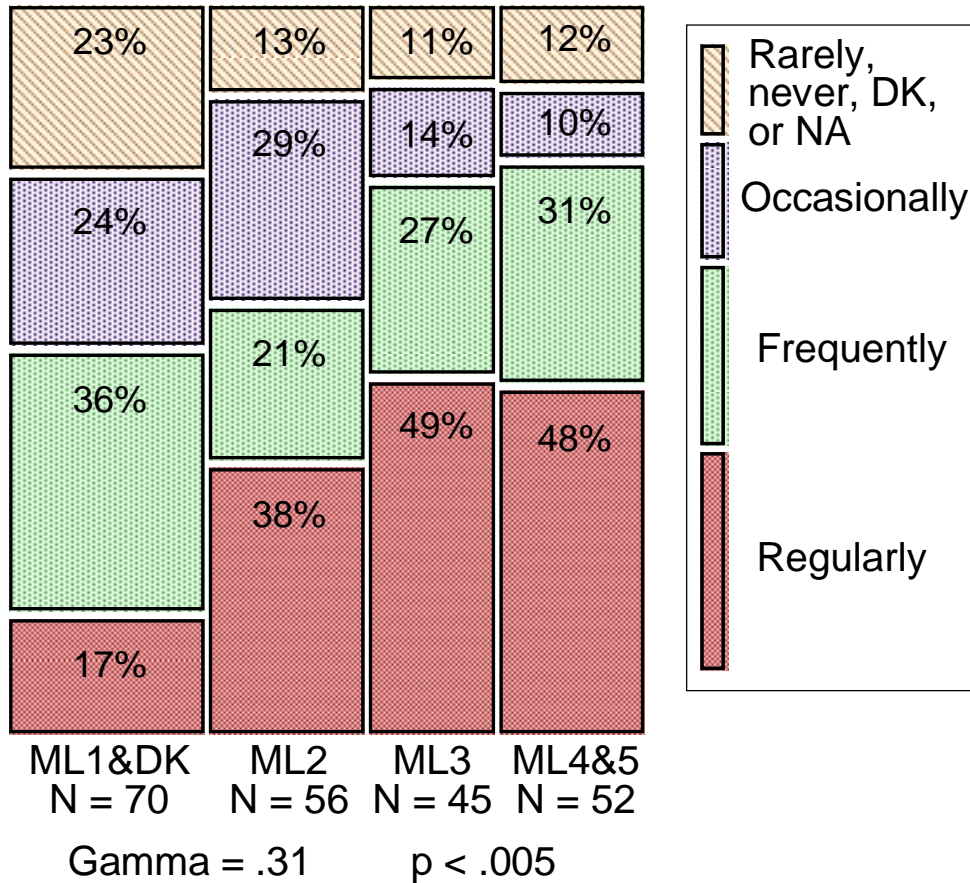
For:

- É Adherence to work processes

- É Effort applied to task

- É Estimation accuracy

- É Cycle time

Proportions sometimes vary across the distributions.

But there are consistent differences by maturity level.
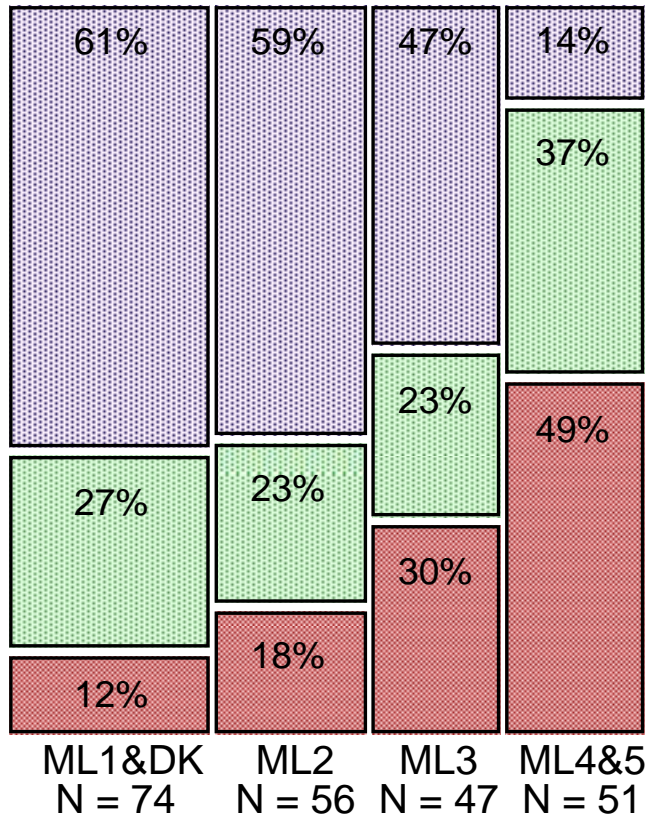
Purpose & scope of the survey

Results

- The respondents & their organizations
- Measurement resources & infrastructure
- Value added by measurement
- Software measures used
- **Data quality & integrity**
- Organizational perspectives on software measurement
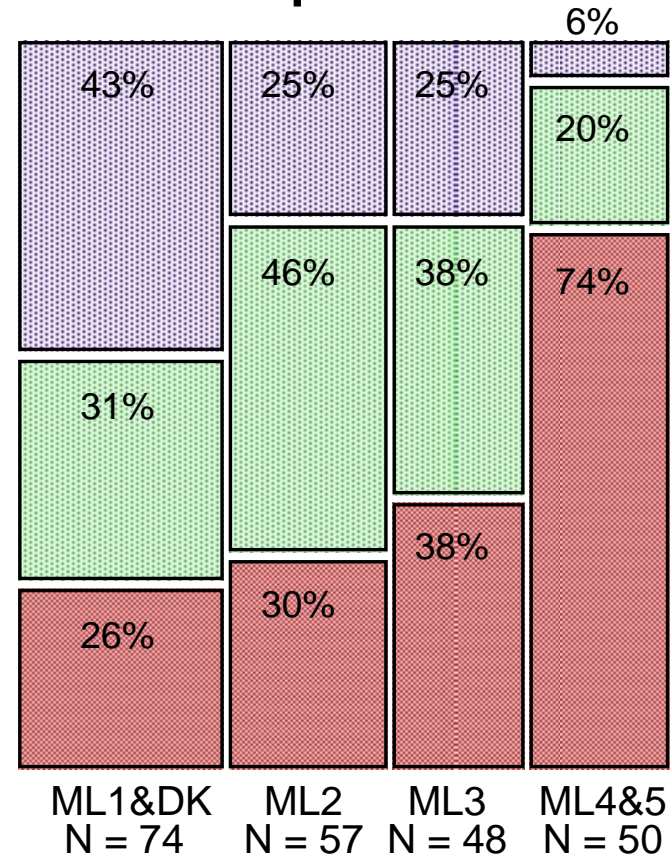
Summary, lessons learned & next steps

Maturity Level: Practices to Ensure Data Quality

Statistical estimates of measurement error
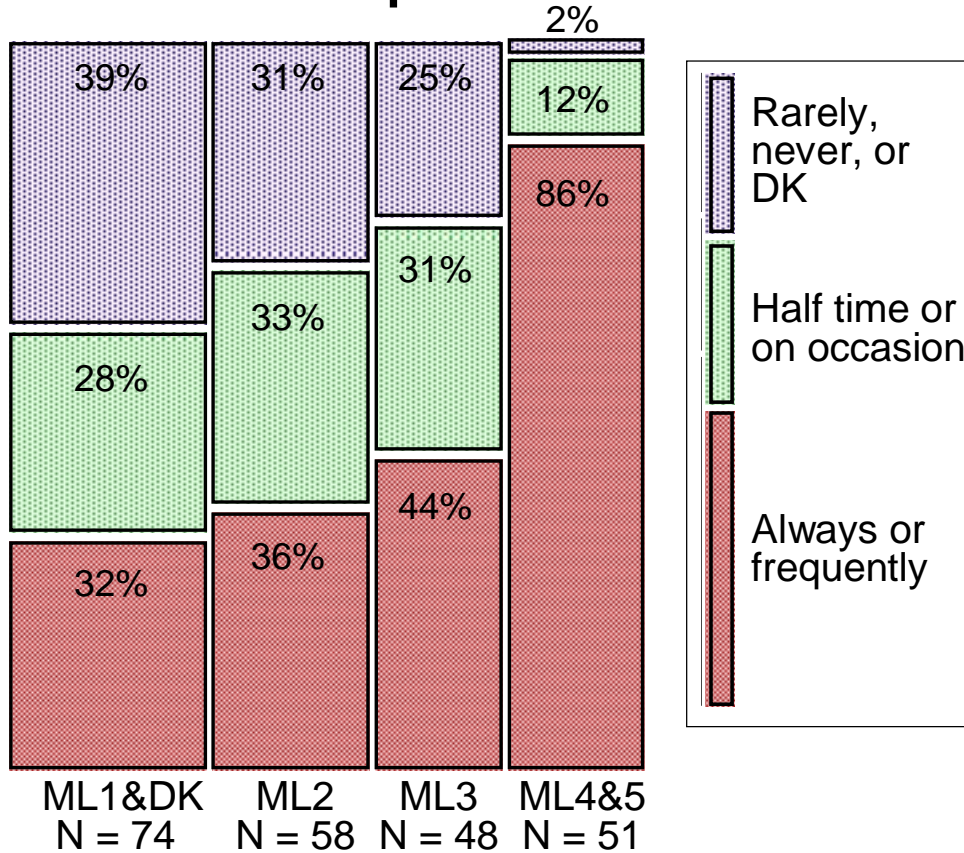

Checks for inconsistent interpretation

Legend:
- Rarely, never, or DK
- Half time or on occasion
- Always or frequently

Statistical estimates of measurement error

	ML1&DK N = 74	ML2 N = 56	ML3 N = 47	ML4&5 N = 51
Rarely, never, or DK	61%	59%	47%	14%
Half time or on occasion	27%	23%	23%	37%
Always or frequently	12%	18%	30%	49%

Gamma = .44   p < .0001

Checks for inconsistent interpretation

	ML1&DK N = 74	ML2 N = 57	ML3 N = 48	ML4&5 N = 50
Rarely, never, or DK	43%	25%	25%	6%
Half time or on occasion	31%	46%	38%	20%
Always or frequently	26%	30%	38%	74%

Gamma = .44   p < .0001

Maturity Level:
Practices to Ensure Data Quality

Checks for unusual distribution patterns

| ML1&DK N = 74 | ML2 N = 58 | ML3 N = 48 | ML4&5 N = 51 |

Legend:
- Rarely, never, or DK
- Half time or on occasion
- Always or frequently

Gamma = .46   p < .0001

For:

- Out of range & illegal values ... Number & distribution of missing data
- Missing data not treated as zero ... Precision & accuracy tests
- Other aspects of alignment & coordination of measurement activities
  - Understandable & consistent measurement definitions
  - Understandable & interpretable measurement results
  - Use of %standard+measurement methods
  - Measurable product & service criteria
  - Measurement used to understand product & service quality
  - Documented data collection process
  - Documented process for reporting results
  - Corrective action taken when thresholds exceeded
  - Understands purposes of the data collected/reported

Proportions sometimes vary across the distributions.

But there are consistent differences by maturity level.
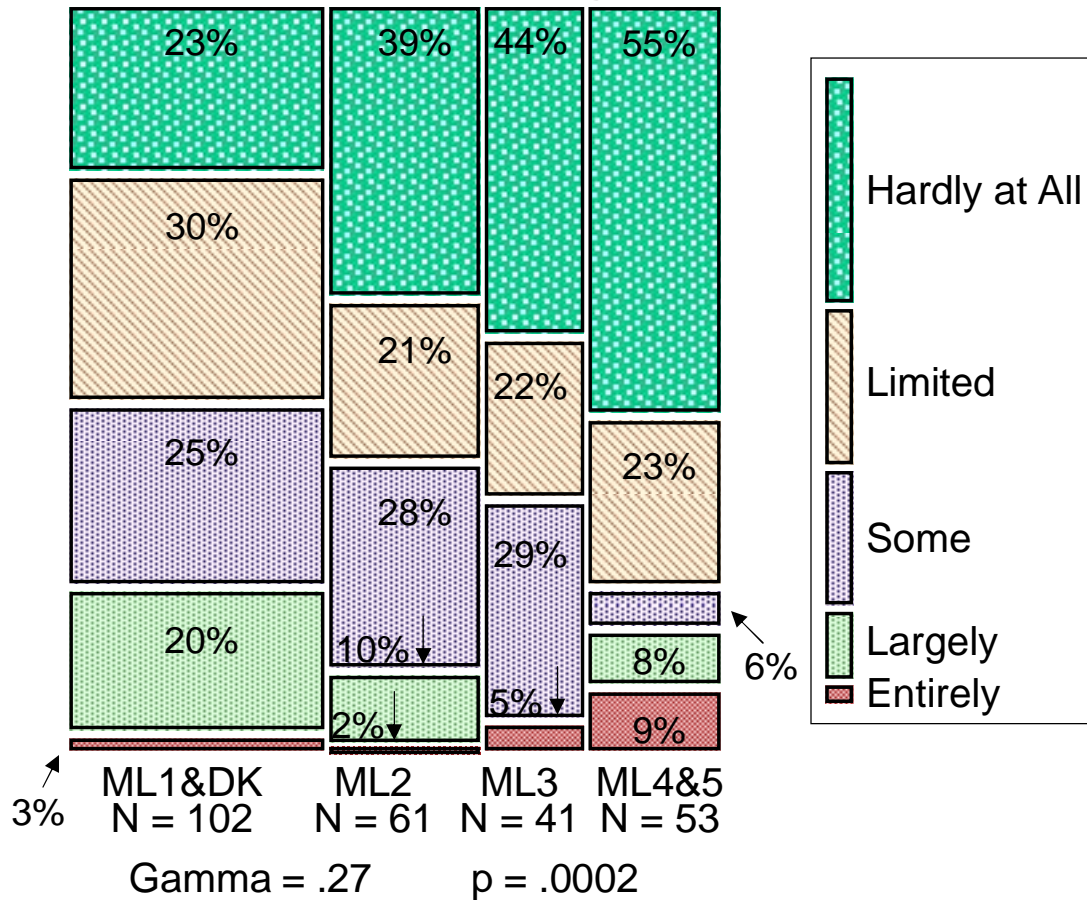
Purpose & scope of the survey

Results

- É The respondents & their organizations
- É Measurement resources & infrastructure
- É Value added by measurement
- É Software measures used
- É Data quality & integrity
- É **Organizational perspectives on software measurement**
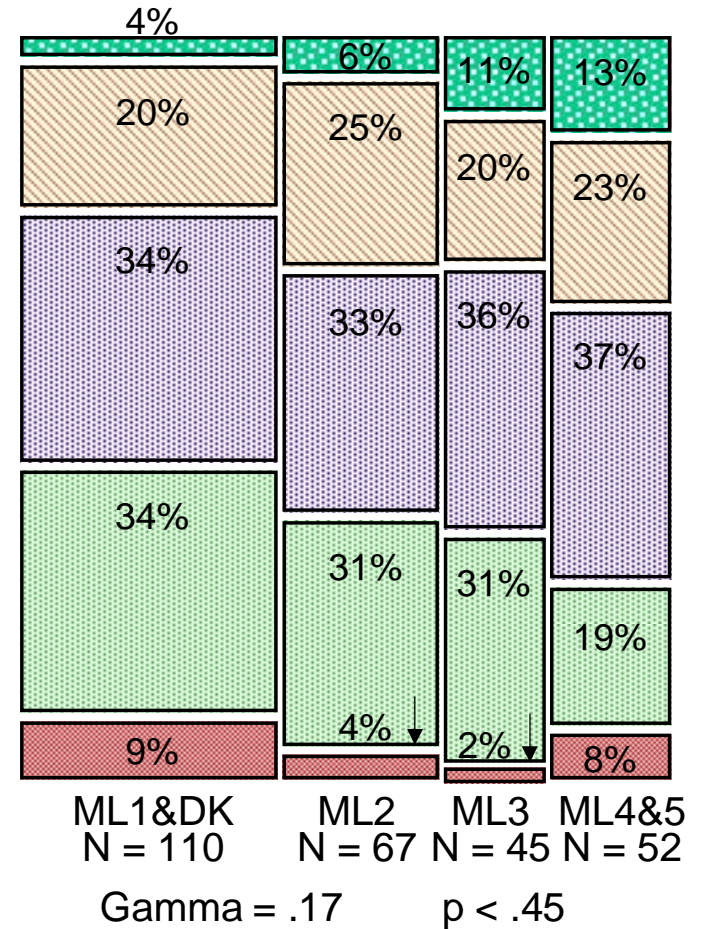
Summary, lessons learned & next steps

Perspectives

Not Relevant for Decision Making

	ML1&DK N = 102	ML2 N = 61	ML3 N = 41	ML4&5 N = 53
Hardly at All	23%	39%	44%	55%
Limited	30%	21%	22%	23%
Some	25%	28%	29%	6%
Largely	20%	10%	5%	8%
Entirely	3%	2%		9%

Gamma = .27   p = .0002

Onerous or Burdensome

	ML1&DK N = 110	ML2 N = 67	ML3 N = 45	ML4&5 N = 52
Hardly at All	4%	6%	11%	13%
Limited	20%	25%	20%	23%
Some	34%	33%	36%	37%
Largely	34%	31%	31%	19%
Entirely	9%	4%	2%	8%

Gamma = .17   p < .45

Legend:
- Hardly at All
- Limited
- Some
- Largely
- Entirely

For:

- É Stated negatively
    - ô Inappropriate collection & use of data
    - ô Resistance to %extra+work
- É Stated positively
    - ô Understandable & interpretable results
    - ô Data collected are regularly analyzed
    - ô Measurement an integral part of the business
    - ô Objective results highly valued

Once again:

- É Proportions sometimes vary across the distributions.
- É But there are consistent differences by maturity level.

Yet resistance to measurement still exists in our field.

- É Even in high maturity organizations

Purpose & scope of the survey

Results

- É The respondents & their organizations

- É Measurement resources & infrastructure

- É Value added by measurement

- É Software measures used

- É Data quality & integrity

- É Organizational perspectives on software measurement

Summary, lessons learned & next steps

esults

Characteristic differences associated with CMMI Maturity level achieved

- É Measurement capability & performance outcomes

- É Common stair step pattern up the maturity levels

- É Some quite substantial

Still, some of the results imply room for improvement

- É Sometimes substantial room

Even in higher maturity organizations

- É Although the expectations for quality & %goodness+may well be higher there too

- É Jim Herbsleb & I saw a similar pattern years ago

  - ô For process champions *versus* practitioners & managers

es

Lower than desired response rate

- Lower maturity level respondents less likely to finish the questionnaire
- Some drop off in higher maturity level respondents later in questionnaire

Not surprising in a relatively long questionnaire ... but exacerbated by:

- Spoofed email invitations & reminder message errors
- Related problems with incremental saving
  - Cookie flushing & assignment of multiple URLs by COTS web survey product
  - Leading to %lost+information
- & (possibly) lack of feedback on time/length remaining

Recurring anomalous dip at maturity level 3

- May be due to bias from relatively small number of ML3 respondents
- Or learning curve effects ... or higher expectations

ssues

There *always* is noise in survey (**&** other measurement) data, e.g.

- Differing interpretations of intended meaning of questions
- Use of "vague quantifiers" in closed ended response categories

"Dong know" & other off scale responses

- Most common at lower maturity levels
- But they also exist at the higher maturity levels
- Perhaps because some folks in larger organizations truly dong know

Regardless, the survey results are consistent with expectations based on CMMI

- a.k.a. predictive validity

Relatively little data yet exist for meaningful comparisons among software & systems engineering projects & organizations

- É Hence tendency to cover too much at once in a single sample survey

Considering variants on matrix sampling strategies for 2008 survey

- É Answer only a subset of questions ... to avoid over-burdening the respondents

%State of the practice+can refer to very different target populations

- É The SEI customer base ... the broader software & systems engineering community ... or those organizations that more routinely use measurement?

- É Of course, the answer depends on the purposes of the survey

Our plans

- É We will track change over time **&** go into further depth about focused topics from the perspective of current measurement practitioners

Considering parallel samples for 2008

- É A short set of questions for tracking the diffusion of measurement through the broader software & systems engineering community
- É Possible focus on issues faced with respect to the adoption & use of high maturity measurement practices

Also fielding a survey on Program Office acquisition capabilities (early 2008)

Of course, there is no shortage of additional topics for the future

- É In the SEI series or in those that we hope to see done by others

k You for Your Attention!

Dennis R. Goldenson
dg@sei.cmu.edu

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213-3890
USA

Software Engineering Institute | **Carnegie Mellon**