

# ECAR Study on Cyberinfrastructure

Mark Sheehan  
ECAR Fellow

CLAC 2008  
Union College  
Schenectady, NY  
June 19, 2008

© 2008 EDUCAUSE. All rights reserved.



# ECAR Research Study

## *Higher Education IT and Cyberinfrastructure: Integrating Technologies for Scholarship*

- Released June 11, 2008 (PDF)
- Available to entire community from outset
- Print version mailed to ECAR members  
~June 26

# Objectives

- Determine “state of the practice” regarding cyberinfrastructure technologies
  - Usage
  - Sourcing/Funding
  - Importance
  - Knowledge about them
  - Collaborative practices

# Objectives

- Explore state of integration of CI resources and services
  - Central/Distributed
  - “Cottage Industry” vs. Enterprise
- Identify practices associated with effective integration
  - Caveats:
    - Evaluations are self-assessed
    - Causality is difficult to establish

# Respondent Demographics

- 1,688 invitations mailed
- 369 qualified respondents
- 22% response rate
- 27% doctorals
- 77% senior-most IT leaders (CIOs)

# Cyberinfrastructure Technologies

- High-performance computing
- Cyberinfrastructure applications and tools
- Data storage and management
- Advanced network infrastructure
- Resources for collaboration within virtual communities

# Cyberinfrastructure Technologies

## High-Performance Computing

- Supercomputers
- Clusters of computers with supercomputer-like performance

# Cyberinfrastructure Technologies

## CI Applications and Tools

- Software for:
  - Simulation
  - Parallelization
  - Visualization
  - Job scheduling
  - Data mining
  - Statistical analysis, etc.
- But not for:
  - Gene sequencing
  - Chemical analysis
  - Other disciplinary applications

# Cyberinfrastructure Technologies

## Data Storage and Management

- Large-scale research data storage systems
  - Real-time use
  - Archival purposes
- Facilities, software, and procedures for periodic backup of research data sets

# Cyberinfrastructure Technologies

## Advanced Network Infrastructure

- High-performance networks on campus
- Connections to off-campus high-performance networks that support such capabilities as
  - Massive data transfers to and from clusters
  - Real-time visualization
  - Use of remote instrumentation
  - Regional or university consortial networks
  - Internet2 and National LambdaRail

# Cyberinfrastructure Technologies

## Resources for Collaboration within Virtual Communities

- Facilities and support for:
  - Teleconferencing
  - Hosting collaborations with off-campus researchers
  - Operation of remotely located research instrumentation and related devices
- Support for identity management and associated middleware in collaborative research activities

# Research vs. Teaching Mission

- Research and teaching are the primary missions, but research is what really drives faculty and institutional success. (CLAC=0)
- Research and teaching are both primary missions, and they are equally important for faculty and institutional success. (CLAC=8)
- Teaching is the primary mission, but faculty research is rewarded. (CLAC=8)
- Teaching is the primary mission, and faculty research does not factor heavily in faculty and institutional success. (CLAC=2)

# Research vs. Teaching Mission

## RESEARCH (CLAC=8)

- Research and teaching are the primary missions, but research is what really drives faculty and institutional success.
- Research and teaching are both primary missions, and they are equally important for faculty and institutional success.

## TEACHING (CLAC=10)

- Teaching is the primary mission, but faculty research is rewarded.
- Teaching is the primary mission, and faculty research does not factor heavily in faculty and institutional success.

# Principal Outcome Variable: Central IT's Effectiveness at Integrating CI Resources

## Survey Question

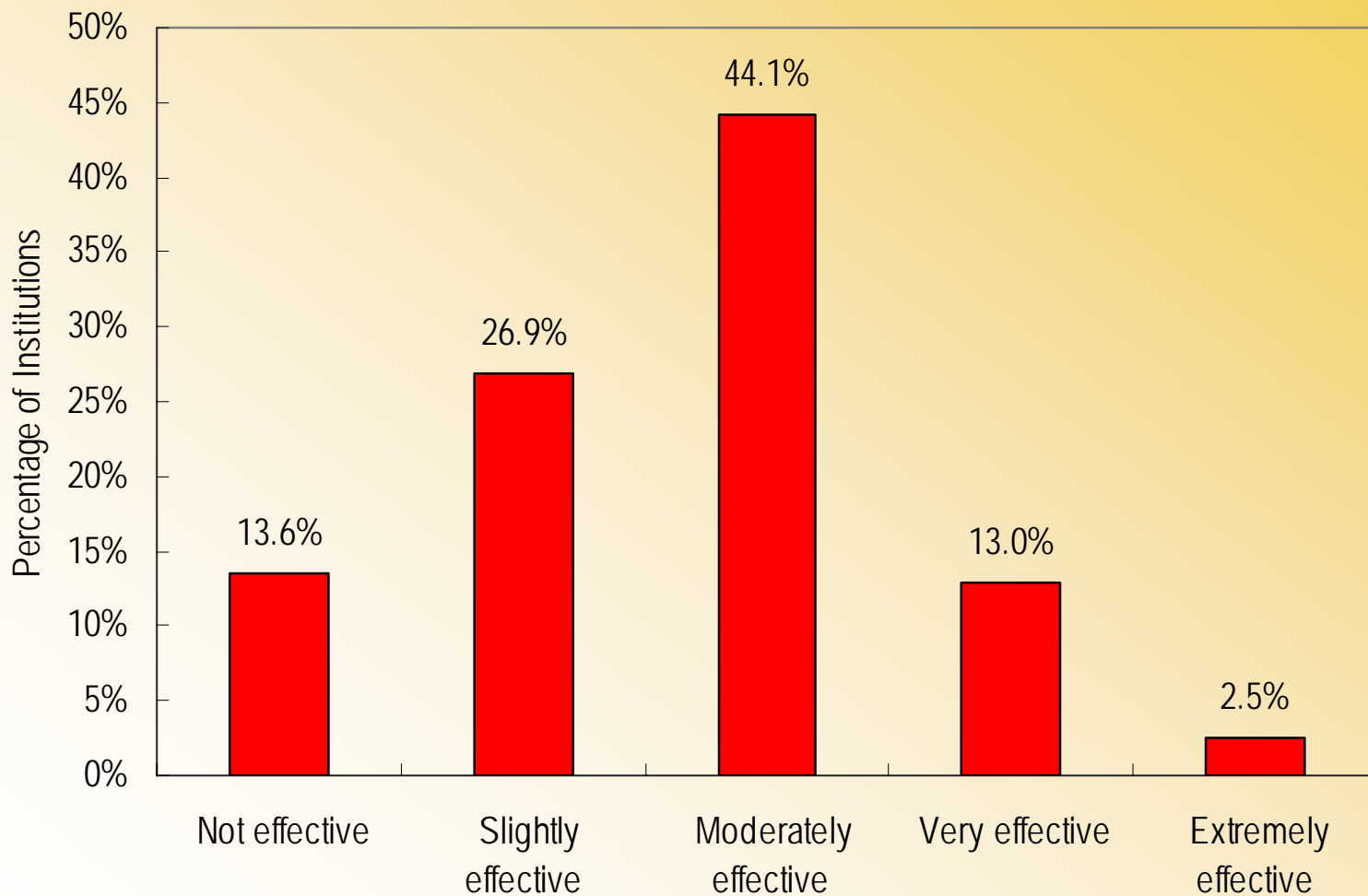
*Rate your central IT organization's effectiveness at integrating the resources we have explored in this survey to provide seamless support for research.*

## Annotation

*“Integrating” means bringing together into a seamless whole a wide variety of human, software, and hardware systems to form a platform for enabling activities in research and in teaching and learning. It involves coordination, synthesis, and teamwork.*

— After Fran Berman, 2004

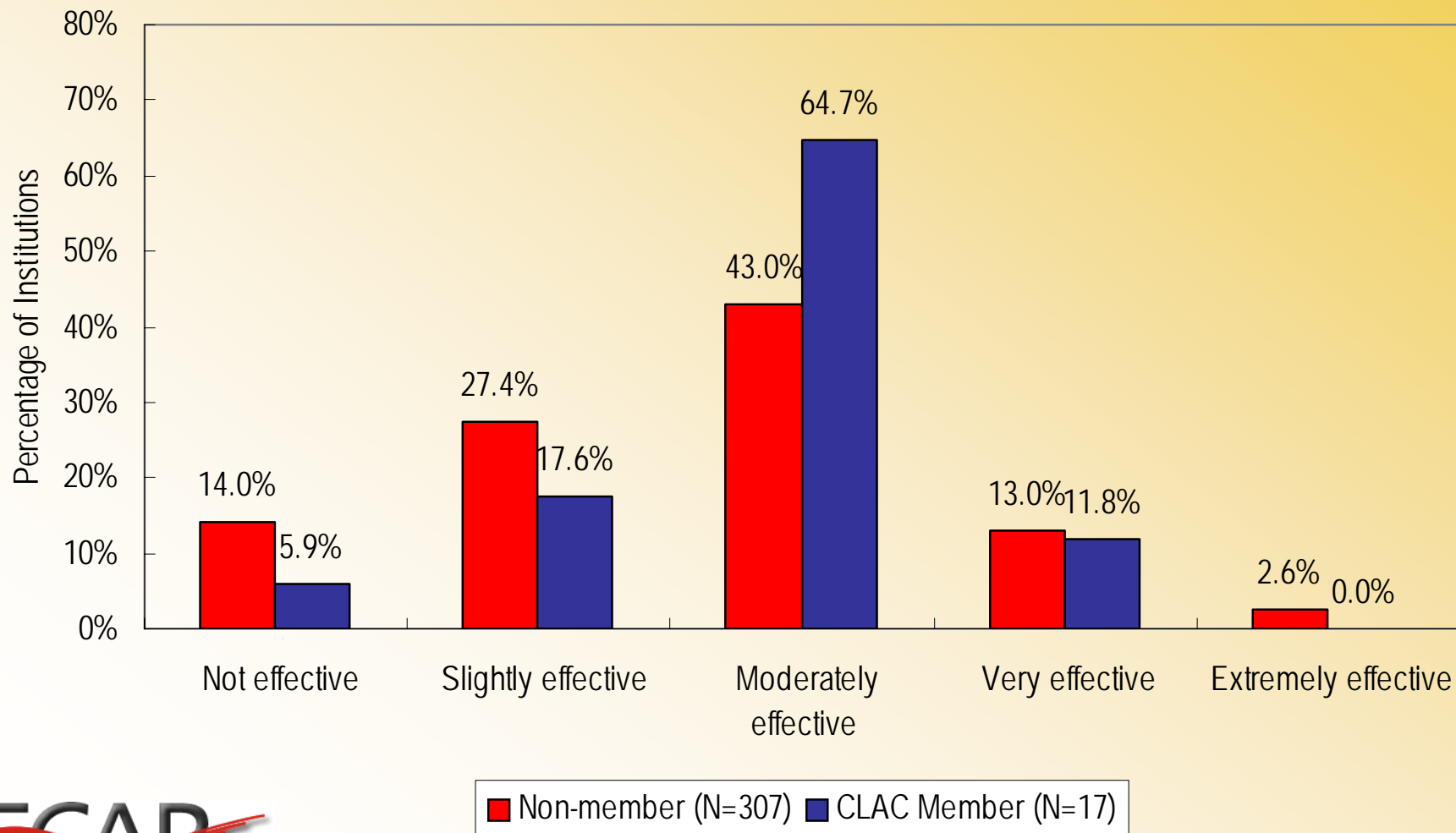
# Effectiveness at Integrating CI Resources



# DISCLAIMER

Several of the following slides deal with responses from CLAC member institutions. In virtually all cases, the data presented lack statistical significance. In the few cases where the findings are statistically significant, the significance is not robust because of the very low numbers of CLAC-member respondents to the survey. ECAR does not consider such findings “reportable.” Be advised....

# Effectiveness at Integrating CI Resources

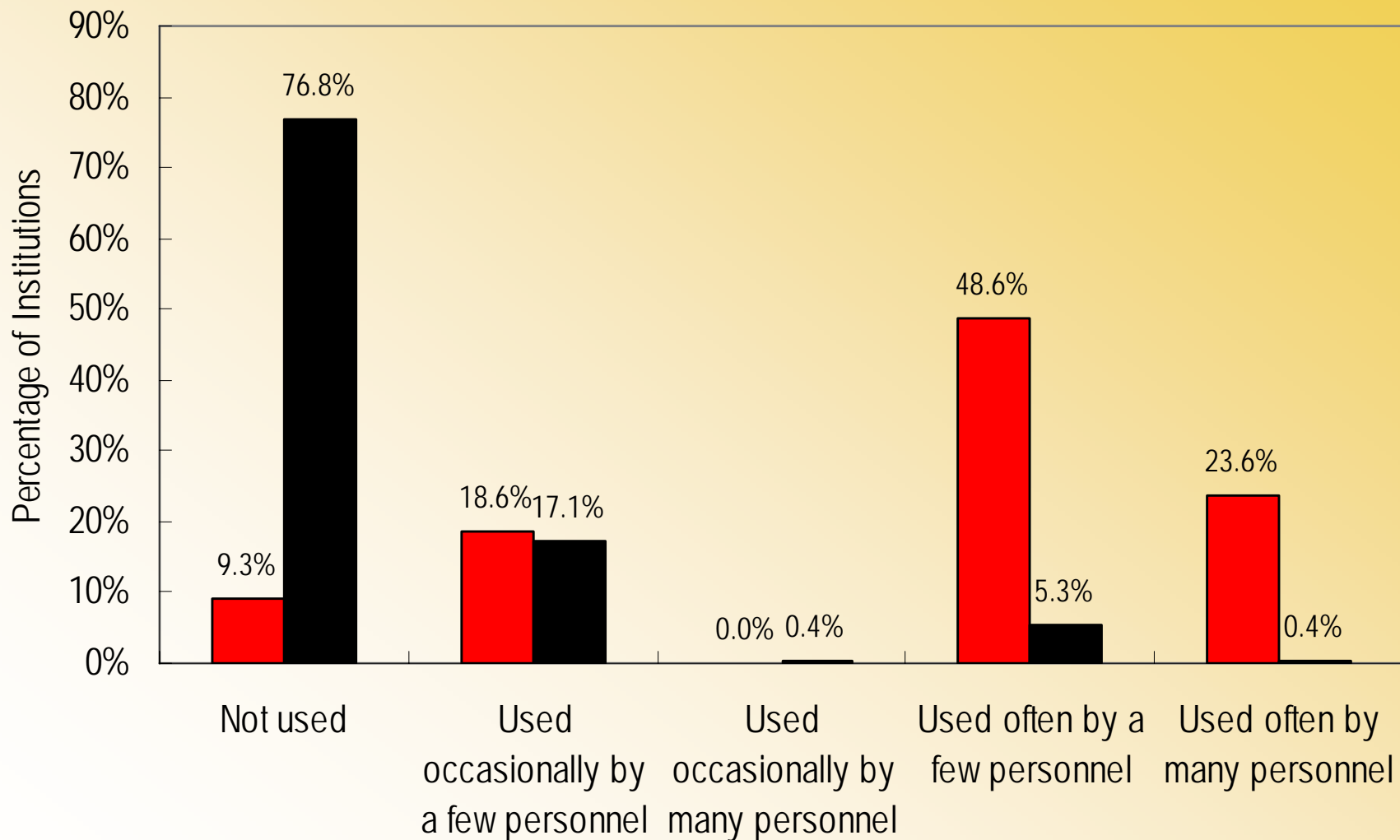


# Usage

- Used occasionally by a few personnel
- Used occasionally by many personnel
- Used often by a few personnel
- Used often by many personnel

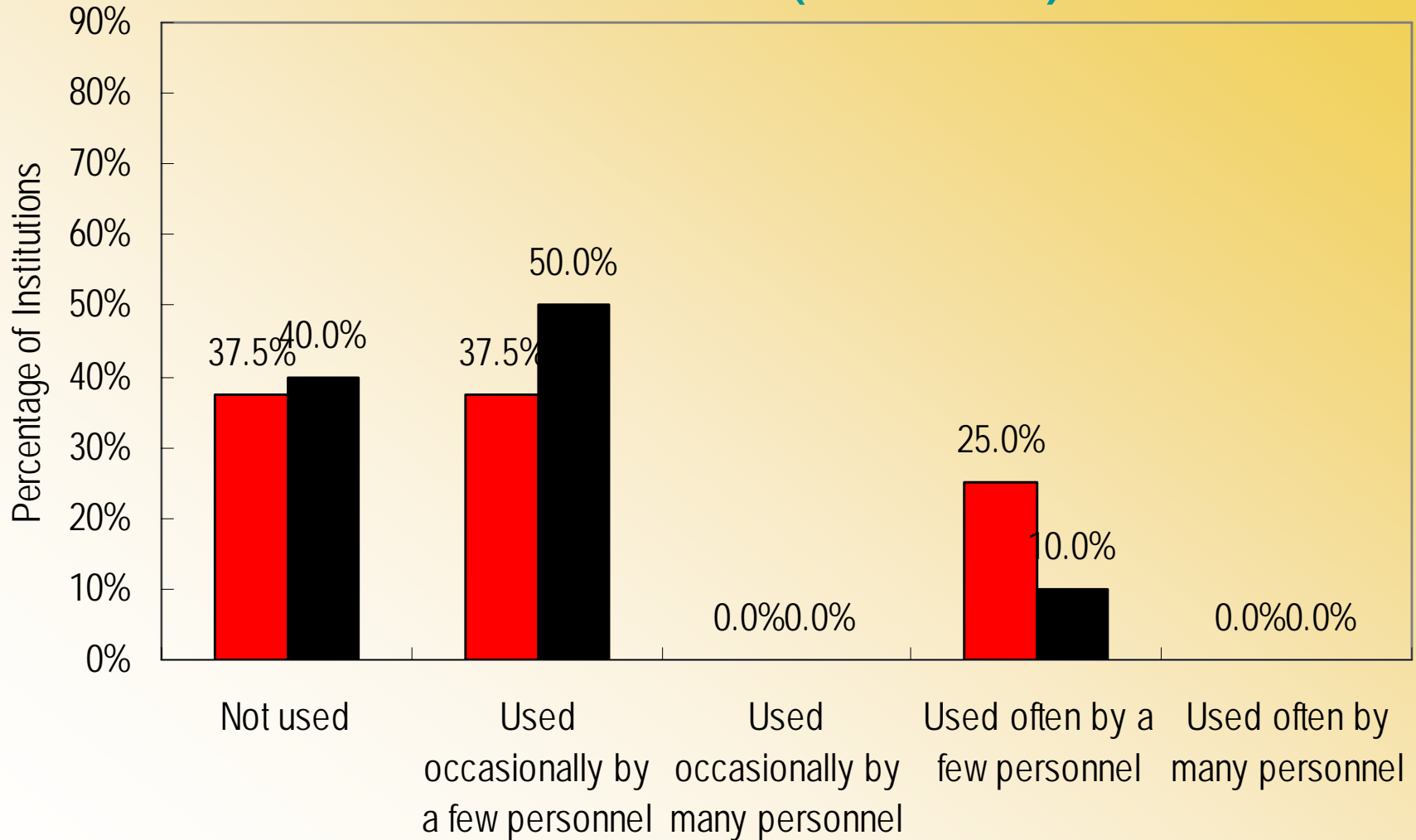


# High-Performance Computing Resources

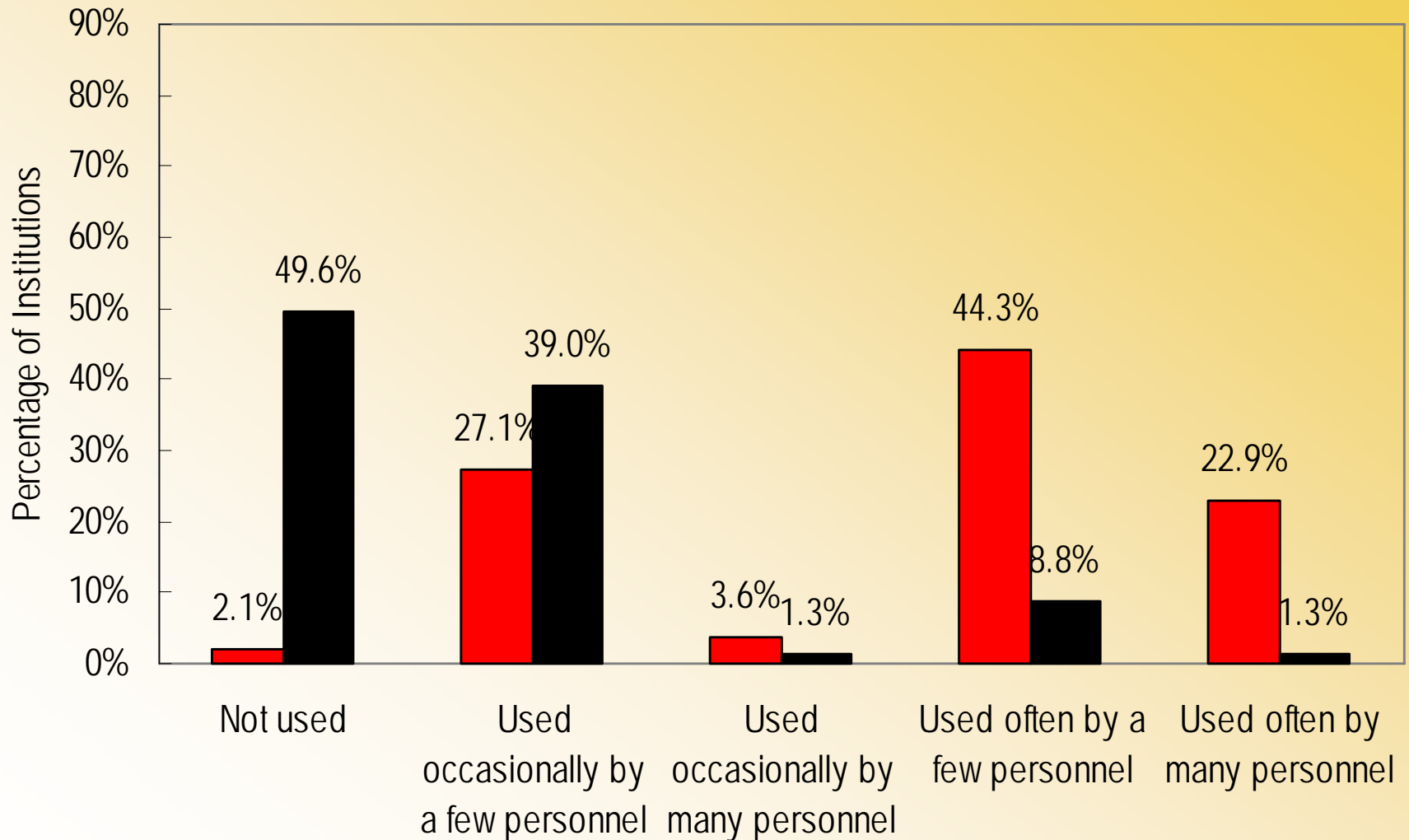


■ Research mission (N=140) ■ Teaching mission (N=228)

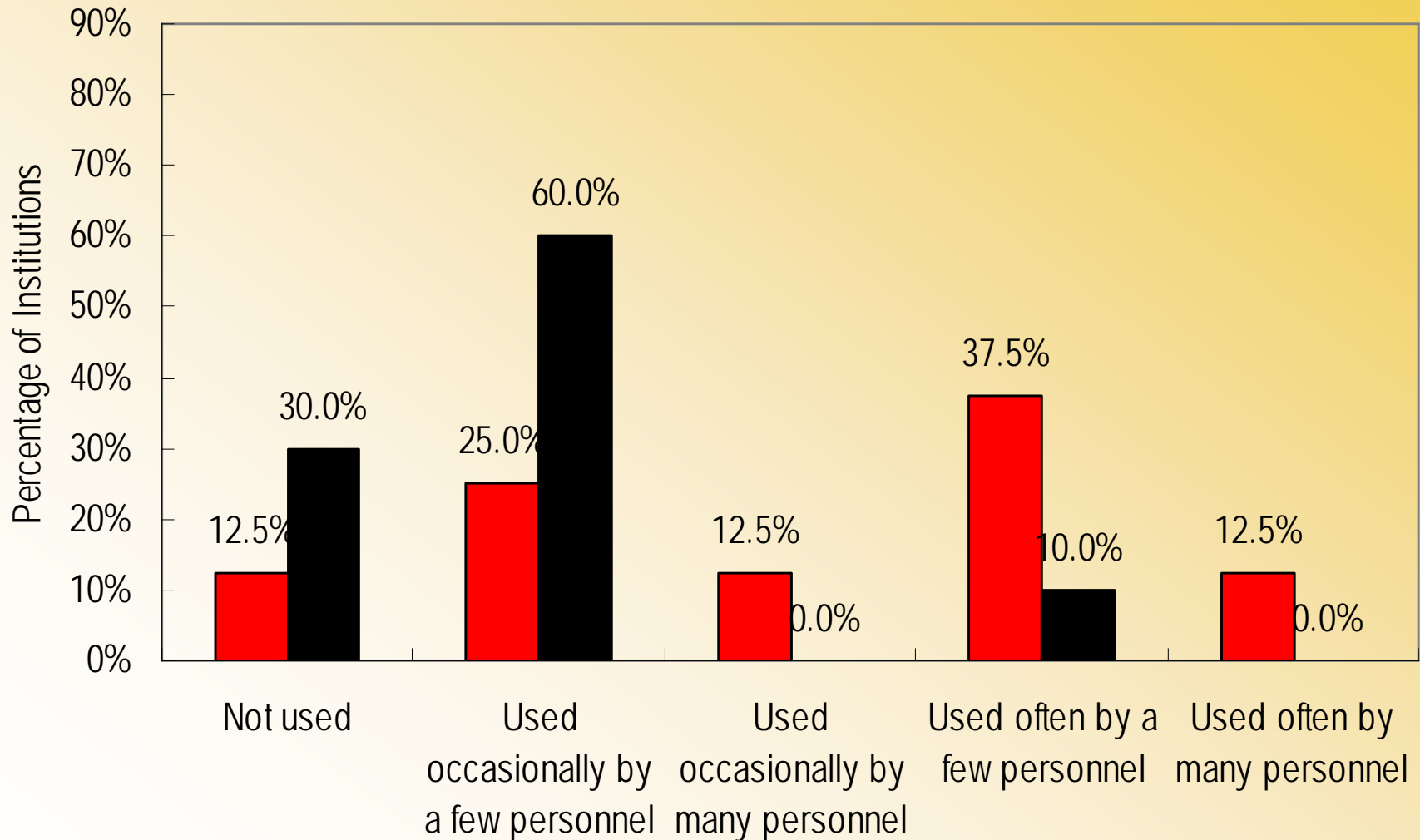
# High-Performance Computing Resources (CLAC)



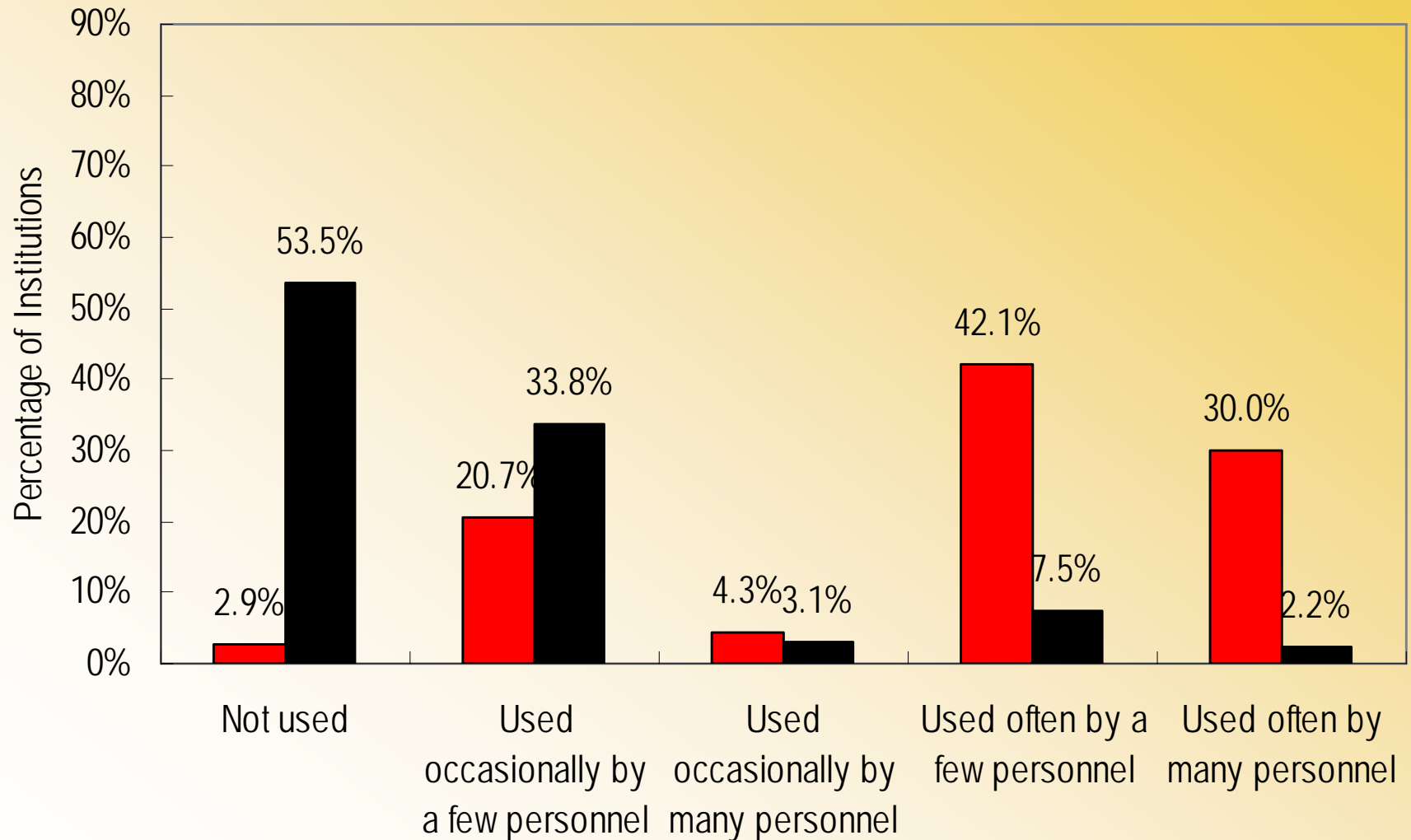
# CI Applications and Tools



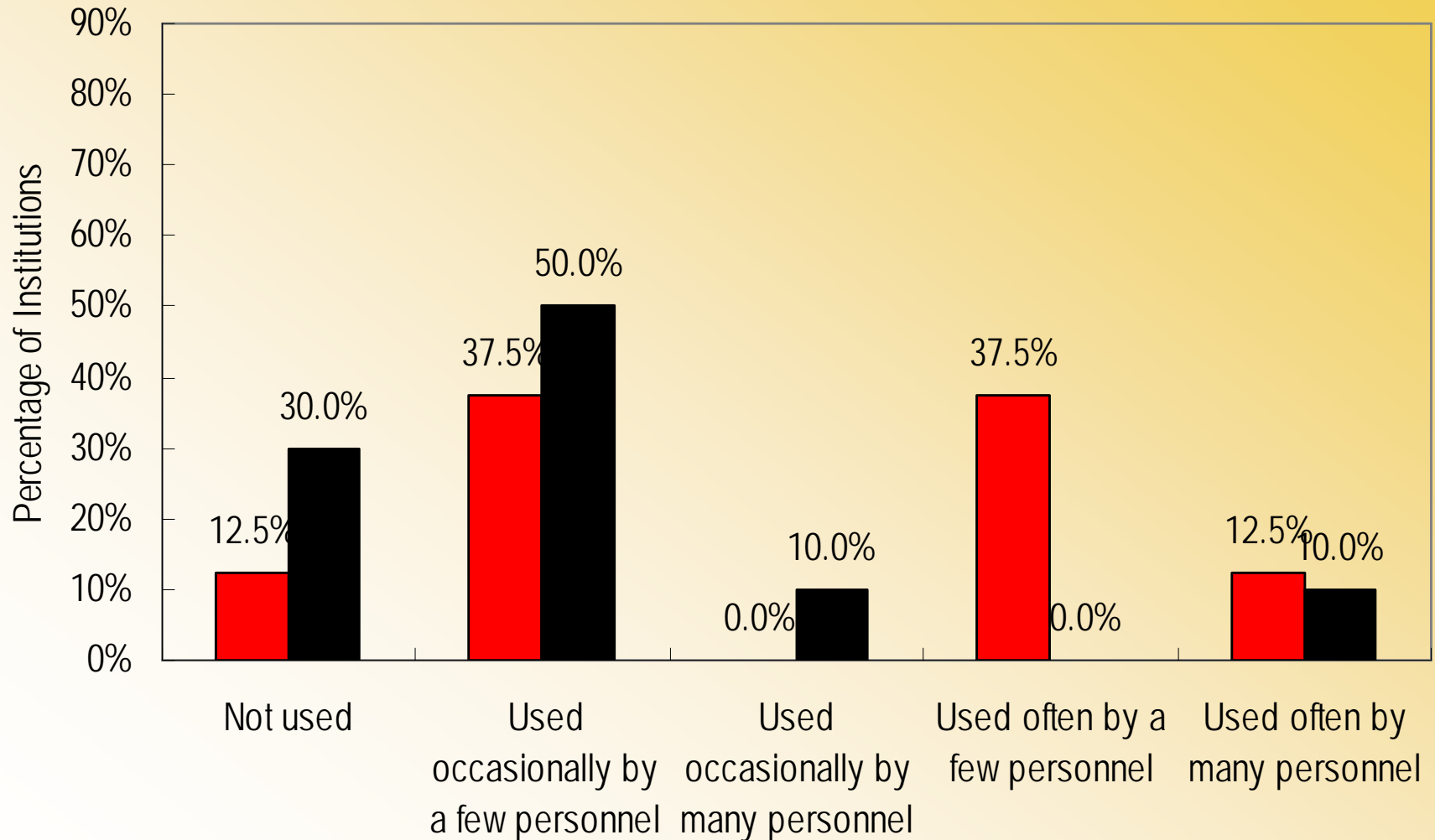
# CI Applications and Tools (CLAC)



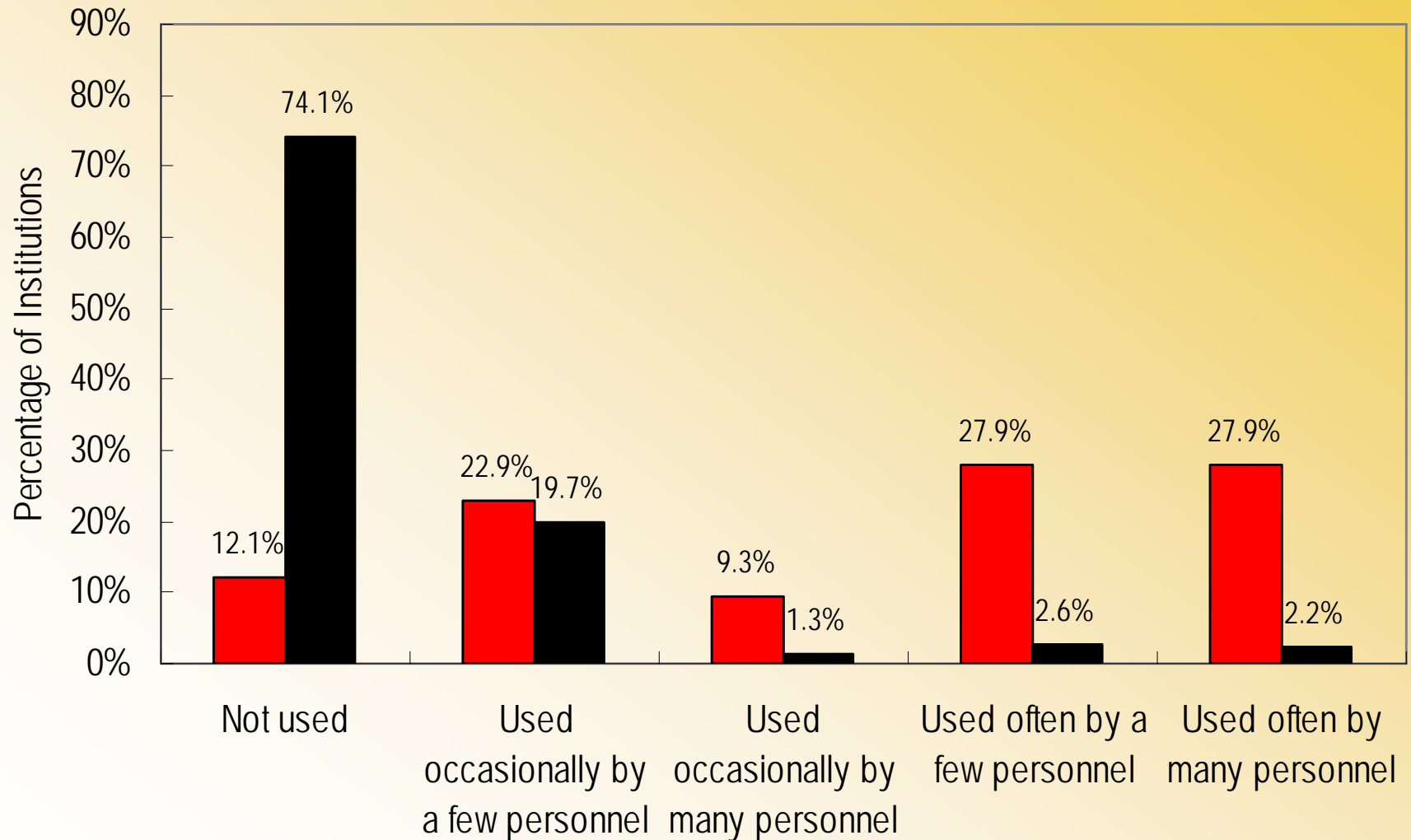
# Data Storage & Management



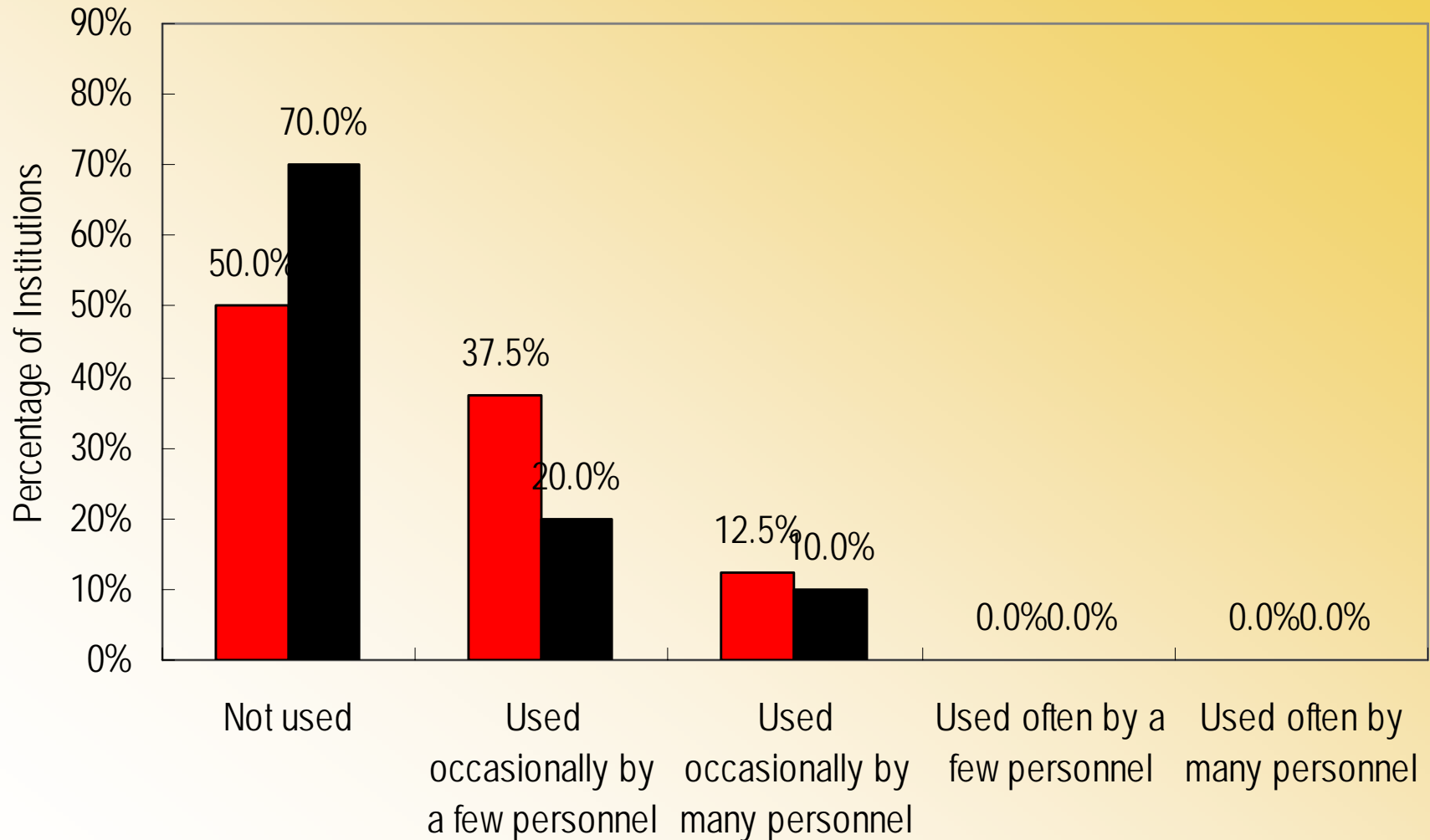
# Data Storage & Management (CLAC)



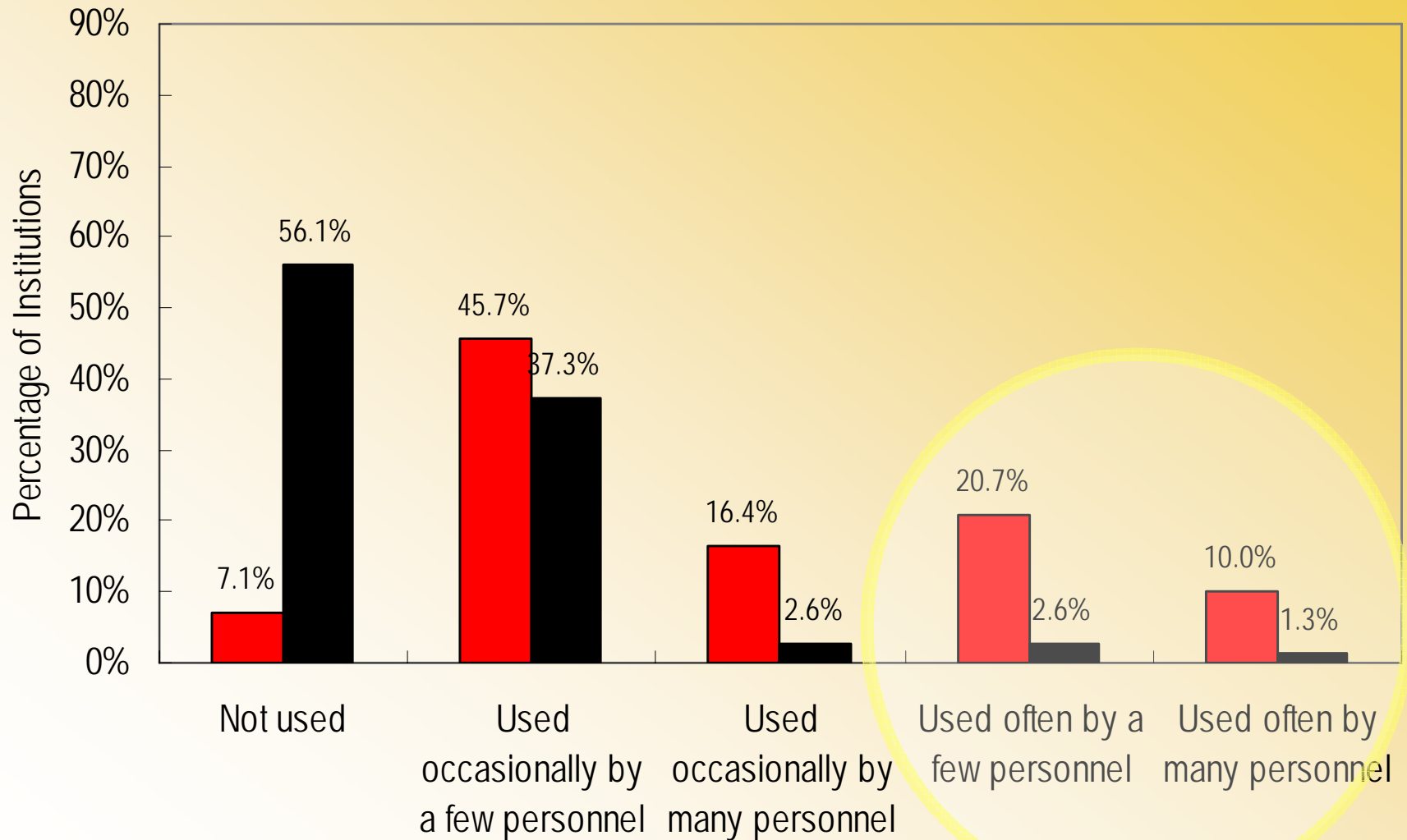
# Advanced Network Infrastructure



# Advanced Network Infrastructure (CLAC)

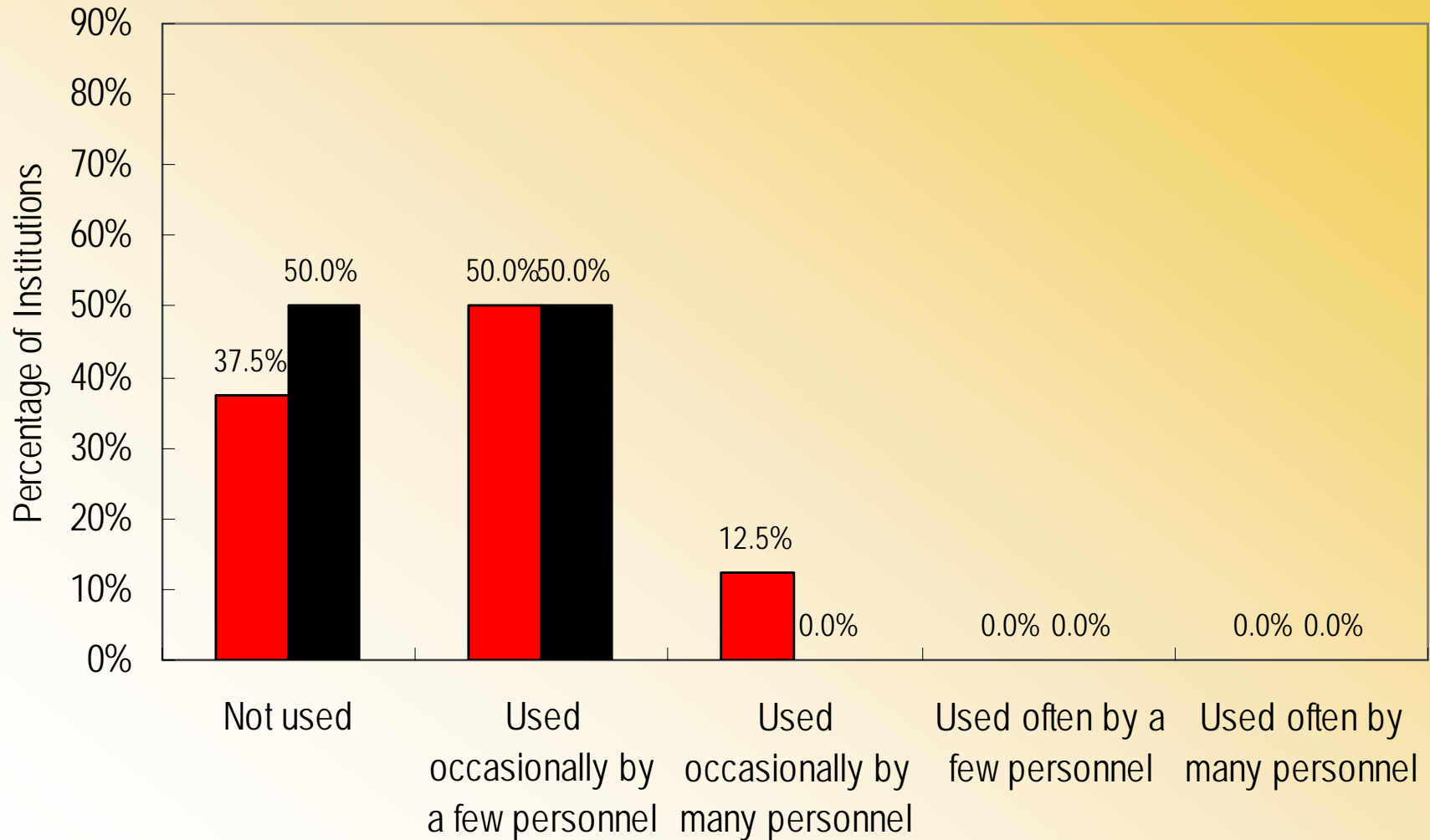


# Resources for Collaboration Within Virtual Communities

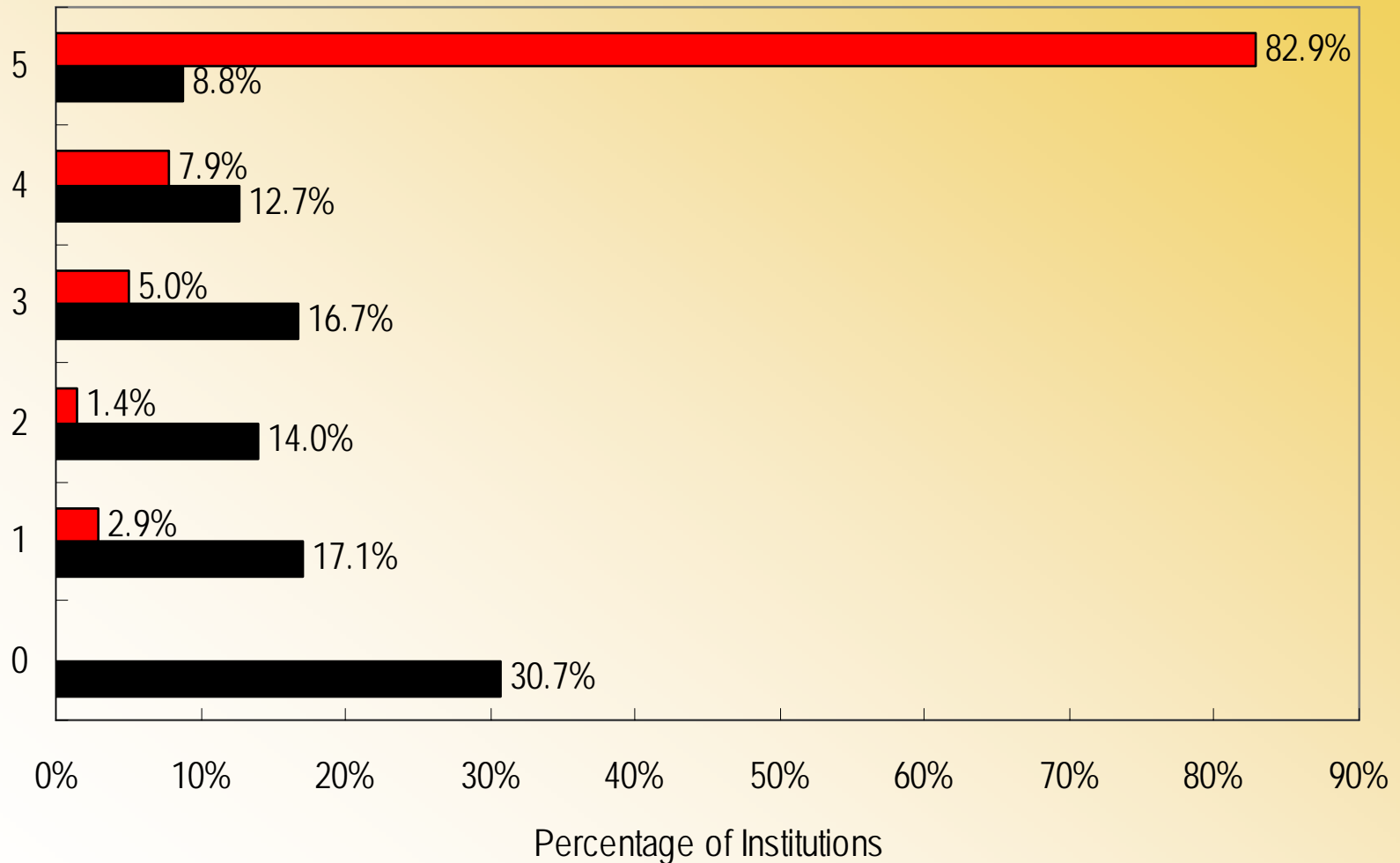


■ Research mission (N=140) ■ Teaching mission (N=228)

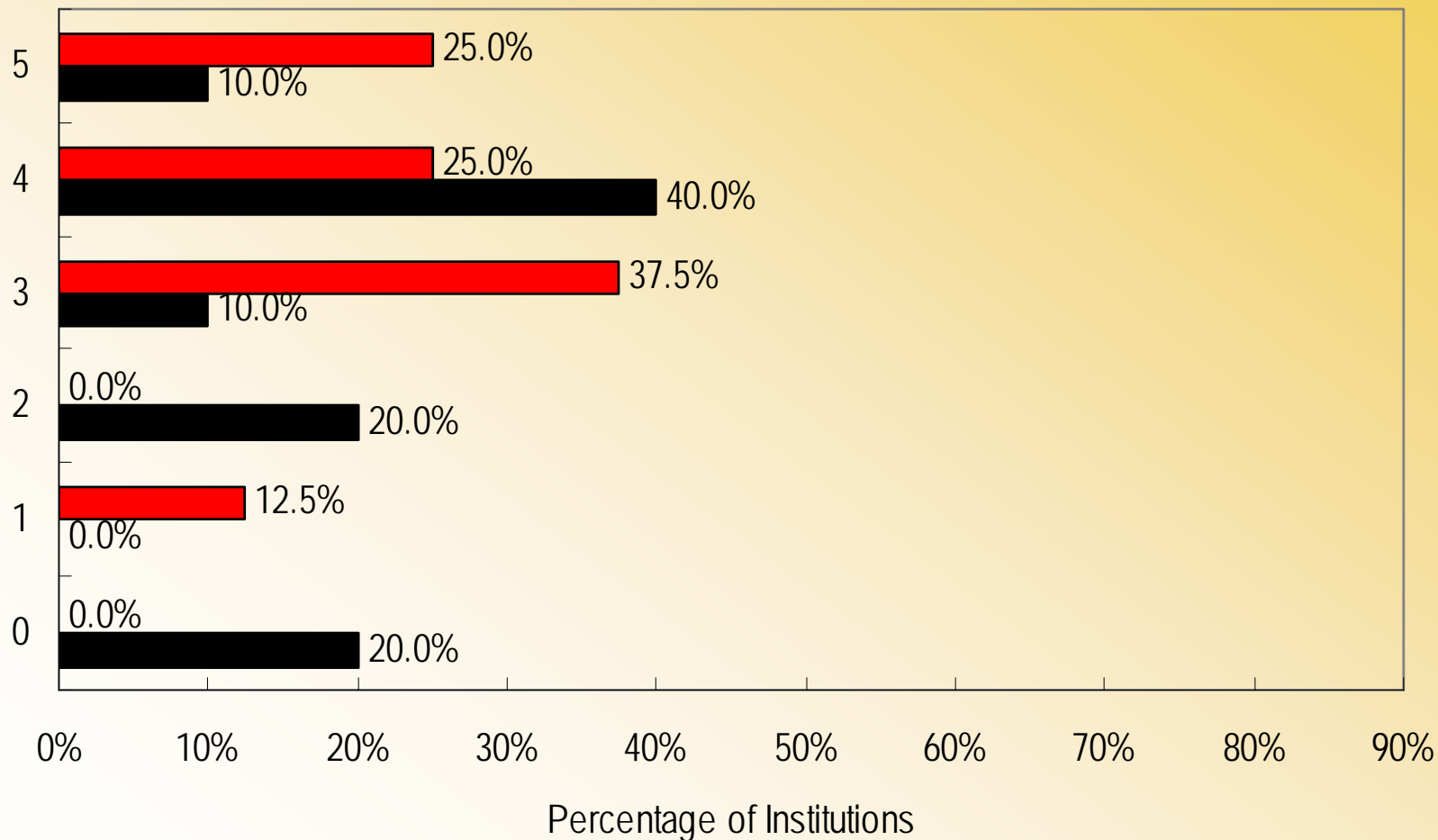
# Resources for Collaboration Within Virtual Communities (CLAC)



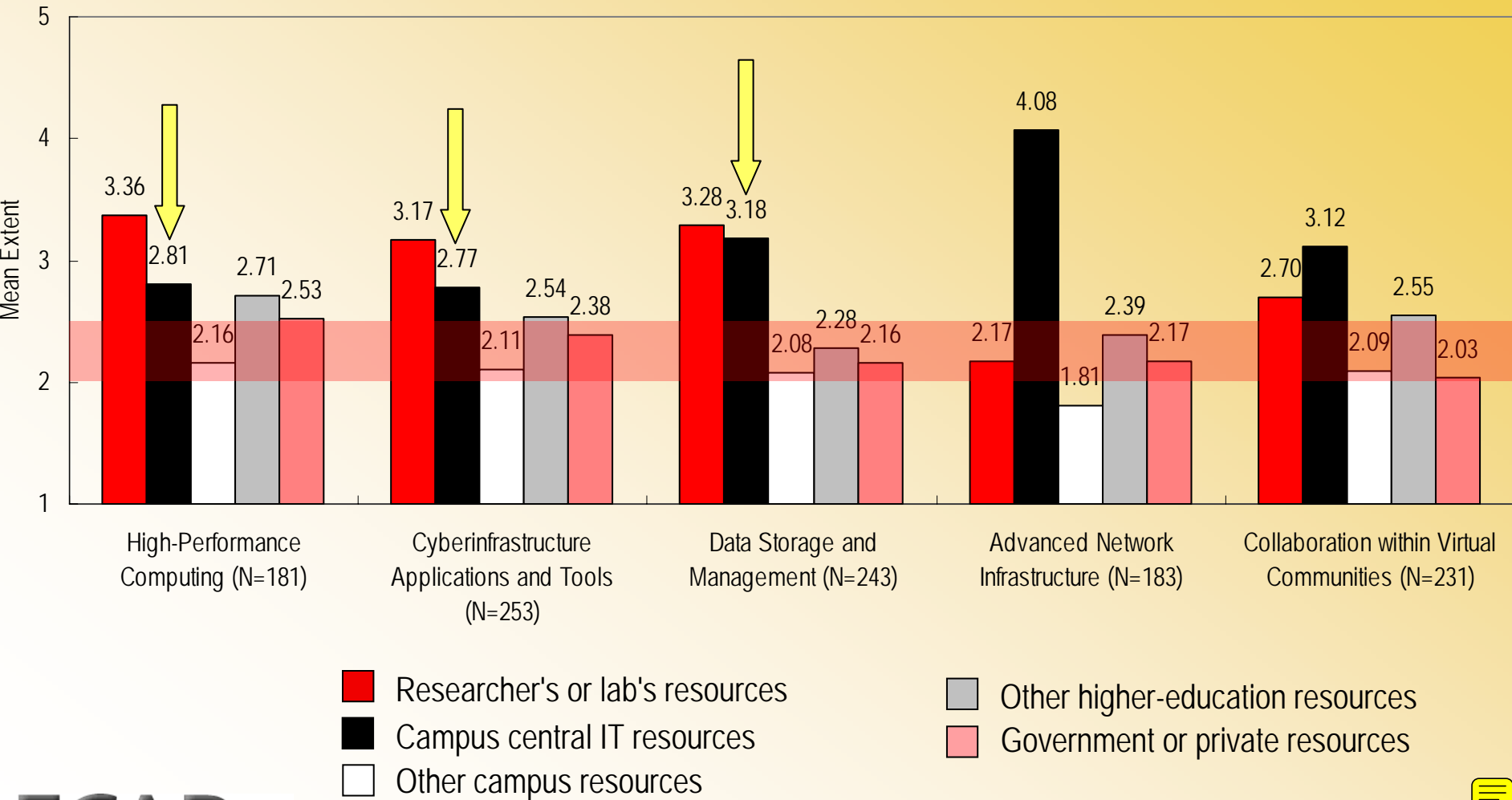
# Number of Technologies Used



# Number of Technologies Used (CLAC)



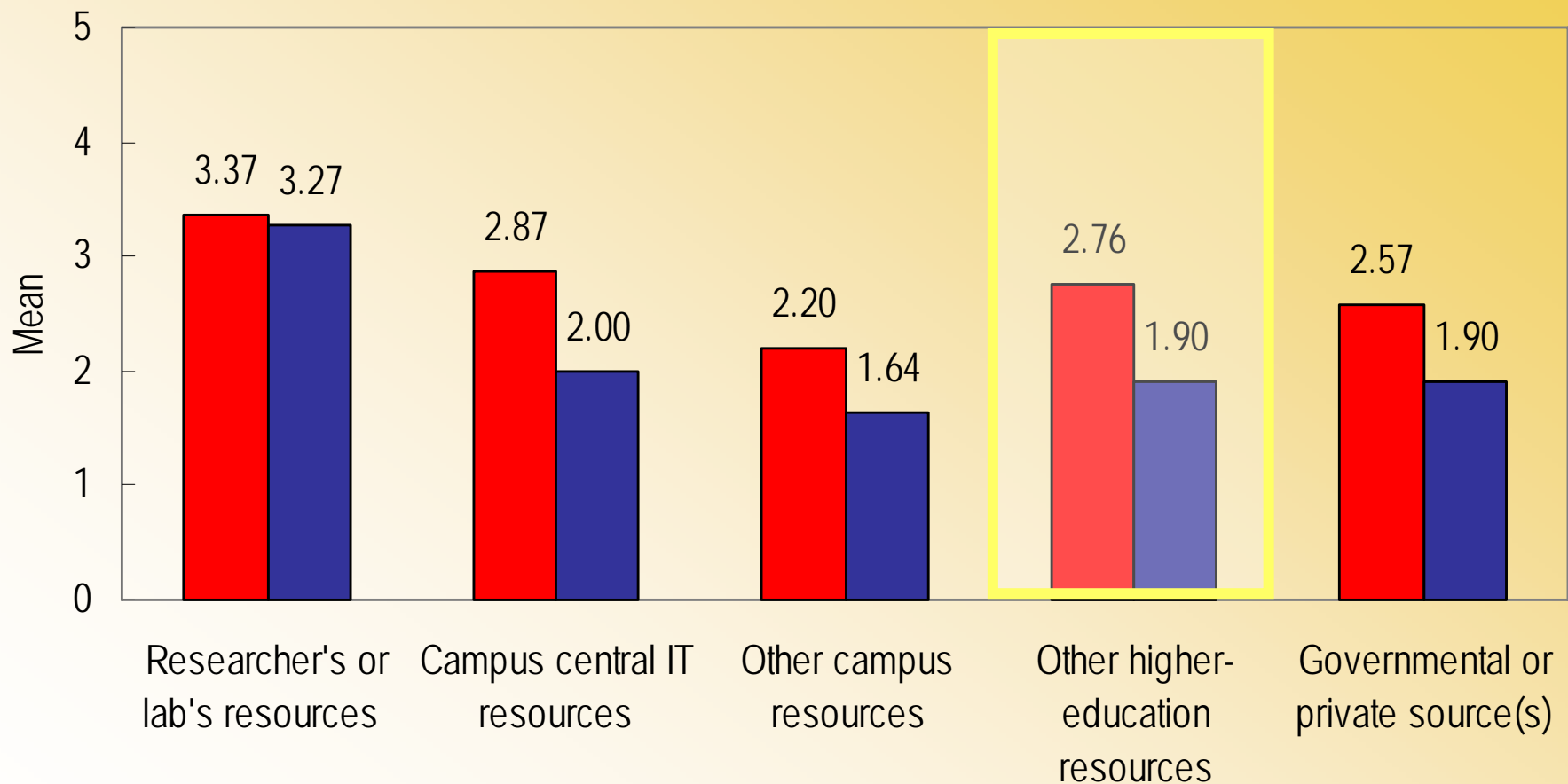
# Sourcing of CI Technologies



Scale: 1=none, 2=small extent, 3=moderate extent, 4=large extent, 5=very large extent



# High-Performance Computing Source

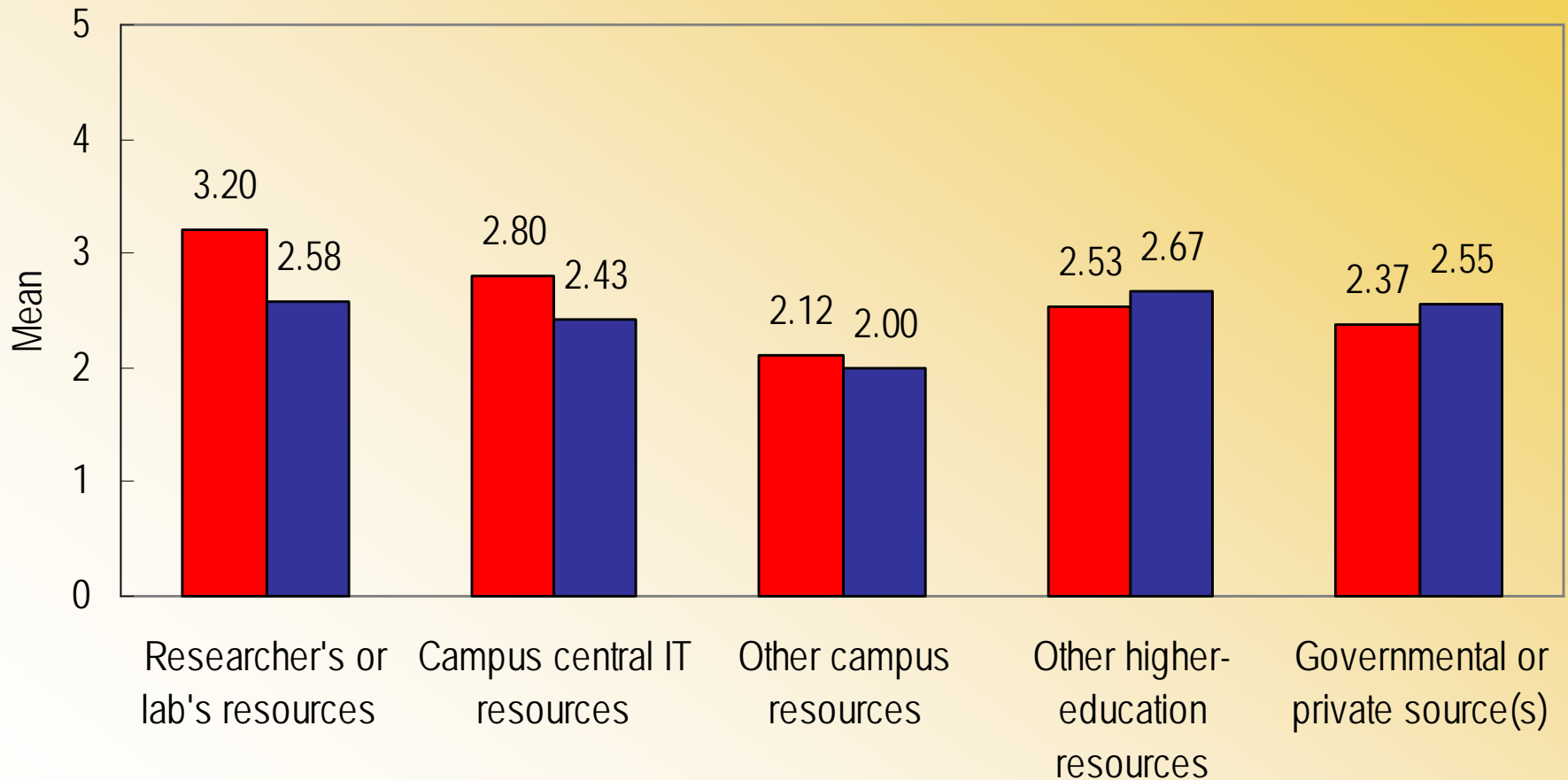


■ Non-Member (N=164) ■ CLAC Member (N=11)

Scale: 1=none, 2=small extent, 3=moderate extent, 4=large extent, 5=very large extent



# CI Applications and Tools Source

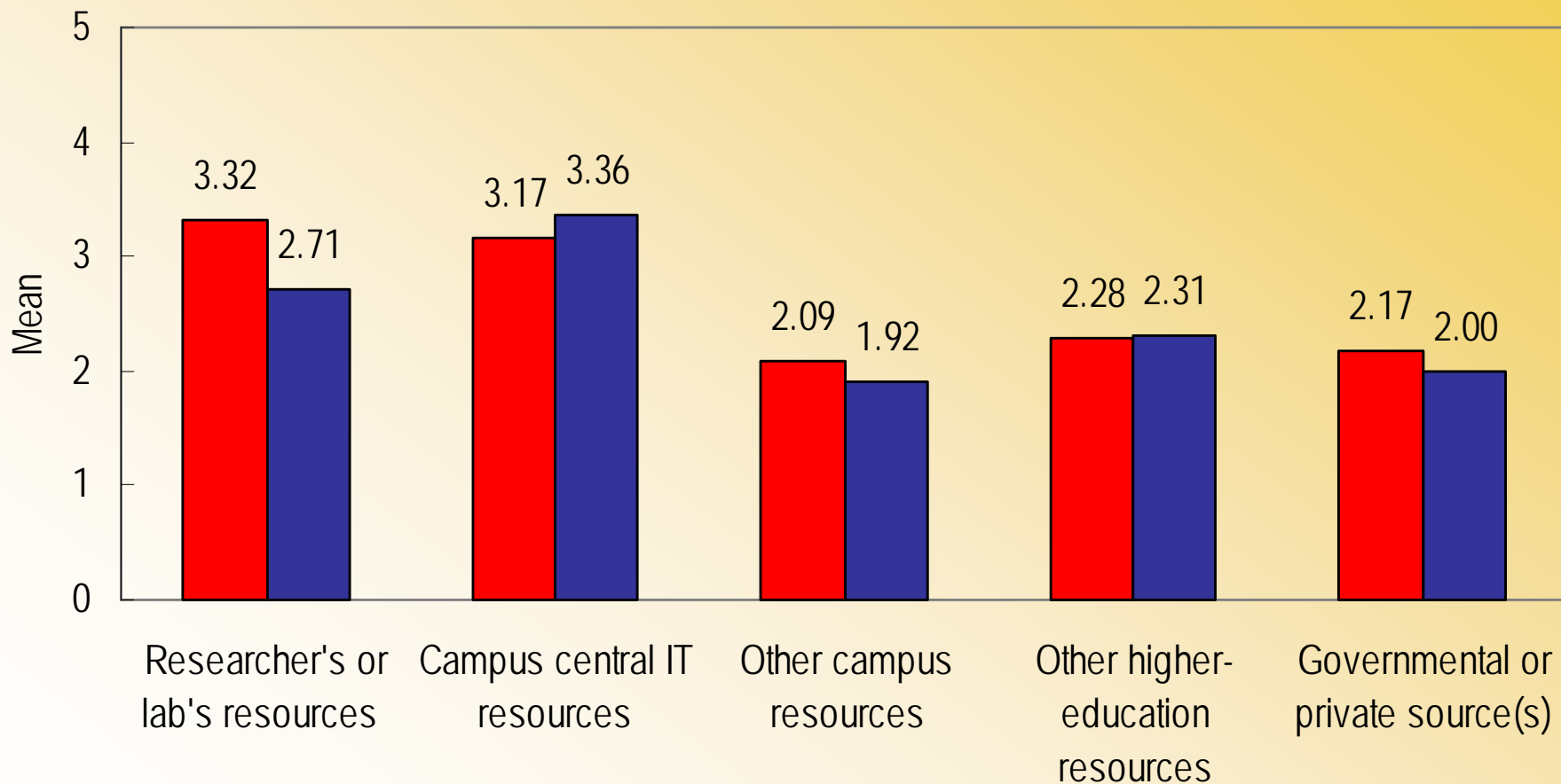


■ Non-Member (N=230) ■ CLAC Member (N=14)

Scale: 1=none, 2=small extent, 3=moderate extent, 4=large extent, 5=very large extent



# Data Storage and Management Source

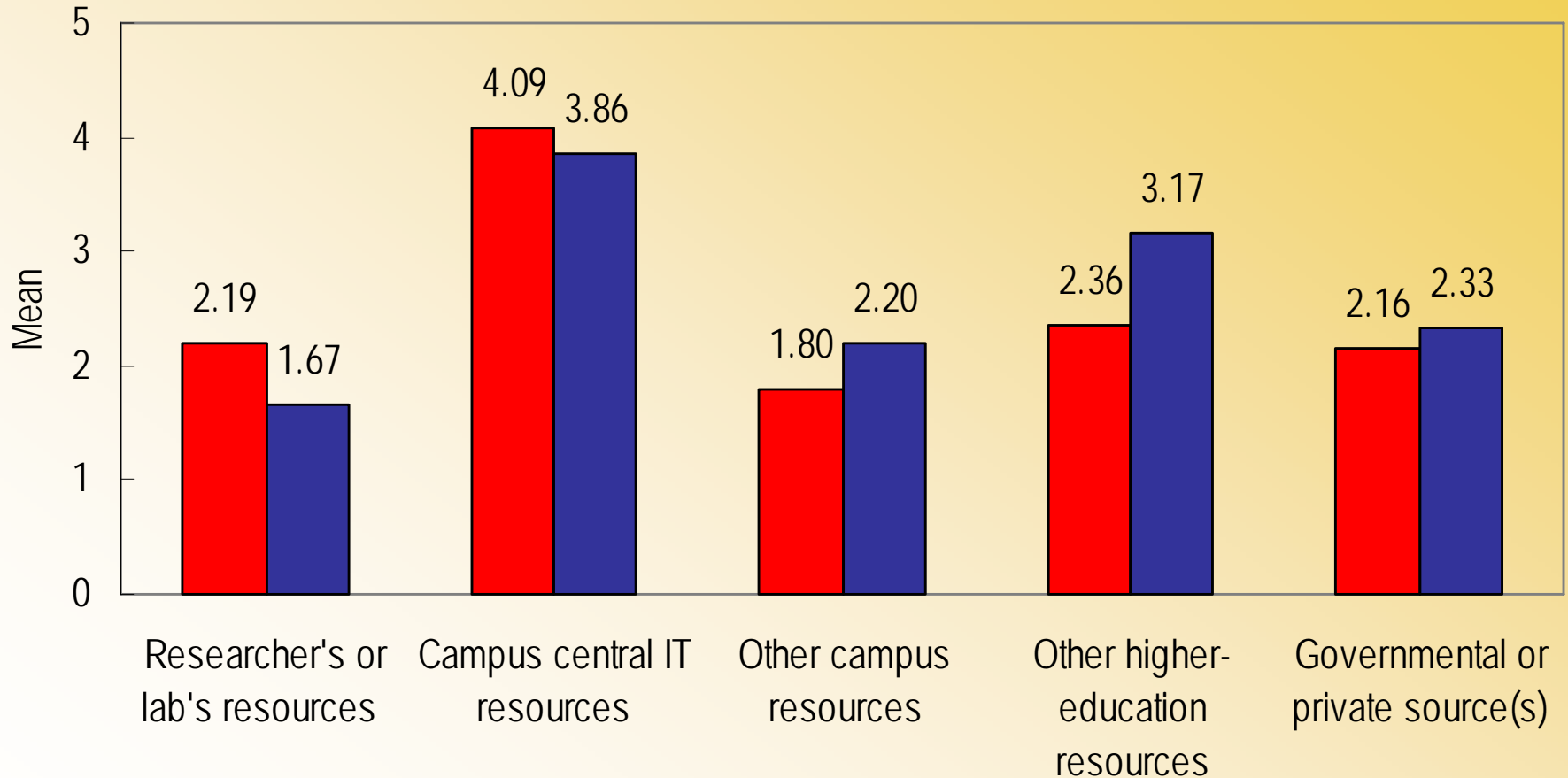


■ Non-Member (N=224)    ■ CLAC Member (N=14)

Scale: 1=none, 2=small extent, 3=moderate extent, 4=large extent, 5=very large extent



# Advanced Network Infrastructure Source

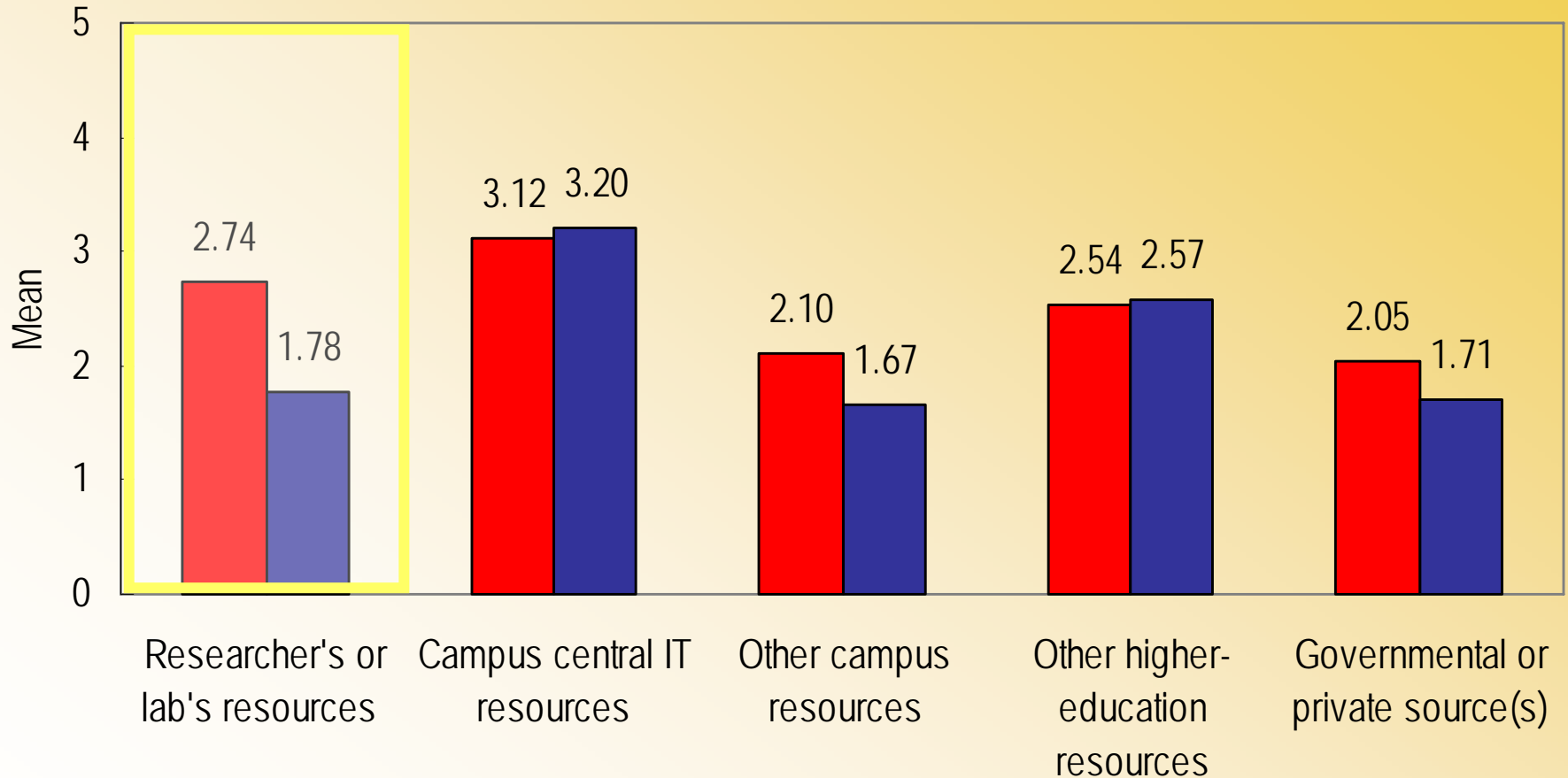


■ Non-Member (N=175) ■ CLAC Member (N=7)

Scale: 1=none, 2=small extent, 3=moderate extent, 4=large extent, 5=very large extent



# Collaboration Tools Source

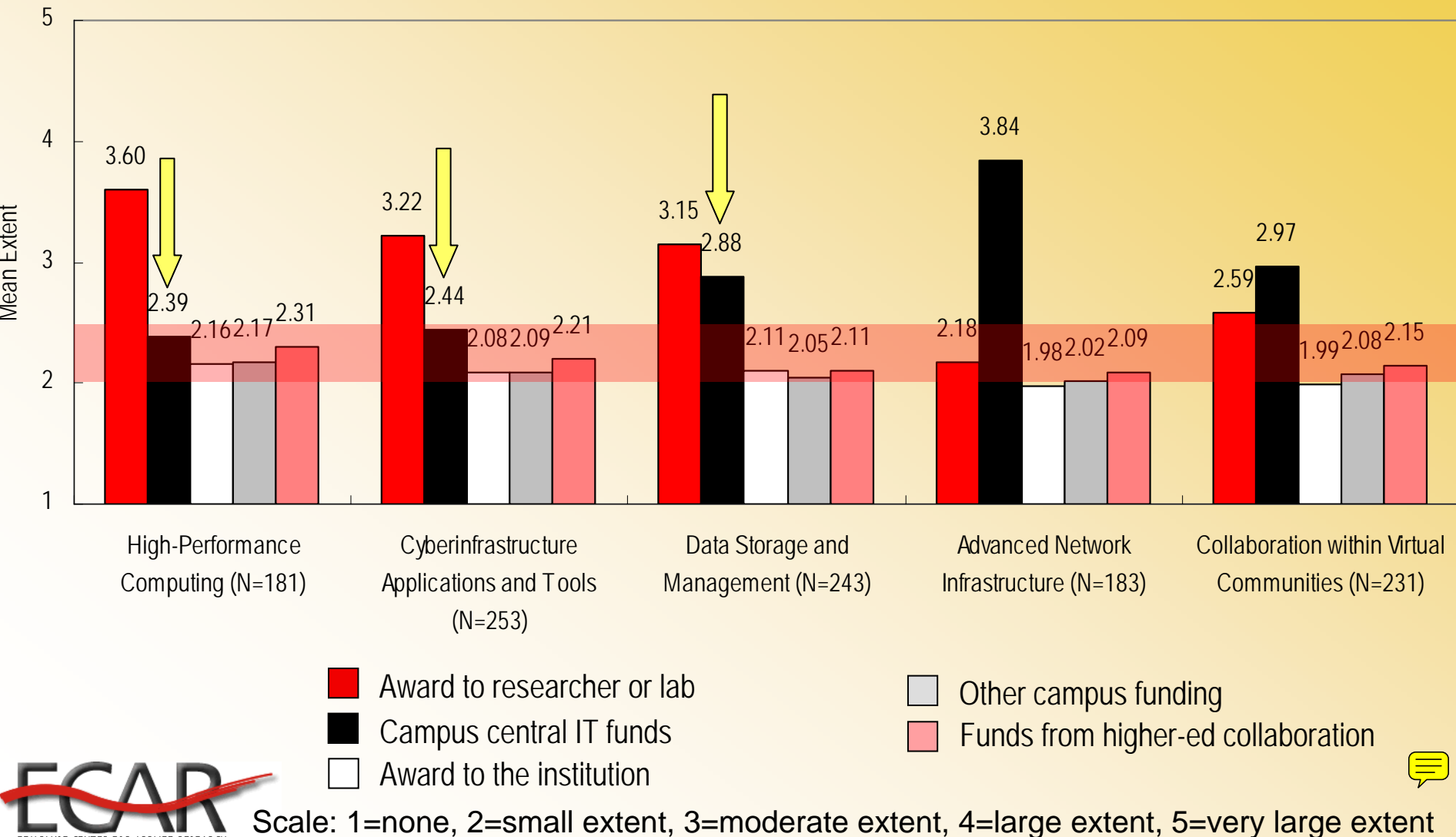


■ Non-Member (N=207) ■ CLAC Member (N=10)

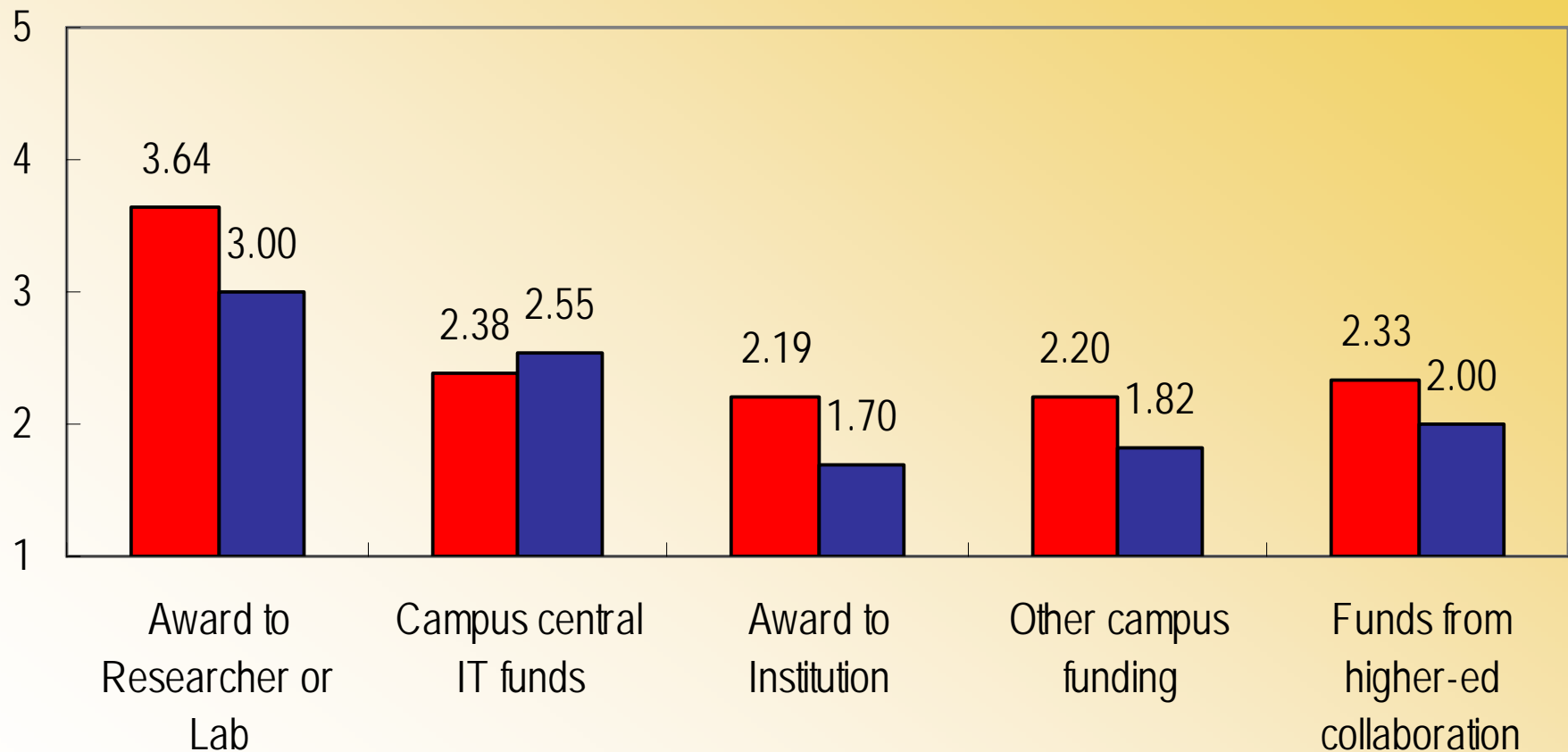
Scale: 1=none, 2=small extent, 3=moderate extent, 4=large extent, 5=very large extent



# Funding of CI Technologies



# High-Performance Computing Funding (CLAC)

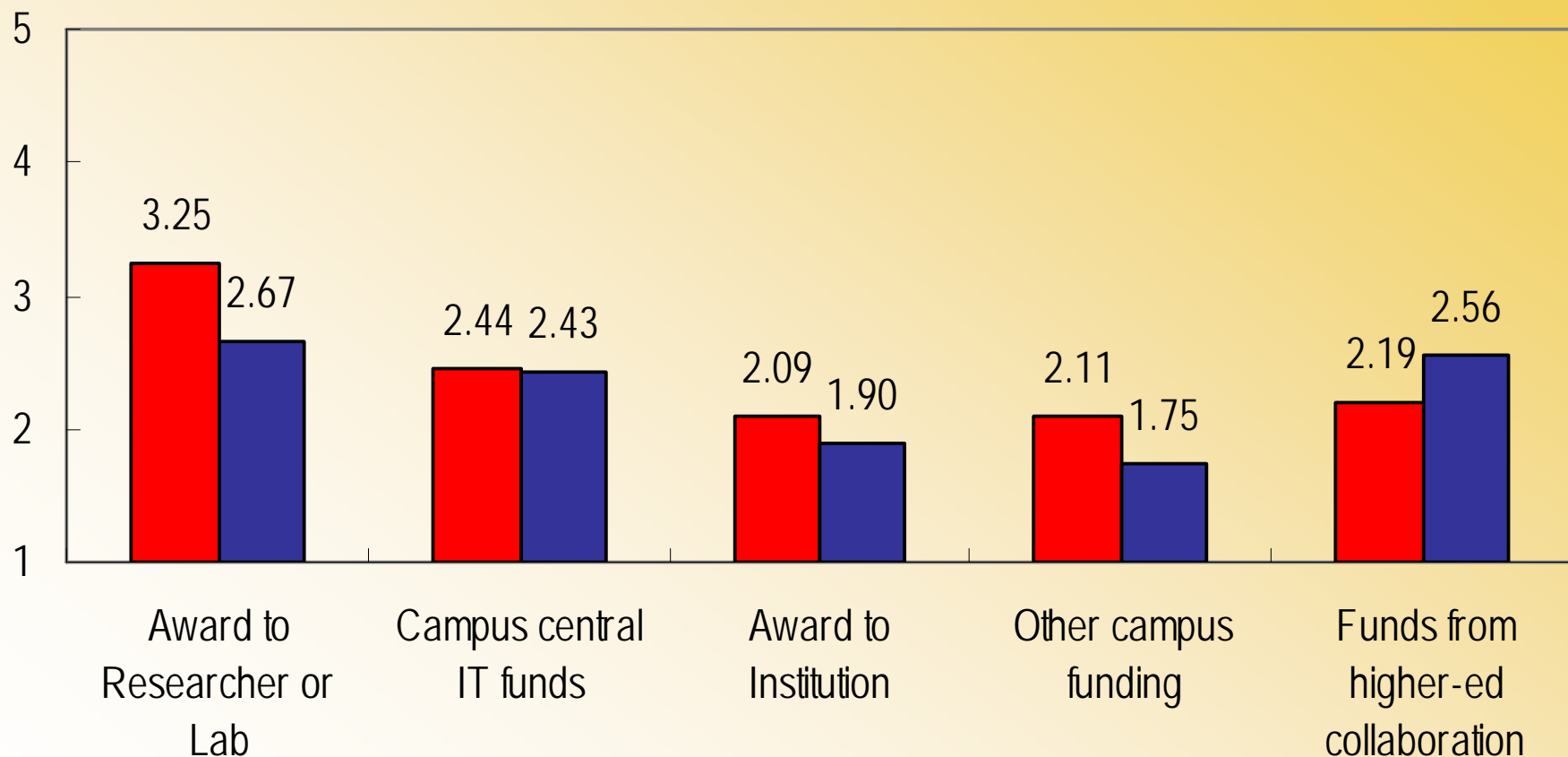


■ Non-member (N=165) ■ CLAC Member (N=11)

Scale: 1=none, 2=small extent, 3=moderate extent, 4=large extent, 5=very large extent



# CI Applications and Tools Funding (CLAC)

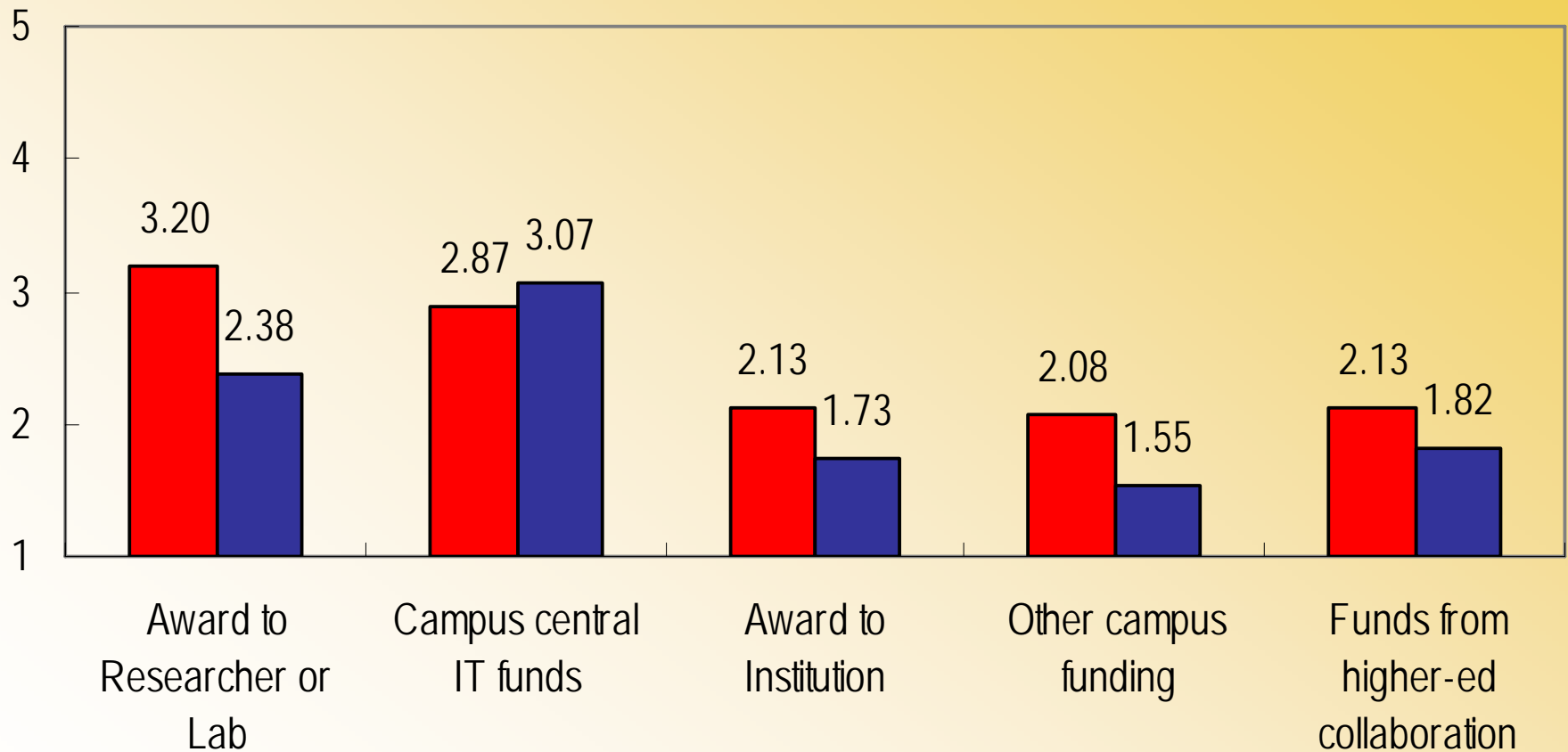


■ Non-member (N=227) ■ CLAC Member (N=14)

Scale: 1=none, 2=small extent, 3=moderate extent, 4=large extent, 5=very large extent



# Data Storage & Management Funding (CLAC)

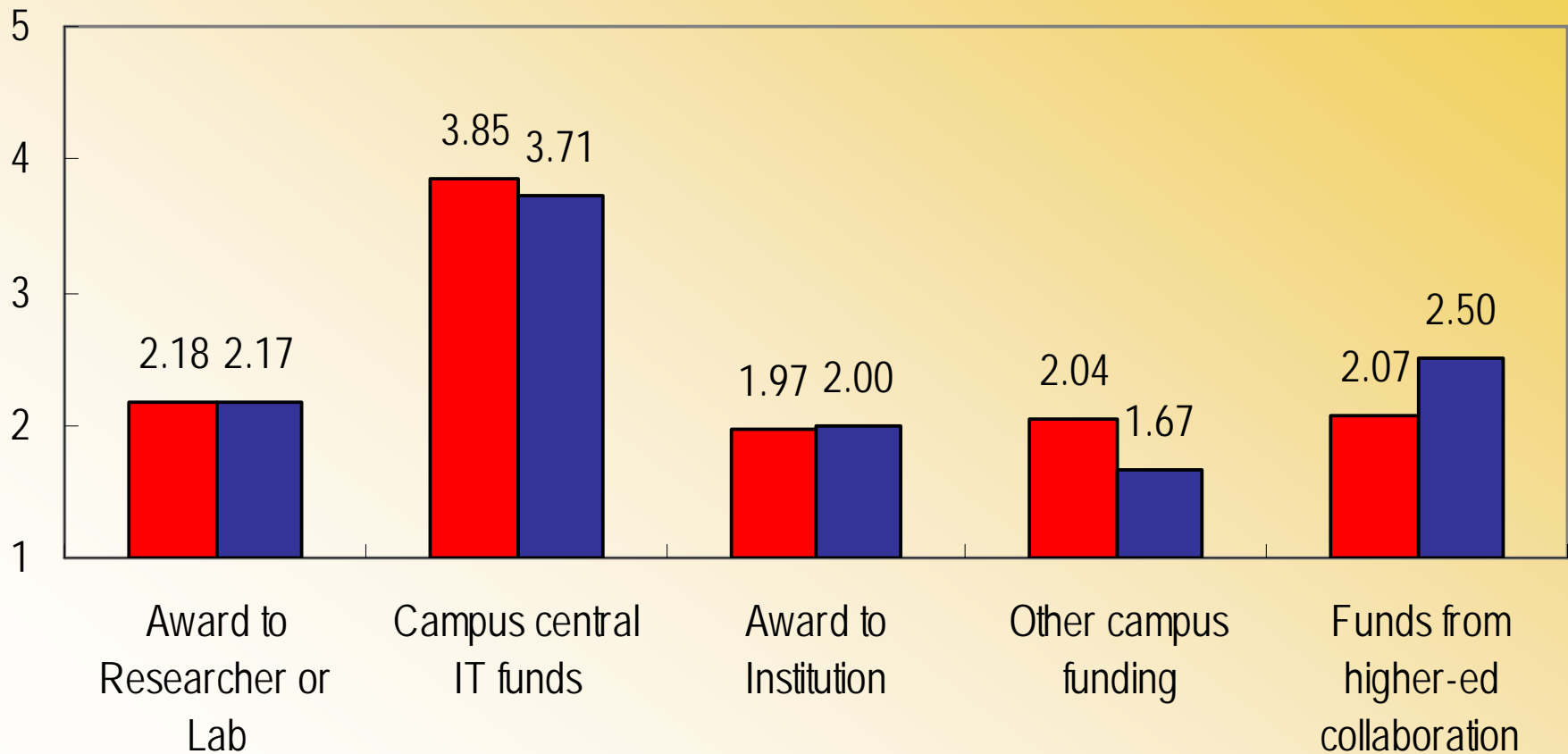


■ Non-member (N=220) ■ CLAC Member (N=14)

Scale: 1=none, 2=small extent, 3=moderate extent, 4=large extent, 5=very large extent



# Advanced Network Infrastructure Funding (CLAC)

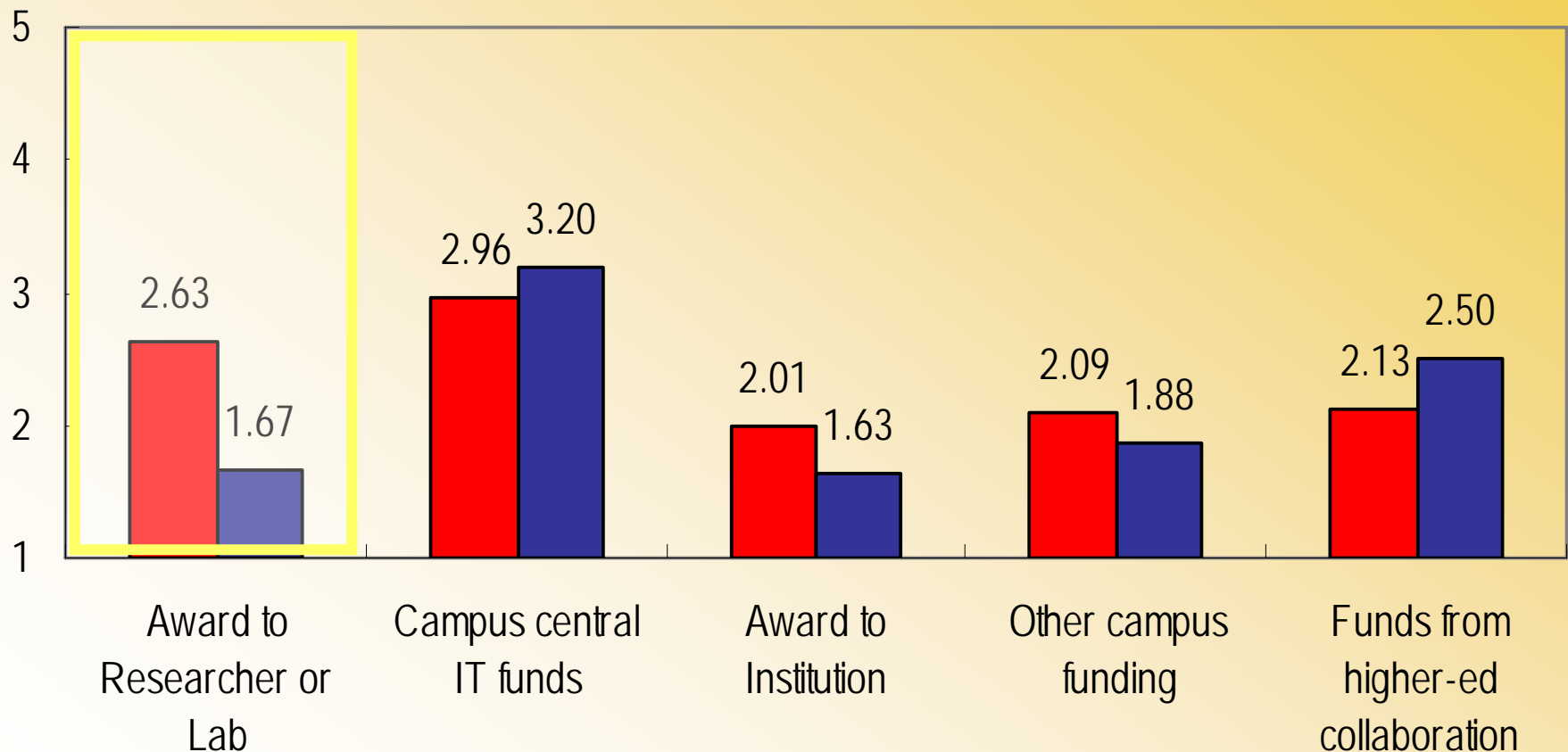


■ Non-member (N=170) ■ CLAC Member (N=7)

Scale: 1=none, 2=small extent, 3=moderate extent, 4=large extent, 5=very large extent



# Collaboration Tools Funding (CLAC)

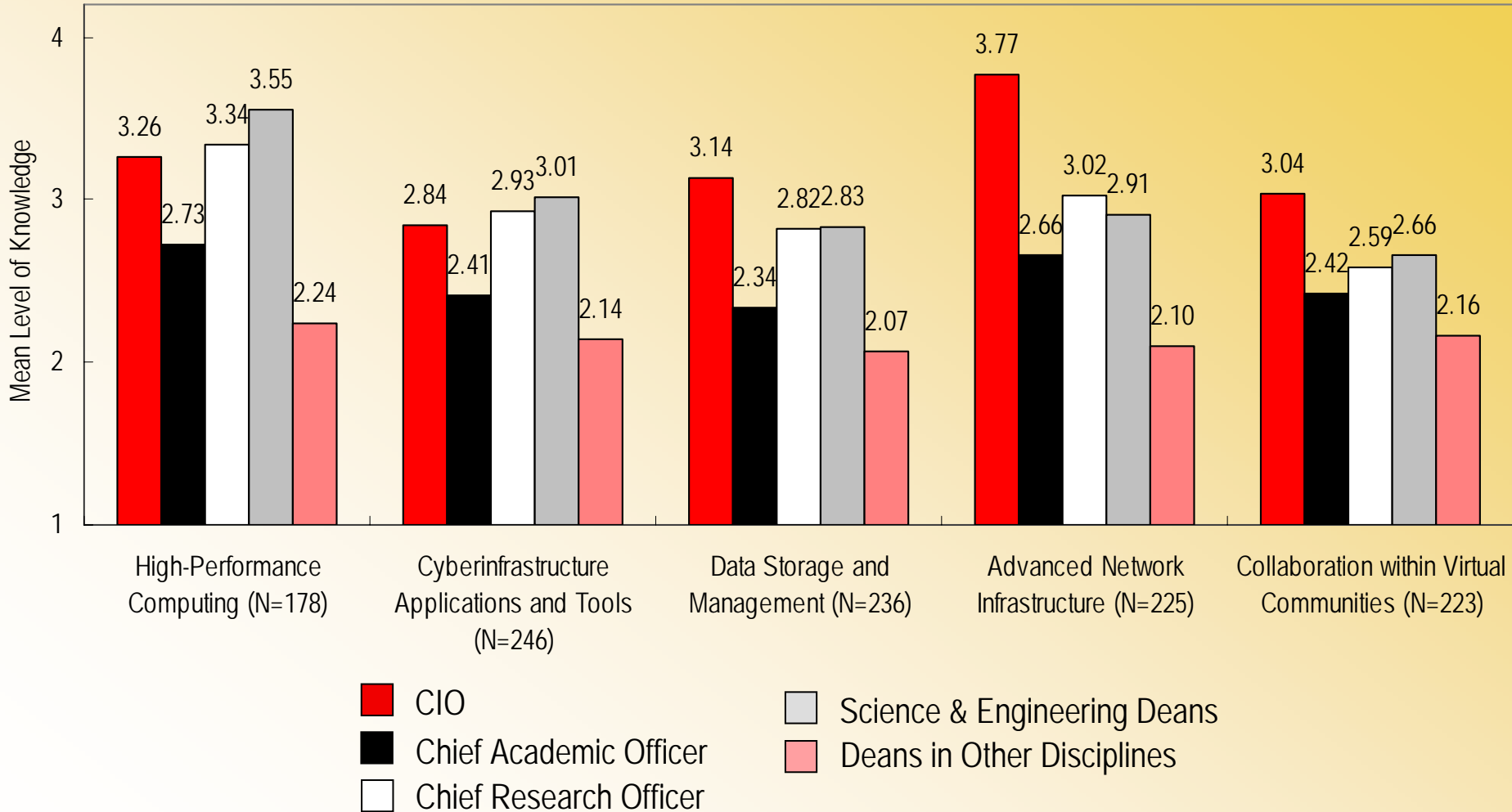


■ Non-member (N=210) ■ CLAC Member (N=10)

Scale: 1=none, 2=small extent, 3=moderate extent, 4=large extent, 5=very large extent



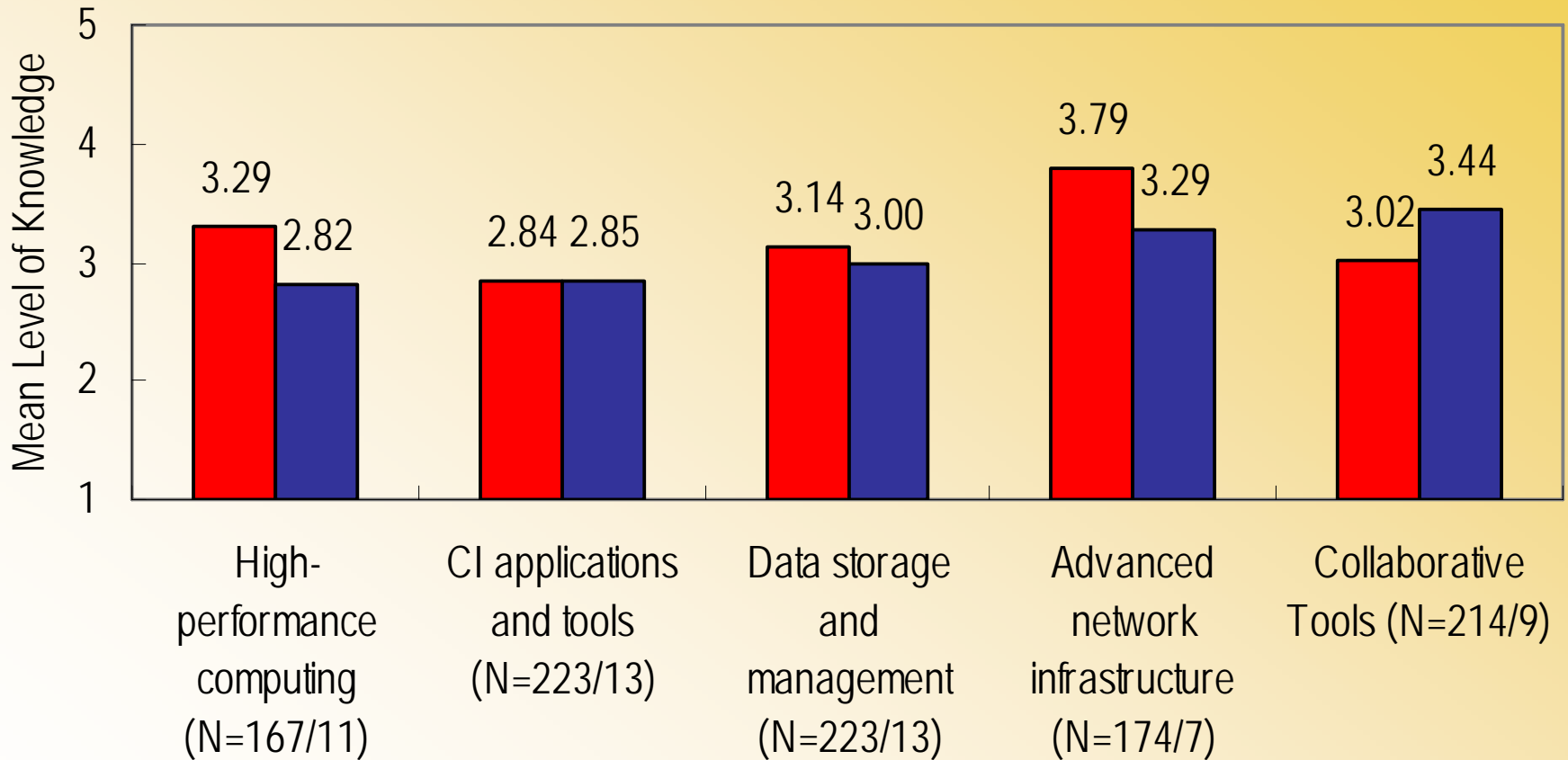
# Campus Executives' Knowledge



Scale: 1=poor, 2=fair, 3=good, 4=very good, 5=excellent



# CIO's Knowledge (CLAC)



■ Non-member ■ CLAC Member

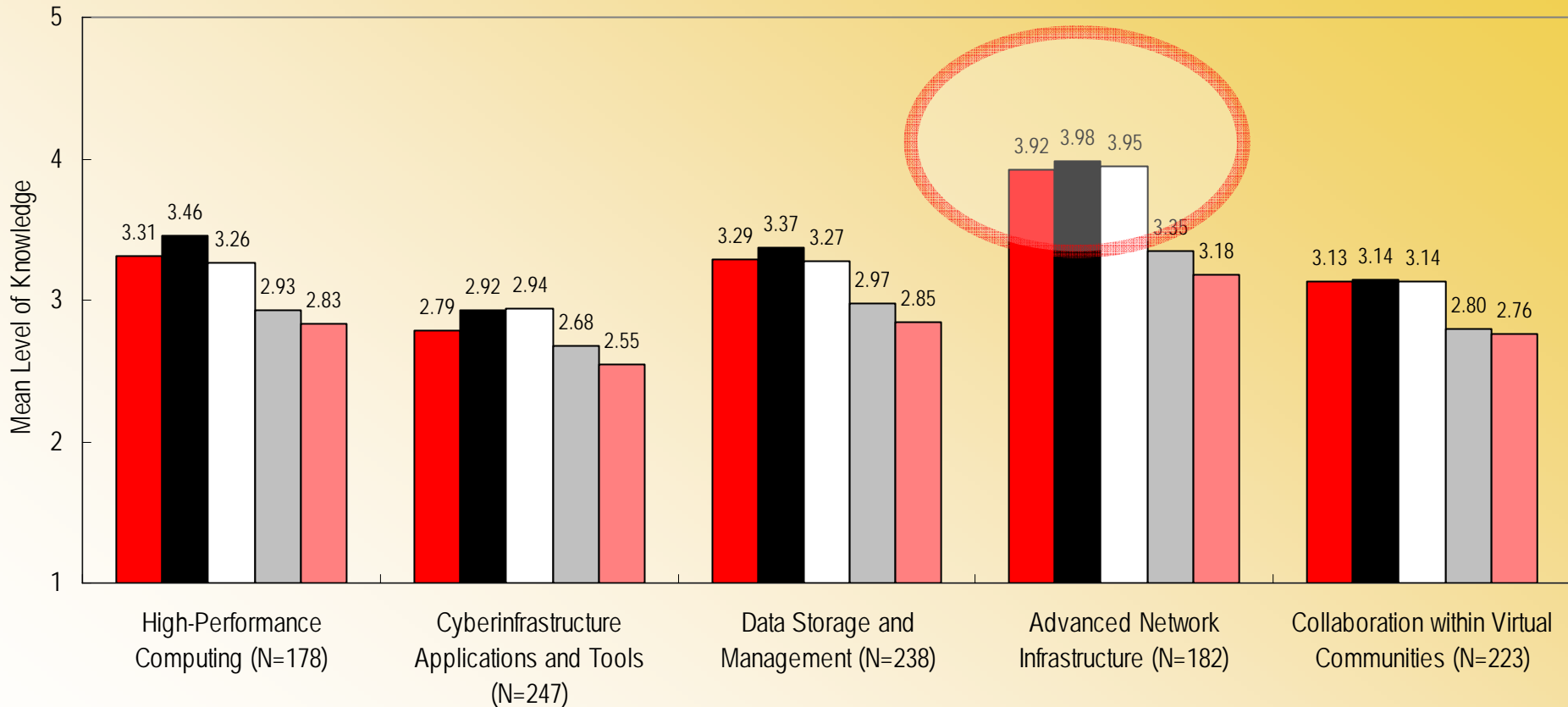
Scale: 1=poor, 2=fair, 3=good, 4=very good, 5=excellent



# Knowledge and Integration of CI Resources

<i>CI Technology</i>	Chief Information Officer	Chief Academic Officer	Chief Research Officer	Science and Engineering Deans	Other Deans
High-Performance Computing					
	✓	✓	✓		✓
Cyberinfrastructure Applications and Tools					
	✓	✓	✓	✓	
Data Storage and Management					
	✓	✓	✓	✓	
Advanced Network Infrastructure					
		✓			✓
Collaboration within Virtual Communities					
	✓	✓			✓

# Details of CIO's Knowledge



- What resources are available to researchers
- Who provides these resources
- Who funds these resources
- Who is using these resources
- What these resources are used for

Scale: 1=poor, 2=fair, 3=good, 4=very good, 5=excellent

# Accountability, Authority, and Resources

# Accountability, Authority, and Resources

## Research-related **Service** Activities

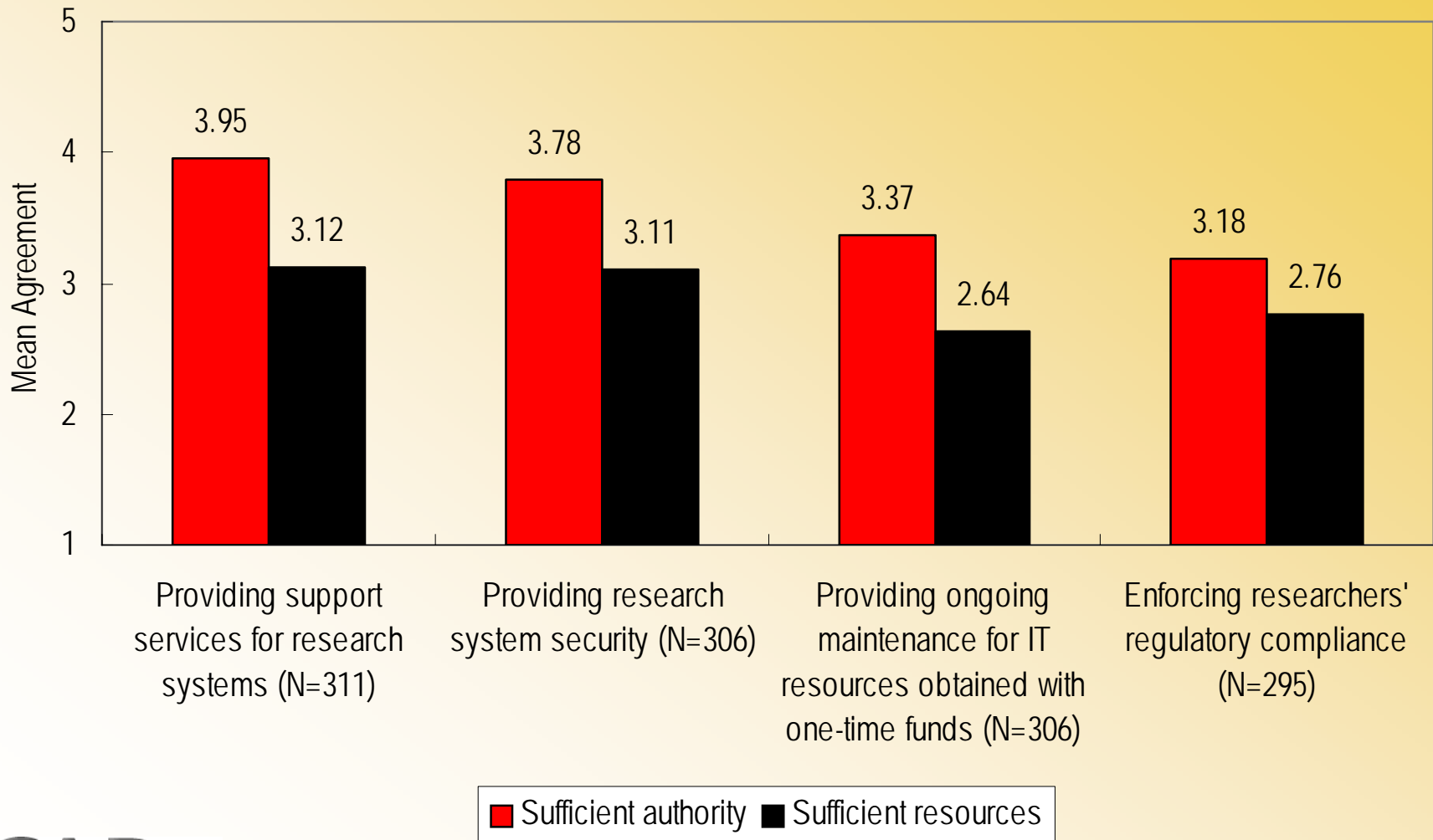
- Providing support services for research IT systems, such as system administration, identity management, and help desk
- Providing security for research systems
- Providing ongoing maintenance and support for IT resources obtained with one-time research funds
- Enforcing the research community's compliance with national regulations regarding privacy of data, such as HIPAA and FERPA in the U.S.

# Accountability, Authority, and Resources

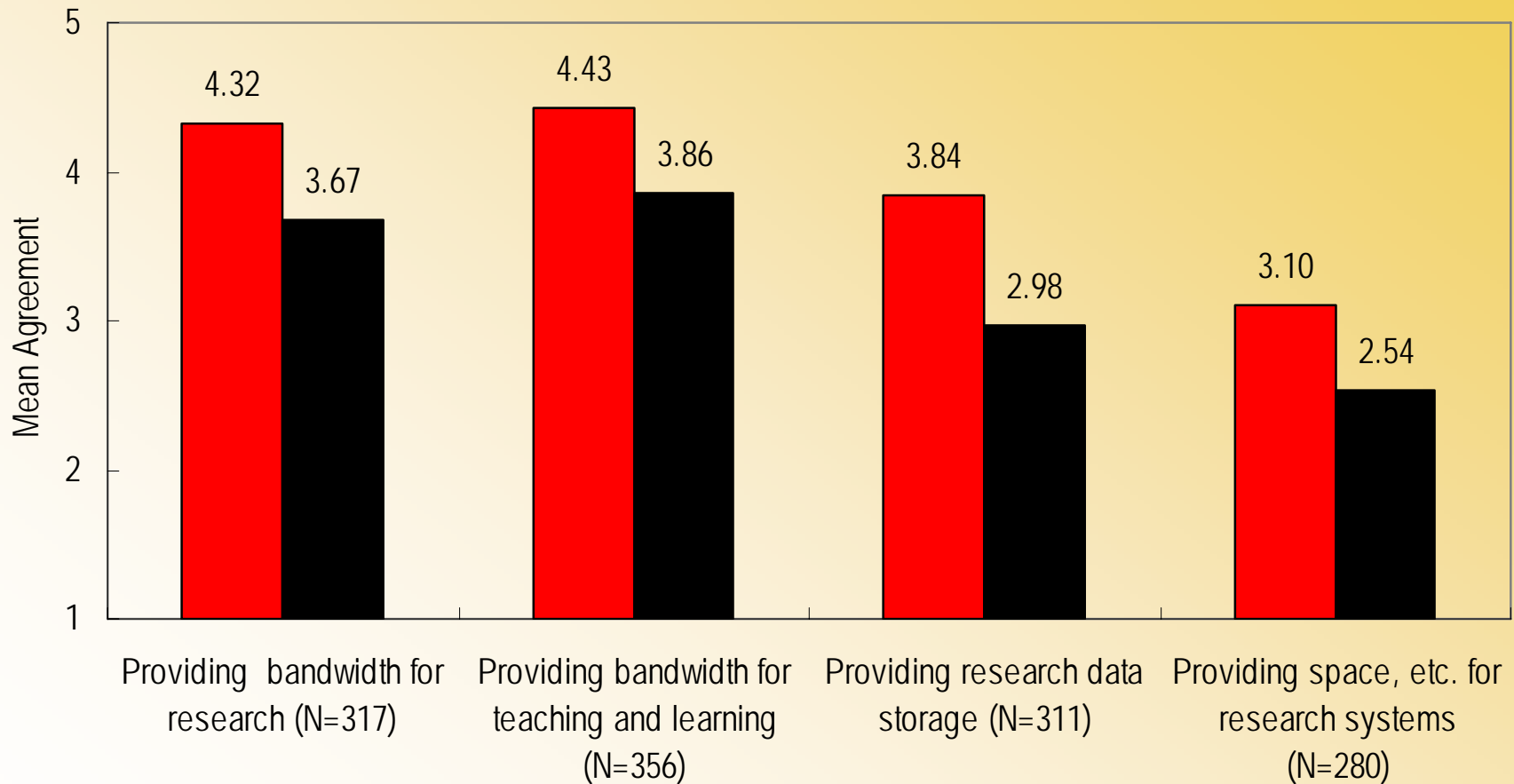
## Research-related **Infrastructure** Activities

- Providing sufficient network bandwidth for research
- Providing sufficient network bandwidth for teaching and learning
- Providing sufficient storage for research data
- Providing space and environmental support for research IT resources owned by campus entities other than central IT

# Research-Related Services



# Research-Related Infrastructure



■ Sufficient authority ■ Sufficient resources

Scale: 1=strongly disagree, 2=disagree, 3=neutral, 4=agree, 5=strongly agree



# Authority, and Resources (CLAC)

- Authority
  - No significant differences
  - Still...
    - CLAC CIOs appear to have a little more authority for **research data storage** and **ongoing maintenance of non-central research systems**
- Resources
  - No significant differences
  - Still...
    - CLAC CIOs appear to agree more strongly that they have sufficient resources for **research data storage, space and environmental support for research systems, ongoing maintenance of non-central research systems, and research bandwidth**



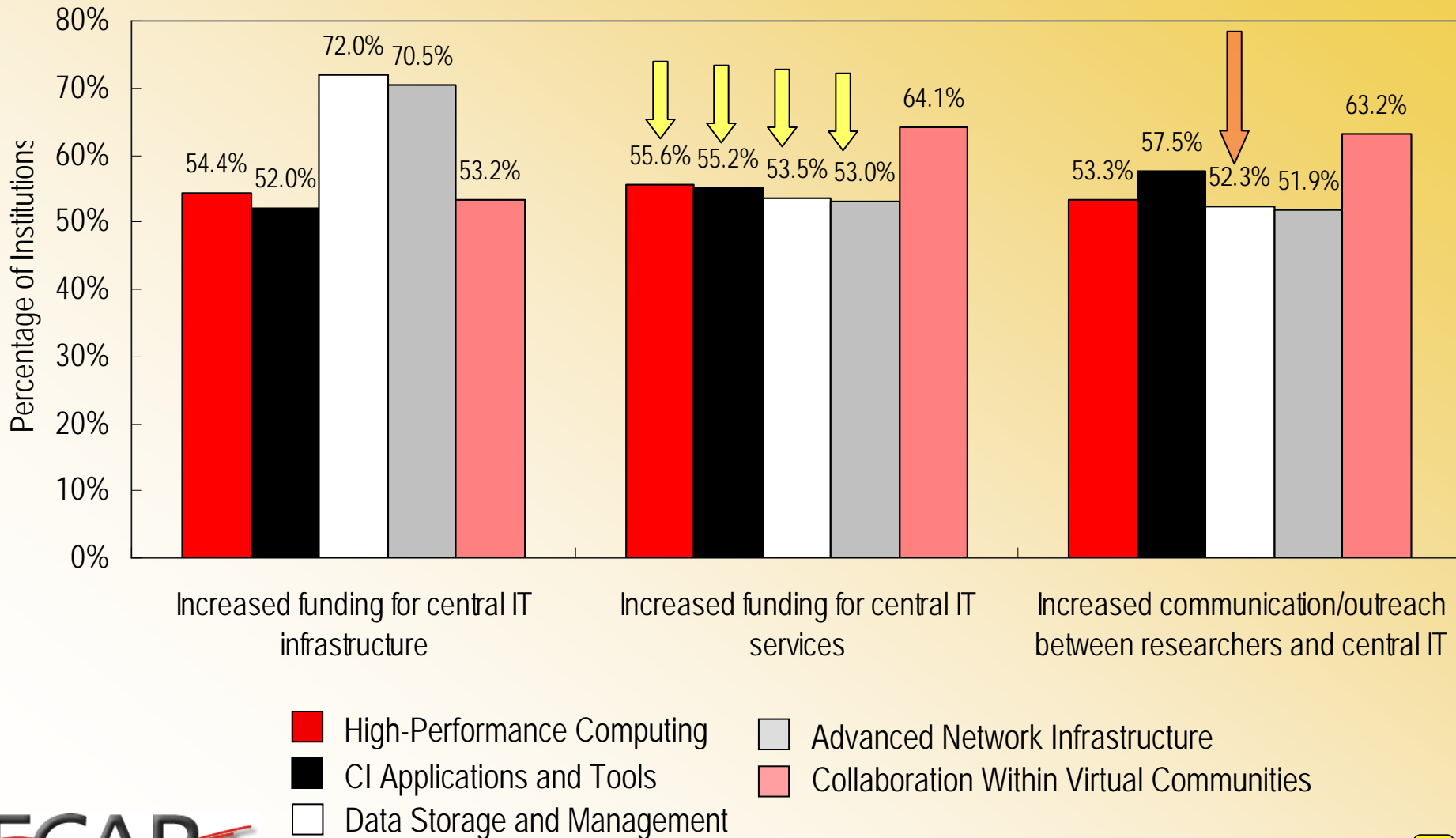
# Aids to Support for Research

**Eight Items** (all but #6 pertain to central IT):

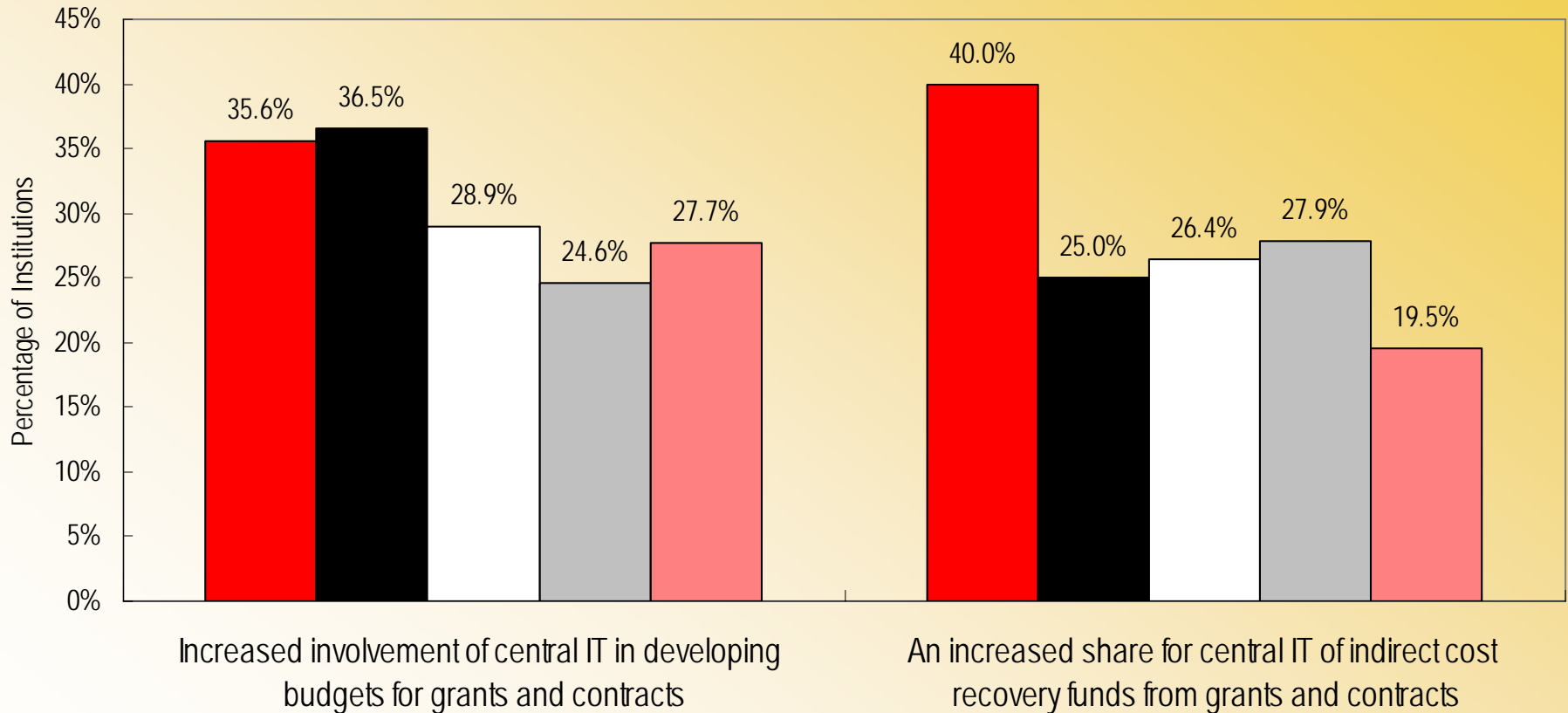
1. Increased funding for infrastructure
2. Increased funding for services
3. Increased communication with researchers
4. Increased involvement in grant writing
5. Increased share of IDCs
6. Greater inclusion of researchers in governance
7. Increased authority for resource acquisition
8. Increased authority for resource management

For each technology, respondents were asked to select the three that would be most helpful.

# Top-Tier Items



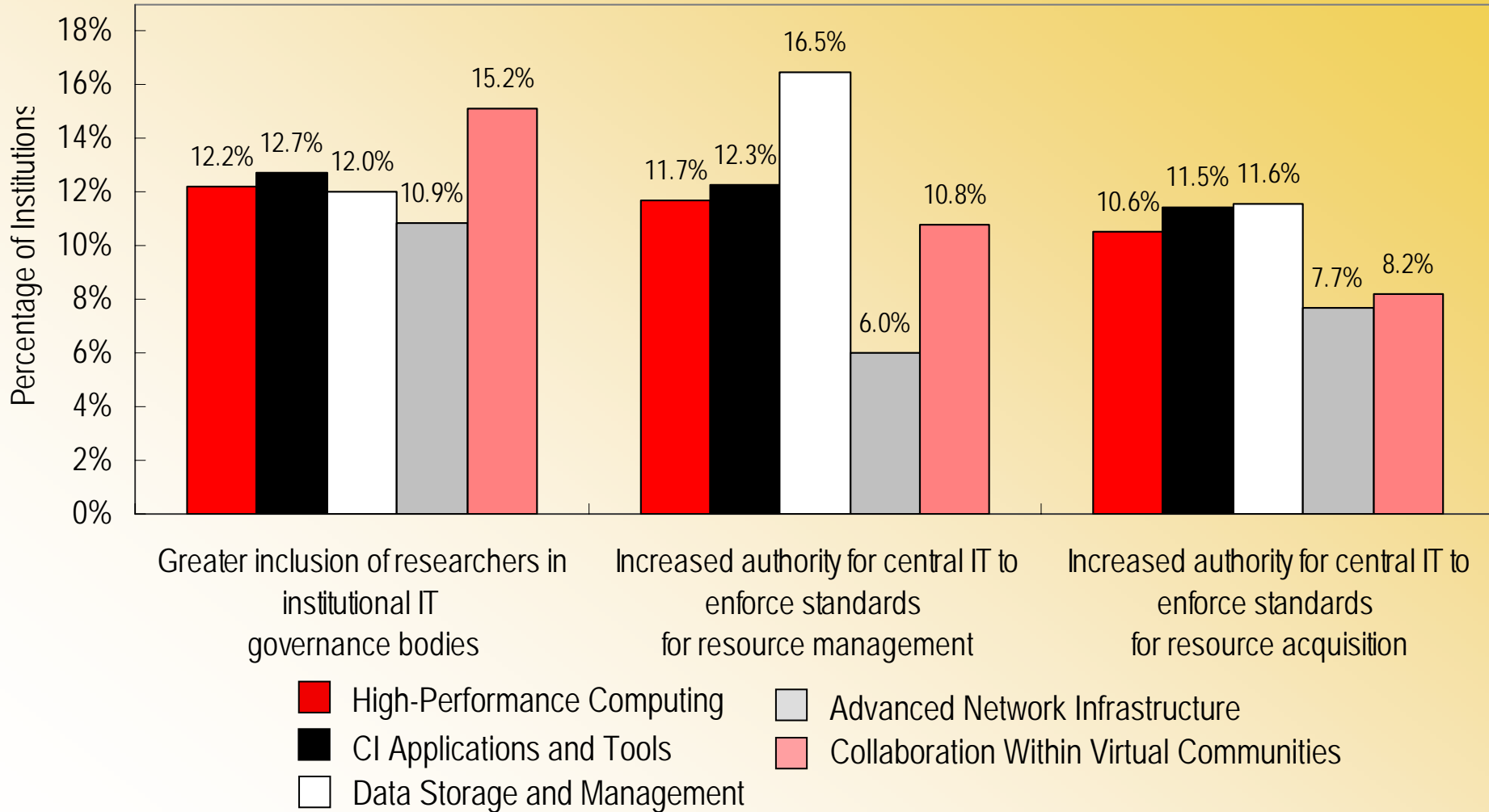
# Middle-Tier Items



- High-Performance Computing
- CI Applications and Tools
- Data Storage and Management

- Advanced Network Infrastructure
- Collaboration Within Virtual Communities

# Lowest-Tier Items



# Where CLAC Differs

(albeit not with statistical significance)

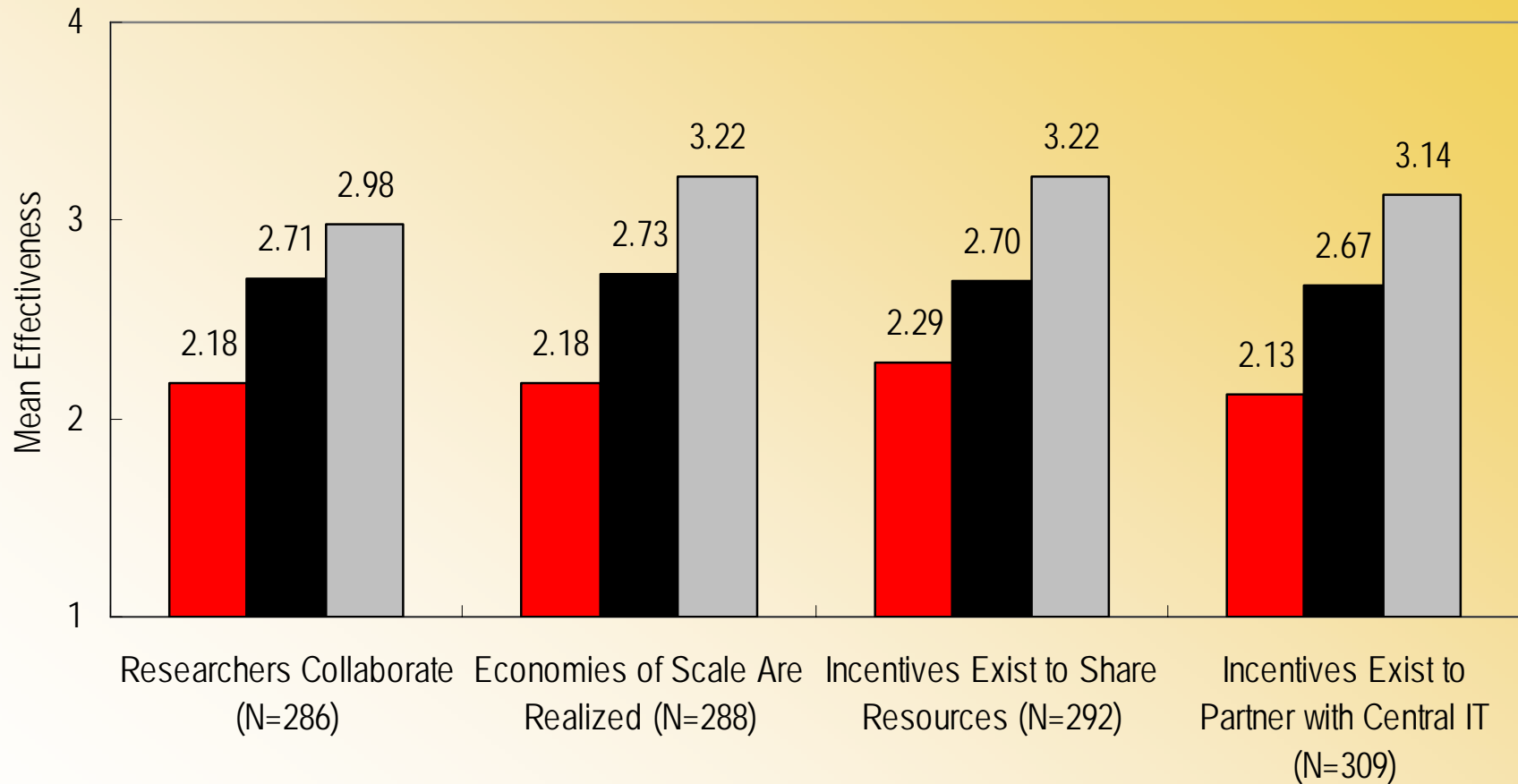
- Comparatively low interest in a share of IDCs for high-performance computing and data storage and management.
- Comparatively high interest in communication and outreach between central IT and researchers in the contexts of data storage and management, advanced networking, and collaboration tools.
- Where others were very interested in increased infrastructure funding in the data storage and management context, CLAC institutions were evenly split.

# Agreement About Collaborative Practices (Overall/CLAC)

Statement	N	Mean Agreement	Std. Deviation
Researchers generally collaborate in the use of CI resources	304	<b>3.03</b> (3.25)	0.947
Institution realizes economies of scale in the use of CI resources	306	<b>2.78</b> (2.81)	0.978
Incentives exist for researchers to share CI resources with other campus researchers	310	<b>2.69</b> (3.18)	1.011
Incentives exist for researchers to partner with central IT for economies of scale	326	<b>2.93</b> (3.24)	1.092

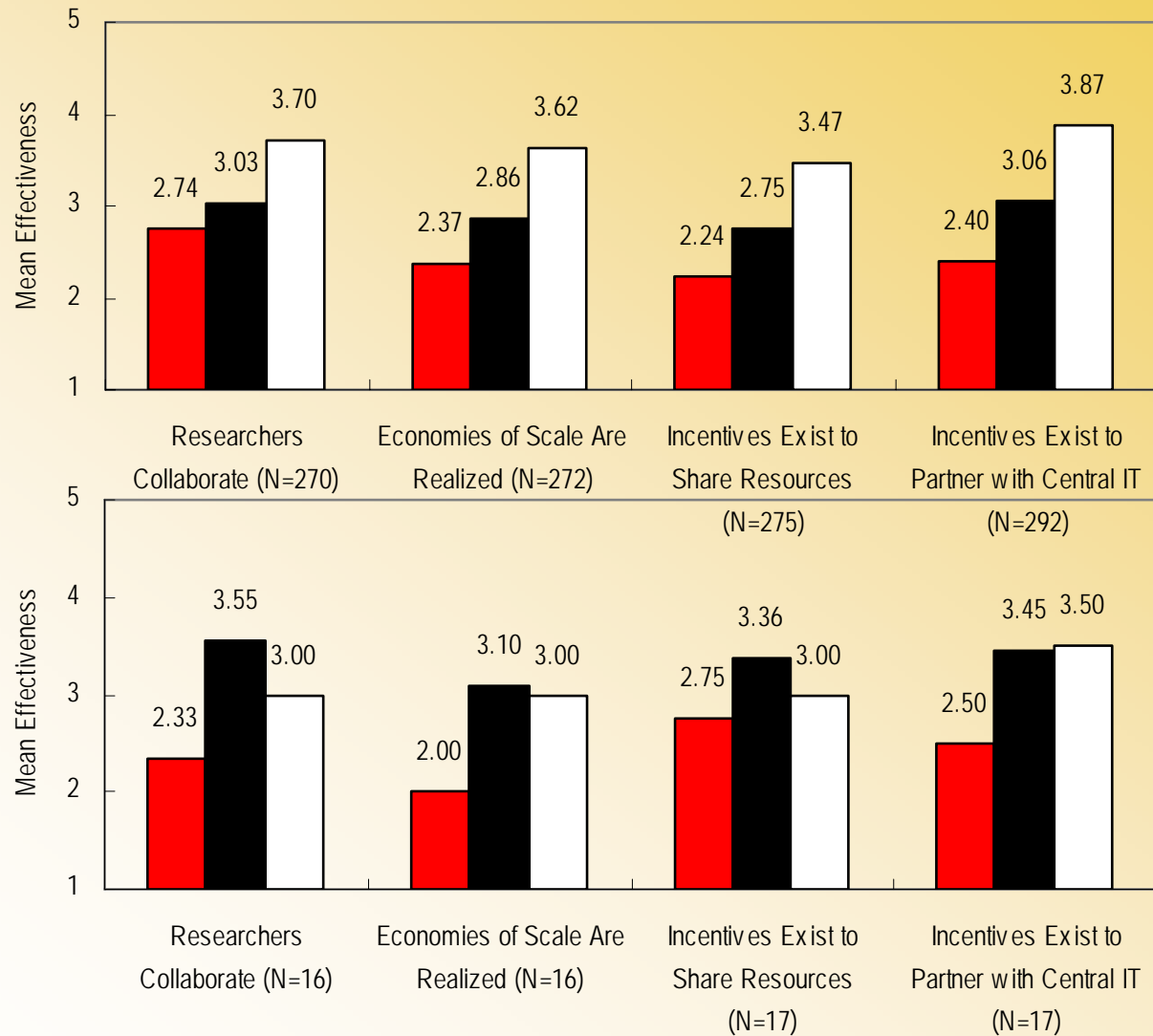


# Effectiveness of Integration of CI Technologies, by Agreement About Collaborative Practices



■ Strongly Disagree/Disagree 
 ■ Neutral 
 ■ Agree/Strongly Agree

# CLAC and Others



■ Strongly Disagree/Disagree 
 ■ Neutral 
  Agree/Strongly Agree

# Recommendations

## RECOMMENDATIONS

Based on its findings in *Higher Education IT and Cyberinfrastructure: Integrating Technologies for Scholarship*, ECAR offers a trio of interrelated recommendations to foster a collaborative and seamless Institutional CI environment:

### 1. Engage the university research leaders and community.

The CIO position is increasingly one that requires outreach, diplomacy, and institutional-level leadership. The CI context seems to be no exception if an institution has as its goal the integration of CI technologies to provide seamless support for research. Our respondent population, however, gave lackluster marks to the collaborative tendencies of researchers, to institutional incentives for collaborative practices, and to their own effectiveness in integrating CI resources. It is clear from our data that CIOs need to exert influence far beyond their own organizations to build an institutional CI platform. As CIOs consider which alliances to build in pursuit of this goal, science and engineering would likely be the best disciplines to start in because of CI's importance to this community and their deans' greater knowledge about key CI resources.

### 2. Encourage open sharing of information about CI technologies.

An important component of the CIO's engagement should be information sharing. Our research suggests that a more open environment, one in which information about the use of CI technologies is readily shared between the research community and the CIO, is one in which those technologies are used more efficiently and are integrated more effectively. An inventory of CI resources facilitates information sharing, and where the inventory is more complete, the CIO's knowledge is significantly better. Respondents indicated that increased communication and outreach between researchers and central IT would be nearly as beneficial as increased funding for central IT infrastructure and services.

### 3. Increase central IT involvement in research grants and contracts budgeting.

Funding, too, seems to affect many aspects of CI integration. Respondents cited increased funding for central IT infrastructure and services and a greater role for central IT in developing budgets for the institution's research grants and contracts as potentially beneficial in their support for effective research use at their institutions. Increased central IT involvement in grant and contract budget development promotes the goals of greater engagement and more open information sharing outlined above.

## Finding Synergies between Central IT and Researchers

One of the goals of CI is to create a rich research IT environment that is highly available to researchers across the institution. As we have seen, central IT has a role in providing and funding such resources and services, but the universe of need and the pace of technological change are well beyond the capacities of even the best-funded central IT organization. Achieving economies of scale in the use of specialized research IT resources depends in part on researchers themselves offering some of the capacity of those resources to their colleagues under a variety of collaborative scenarios, and it relies on them as well to work with central IT where appropriate to ensure optimal management of shared or shareable resources.

Nevertheless, our respondents, on average, gave their institutions' researchers mediocre scores for collaboration and gave their institutions somewhat poorer scores for achieving economies of scale and for incenting collaborative behavior. We discovered, however, that some institutions are doing better than others. For example, we found respondents more positive about collaborative practices and incentives to collaboration at institutions where central IT

provides and funds various CI technologies and where the CIO is more knowledgeable about them.

These same factors—central IT as provider and funding source for CI technologies, level of knowledge of the CIO and other officers about CI technologies, and the status of CI technology inventories—appear in our findings in relation to respondents' self-assessment of central IT's effectiveness at integrating CI resources for support of research, suggesting that what helps collaboration grow is also helpful in central IT's CI integration efforts. Indeed, we found that, on average, institutions giving higher marks to researchers for collaboration and to the institution for realizing economies of scale, and those agreeing more strongly that incentives for those practices are in place, reported that their central IT organizations were significantly more effective at integrating CI technologies to provide seamless support for research.

## Endnote

1. Daniel E. Atkins, et al., *Revolutionizing Science and Engineering through Cyberinfrastructure* (Washington, D. C.: NSF, 2003). <http://www.nsf.gov/od/ocv/reports/atkins.pdf>



# Recommendations

- Engage university research leaders and community
  - Science & engineering deans and CRO are most knowledgeable
  - CAO may be most influential
  - Direct CIO engagement
    - Knowledge: the more detailed, the better

# Recommendations

- Encourage open sharing of information about CI technologies
  - Build CIO's knowledge
  - Build other executives' knowledge
  - Build partnerships between central IT and researchers
  - Encourage collaboration among researchers

# Recommendations

- Increase central IT involvement in research grants and contracts budgeting
  - Gain buy-in of CIO
  - Ensure sufficient fund base for support activities
  - Expand CIO's detailed knowledge of CI technology use
  - Optimize equipment selection to facilitate sharing and economies of scale

Mark Sheehan  
msheehan@educause.edu

# Panel presentation

These slides were presented during the panel discussion about cyberinfrastructure that followed the keynote address.

# *EDUCAUSE Review*

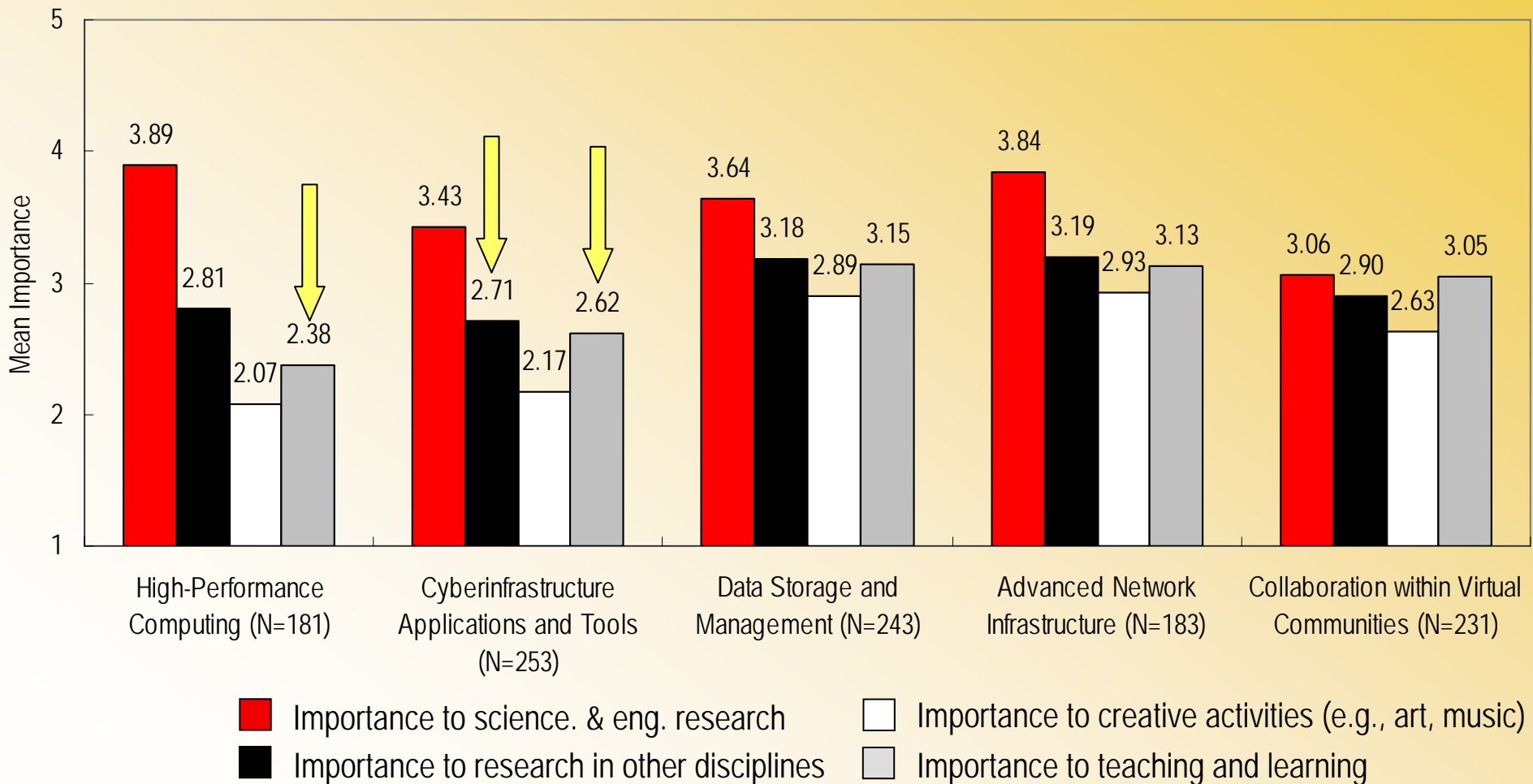
July/August 2008—Cyberinfrastructure issue

## **Cyberinfrastructure: Changing a Cottage Industry**

Mark C. Sheehan

Drawn from a recent ECAR research study, this article addresses the importance of five CI technologies to various academic areas in research and in teaching and learning at present and how survey respondents think the importance of these technologies might change in the near future.

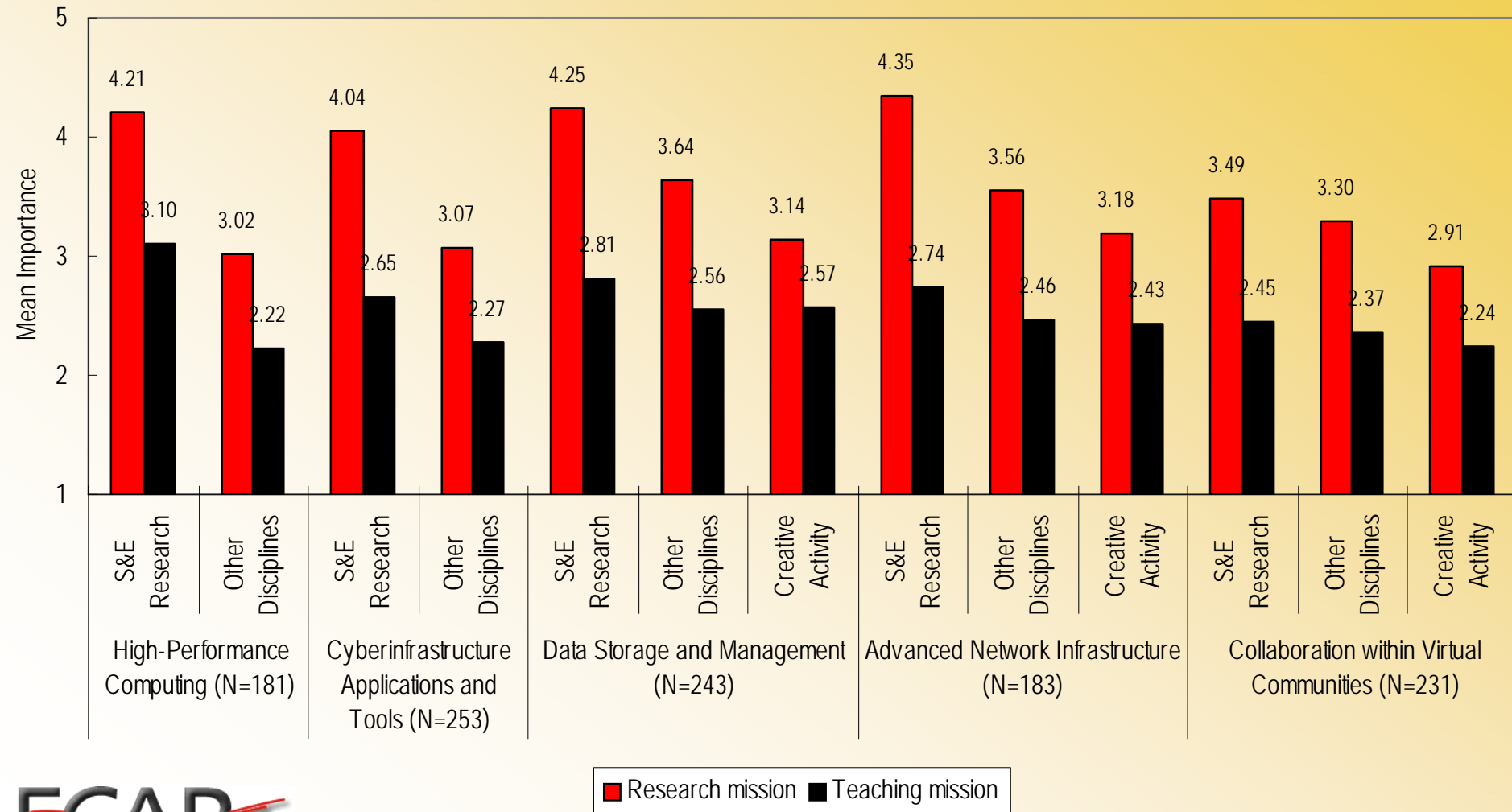
# Importance of CI Technologies



# CLAC's View

- CLAC respondents saw comparatively less importance for HPC in:
  - Science & Engineering Research
  - Research in Other Disciplines
  - Creative Activity

# Importance of CI Technologies



**Table 1. Respondents Anticipating Moderate or Great Increase in Importance of Cyberinfrastructure Technologies (N=364)** 71

	Research Institutions		Teaching Institutions	
<b>Importance to Research</b>		Percentage of Institutions		Percentage of Institutions
	Data Storage and Management	92.8%	Data Storage and Management	35.5%
	Advanced Network Infrastructure	88.4%	Collaboration within Virtual Communities	25.8%
	Cyberinfrastructure Applications and Tools	87.0%	Advanced Network Infrastructure	23.9%
	Collaboration within Virtual Communities	85.0%	Cyberinfrastructure Applications and Tools	18.4%
	High-Performance Computing	82.6%	High-Performance Computing	16.9%
<b>Importance to Teaching</b>		Percentage of Institutions		Percentage of Institutions
	Data Storage and Management	80.6%	Data Storage and Management	51.8%
	Collaboration within Virtual Communities	75.7%	Collaboration within Virtual Communities	45.5%
	Advanced Network Infrastructure	71.3%	High-Performance Computing	31.5%
	Cyberinfrastructure Applications and Tools	59.0%	Cyberinfrastructure Applications and Tools	31.5%
	High-Performance Computing	48.1%	Advanced Network Infrastructure	31.1%

