Collaborative Research Across Boundaries: A Report on Results of the KDI Initiative

Jonathon Cummings
Sloan School of Management, MIT

Sara Kiesler
Human-Computer Interaction Institute, CMU

February 15, 2004

This research was supported by
NSF Award #IIS-9872996

KDI Initiative

- Two-year innovative NSF research program whose goal was to foster multidisciplinary research including computer and information sciences as well as the other sciences

- Highly competitive; 40 / 697 (6%) proposals funded in 1998, and 31 / 554 (6%) pre-proposals (163 full proposals) funded in 1999

- http://www.researchstudy.org/kdi

KDI Projects (N=71)

- Projects had a principal investigator (PI) and up to five co-principal investigators (co-PIs) (average was three), who all held PhD's in their field of study (N=291 PIs/co-PIs)

- Scientists resided in over 100 universities and represented close to 40 disciplines, e.g., computer science (16%), electrical engineering (13%), other engineering disciplines (12%), psychology (12%), physics (9%), mathematics (9%)

Evaluating the KDI Initiative

- Workshop held in Spring of 2002 to discuss KDI research projects, identify their problems and successes, and generate metrics for evaluating projects

- Post-workshop survey conducted in Fall of 2002 to assess projects using metrics generated in the workshop

Overview of Data Collection

Step 1 – Elicited proposal, participant names (e.g., post-docs, graduate students, staff), CVs, and annual reports from PIs (62/71; 87% responded)

Step 2 – Workshop held for PIs and co-PIs in the Spring of 2002 to better understand how projects were organized (52/71; 73% projects represented)

Step 3 – Survey administered to PIs and co-PIs in the Fall of 2002 to assess coordination and project outcomes (62/71; 87% projects represented)

Workshop Lessons

- Investigators manage and leverage funded research projects in creative, flexible ways
- Examples of successes are not easily quantifiable and should be elicited in concrete terms other than publications
- Many project outcomes are realized in the long term rather than the time span of grant

Survey Sample

- N=62 projects included in sample where PI or co-PI completed questions about coordination and innovation in survey
- Data from only one respondent (PI given priority) used for each project in analyses (breakdown of surveys used in a few slides)
- N=26 co-located projects (PIs and co-PIs at same university) vs N=36 geographically dispersed projects (PIs and co-PIs at two or more universities)

Coordination Measures

(a) Faculty supervised tasks / studies (84%)

(b) Post-doc supervised tasks / studies (44%)

(c) Grad student supervised tasks / studies (34%)

(d) Held conference or workshop (55%)

(e) Held seminar or invited outside speakers (60%)

(f) At least monthly face-to-face meetings (55%)

(g) At least monthly phone calls or email (84%)

(h) Worked during conference or workshop (52%)

(i) Worked during sabbatical leave (21%)

(j) Flew on airplane to another site to work (52%)

(% yes)

Outcome Measures

(a) Started new field or research area (58%)

(b) Came up with new spinoff projects (58%)

(c) Developed new methodologies (66%)

(d) Recognized with award (19%)

(e) Created new software (71%)

(f) Created new hardware (13%)

(g) Generated new datasets (47%)

(h) Submitted patent application (15%)

(% yes)

People Measures

(a) Student finished thesis/dissertation (76%)

(b) Student/post-doc got academic job (48%)

(c) Student/post-doc got industry job (42%)

(d) Formed partnerships with industry (27%)

(e) Formed community relationships through research (27%)

(f) Formed collaborations with different researchers (65%)

(% yes)

Outcome (Factor Analysis)

(0 : no – 1 : yes)	Ideas	Tools
(a) Started new field or research area	0.66	-0.01
(b) Came up with new spinoff projects	0.74	0.08
(c) Developed new methodologies	0.46	-0.02
(d) Recognized with award	0.72	0.07
(e) Created new software	-0.14	0.68
(f) Created new hardware	0.05	0.61
(g) Generated new datasets	0.15	0.56
(h) Submitted patent application	0.00	0.72
(items were standardized before combining into innovation measures)	α=.55	α=.51

($r_{Ideas/Tools}$ = .06)

Outcome Examples (Ideas)

- Algorithm for large scale predictive species distribution
- 3d optical and magnetometer signals for speech
- Combined game logic and algorithms
- Blood flow simulation around prosthetic heart valves
- Multi-electrode recording MEG and fMRI
- Structure-based vehicle detection/tracking
- Use of eye movements to study language production
- Digital language archives
- Brain machine interfaces
- Performance analysis of comparative gene finders
- Digital technologies for archaeology
- Application of conservation laws to interface motion
- Ecological informatics

Outcome Examples (Tools)

- System to support manual manipulation of virtual objects
- Web interface for analyzing and mapping species
- Speech data collection and analysis program
- Hardware technology for tracking facial expressions
- Web-based tutoring system for cognitive science
- Dynamic lead time production scheduling software
- Developed program to calculate fluid-structure interaction
- Software environment for polycrystal sample generation
- Code for analysis of multielectrode data
- Language acquisition by autonomous robot program
- Software for relativistic astrophysics simulation
- Database of special functions of applied mathematics
- Open source resource for others doing surgical simulation

People (Factor Analysis)

(0 : no – 1 : yes)	Training	Outreach
(a) Student finished thesis/dissertation	0.84	-0.03
(b) Student/post-doc got academic job	0.59	0.00
(c) Student/post-doc got industry job	0.68	0.15
(e) Formed partnerships with industry	0.34	0.56
(f) Formed community relationships through research	-0.23	0.68
(g) Formed collaborations with different researchers	0.10	0.65
(items were standardized before combining into people measures)	α=.55	α=.28

($r_{Training/Outreach}$ = .24)

People Examples (Training)

- Grad finished thesis on learning in a flight training setting
- Undergrad thesis published in top journal
- KDI research incorporated in doctoral dissertation
- Student moved to aviation research unit of major company
- Two masters theses completed
- Post-doc got assistant professor job
- Completed Ph.D. thesis on topic under my supervision
- Ph.D. student defended thesis on modeling traffic dynamics
- Got an academic job teaching linguistics at Northwestern
- Grad assistant received a prestigious post-doc at Yale
- Student got a job at Rand Corporation
- One of our researchers got an academic appointment
- First ever Stanford BS honors thesis in MS&E

People Examples (Outreach)

- Museum community developed better access to software
- New collaborations with researchers at CONABIO Mexico
- Partnered w/ Sun Microsystems to obtain computer system
- Worked with lawyers at the Courtroom 21 project
- Collaboration with IBM Watson Research Center
- Formed alliance and transferred technology developed
- Met researchers via seminar series
- One spinoff is a partnership from Pfizer
- Strengthened relationships with government community
- Project software has been deployed in the IT industry
- Supervised talented high school students
- Formed close ties to IBM and Intel
- New community formed around project results

Surveys Used (N=62)

	co-located collaborations	dispersed collaborations	
co-PI completed survey	n = 12 ($r_{co\text{-}PI/co\text{-}PI}$ = .35/.34/.36)	n = 13 ($r_{co\text{-}PI/co\text{-}PI}$ = .45/.46/.42)	(n=25)
PI completed survey	n = 14 ($r_{PI/co\text{-}PI}$ = .42/.44/.38)	n = 23 ($r_{co\text{-}PI/co\text{-}PI}$ = .46/.46/.47)	(n=37)
	(n=26)	(n=36)	

(within-project correlation for overall/ideas/tools)

Control Variables (N=62)

(a) Year Started (M = 1998.47, SD = 0.50)

(b) Budget (M = $1,525,315, SD = $615,476)

(c) Univ R&D Exp (M = $258M, SD = $167M)

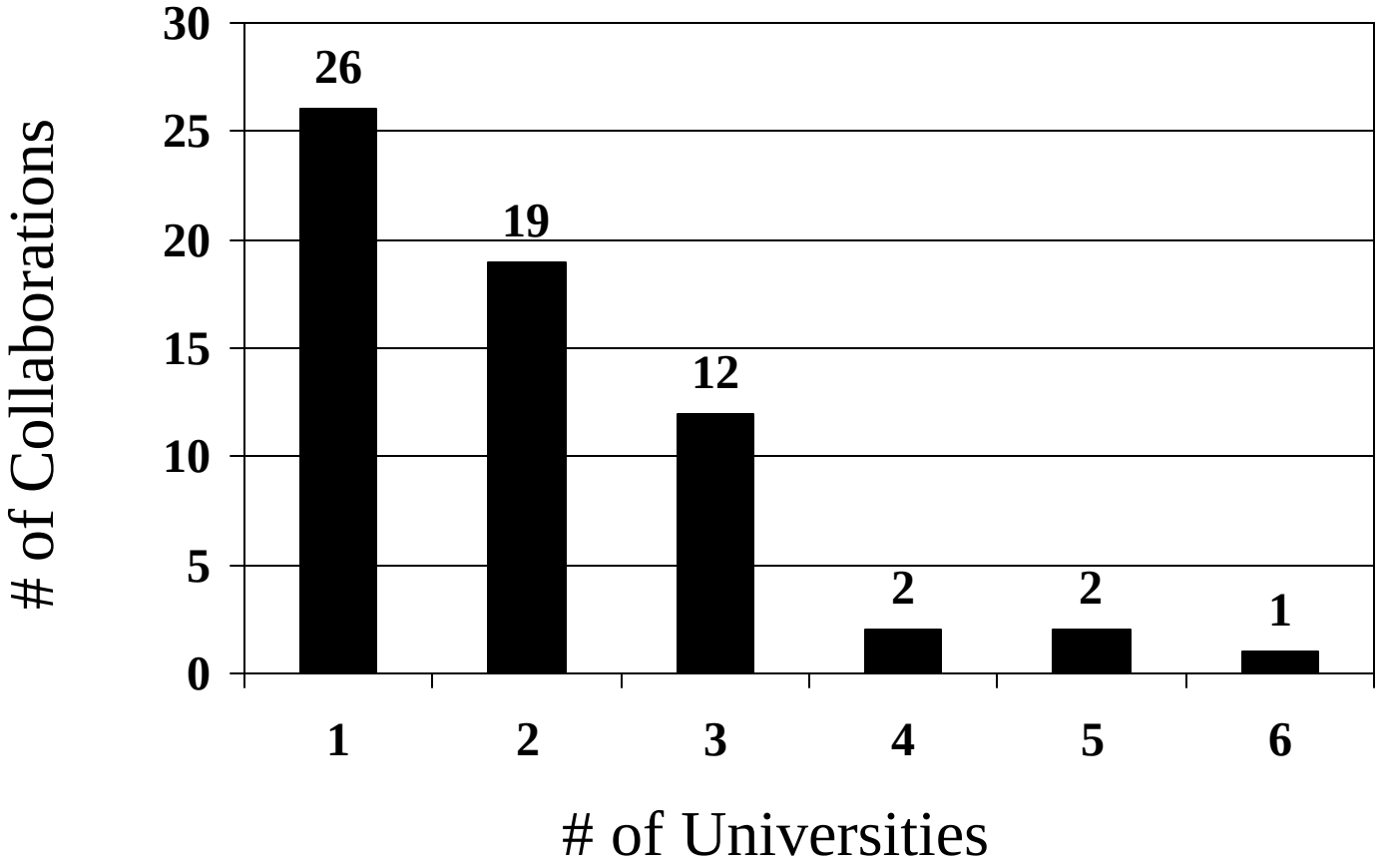
(d) # of PIs (M = 3.87, SD = 1.19)

(e) # of Post-docs (M = 0.94, SD = 2.08)

(f) # of Grad students (M = 2.66, SD = 4.05)

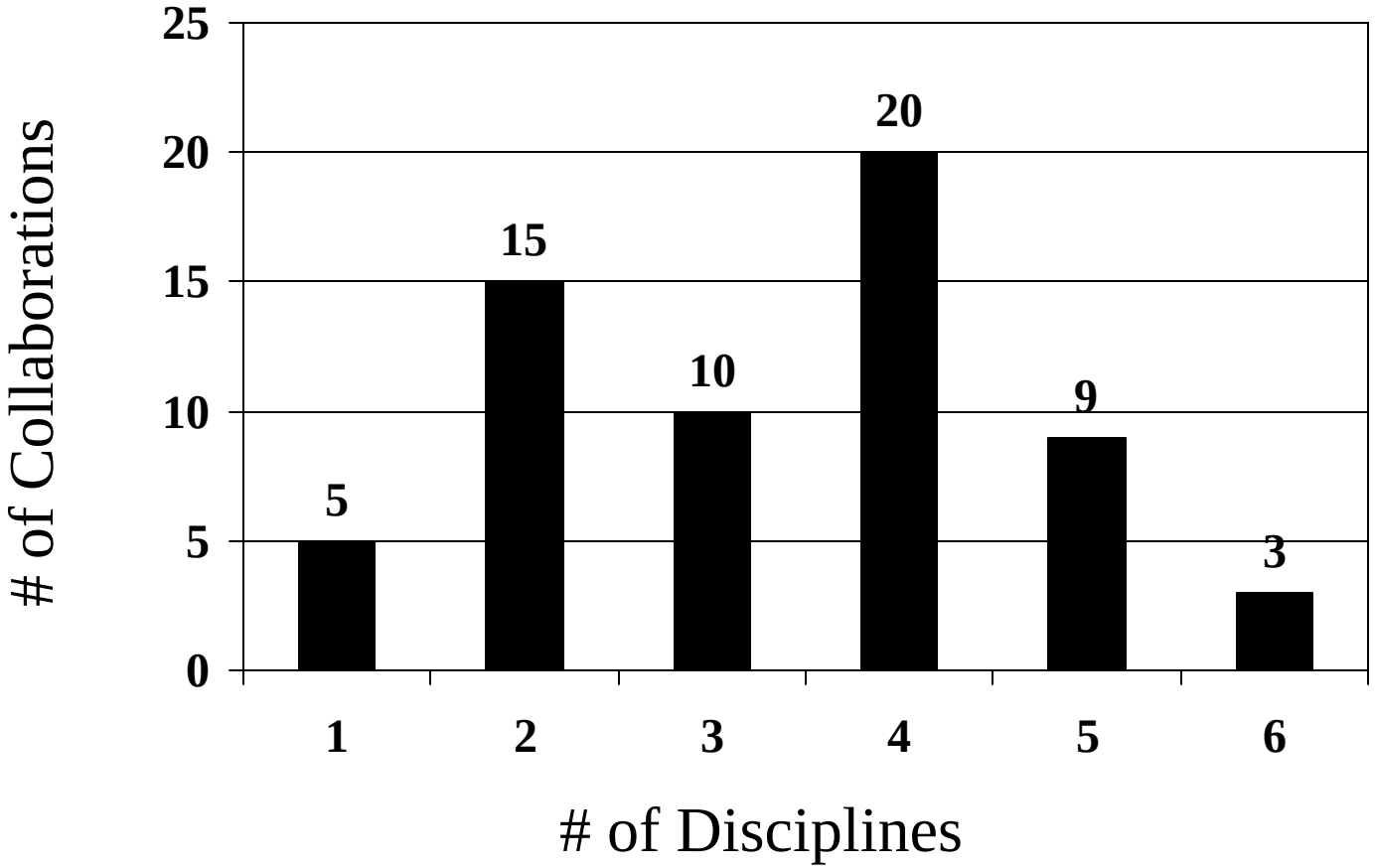
Geographic Dispersion (N=62)

of Universities (PIs) (M = 2.00, SD = 1.15)



Multi-Disciplinary (N=62)

of Disciplines (PIs) (M = 3.35, SD = 1.34)



Regression Analyses

Model 1 – geographic dispersion, multidisciplinary, and outcome measures were entered as DVs and *controls* were entered as IVs

Model 2 – each *coordination* mechanism was entered separately as a DV and geographic dispersion and multidisciplinary were entered as IVs (with controls)

Model 3 – *outcome* measures were entered as DVs and each coordination mechanism was entered as an IV (with controls, geographic dispersion, and an interaction term of coordination*dispersion)

Model 1 – Controls

	DV = Dispersion	**DV = Disciplines**
IV = Year Started	-0.18	0.19
IV = Budget	**0.27*t***	0.12
IV = Univ R&D Exp	-0.15	0.13
IV = # of PIs	-0.02	**0.35****
IV = # of Post-docs	0.07	0.10
IV = # of Grad students	-0.08	-0.04
IV = # of Disciplines (PIs)	**0.35***	**--**
IV = # of Universities (PIs)	**--**	**0.27***

Note: OLS standardized estimates, ^{t}p<.10,*p<.05,**p<.01, **N=62**

Model 1 – Controls

	DV = Ideas	**DV = Tools**	**DV = Training**	**DV = Outreach**
IV = Year Started	0.12	-0.09	-0.06	0.01
IV = Budget	-0.02	0.16	-0.03	-0.08
IV = Univ R&D Exp	-0.03	-0.07	0.07	-0.06
IV = # of PIs	-0.07	-0.05	**-0.26*t***	0.07
IV = # of Post-docs	0.20	**-0.32*t***	0.06	0.09
IV = # of Grad students	-0.09	0.17	-0.08	-0.03
IV = # of Disciplines (PIs)	-0.04	0.08	-0.16	0.22
IV = # of Universities (PIs)	**-0.38****	0.03	-0.18	-0.16

Note: OLS standardized estimates, ^{t}p<.10,*p<.05,**p<.01, **N=62**

Model 2 – Coordination

(note that DV switched to row)	**IV = Dispersion**	**IV = Disciplines**
DV = Faculty supervised tasks / studies	**-0.29**^{t}	-0.03
DV = Post-doc supervised tasks / studies	**-0.27**^{t}	0.12
DV = Grad student supervised tasks / studies	**-0.34***	-0.06
DV = Held conference or workshop	**0.32***	-0.18
DV = Held seminar or invited outside speakers	**-0.49****	-0.01
DV = At least monthly face-to-face meetings	**-0.47****	-0.07
DV = At least monthly phone calls or email	0.13	-0.05
DV = Worked during conference or workshop	**0.26**^{t}	0.07
DV = Worked during sabbatical leave	0.05	-0.03
DV = Flew on airplane to another site to work	0.15	0.15

Note: OLS standardized estimates, ^{t}p<.10,*p<.05,**p<.01, **N=62**

Model 3 – Outcome

(significant interaction of dispersion*coordination)	**DV = Ideas**	**DV = Tools**
IV = Geographic Dispersion	**-.42****	-.08
IV = Faculty supervised tasks / studies	**.33***	**.27^{t} (-.34*)**
IV = Post-doc supervised tasks / studies	.06	-.04 **(-.43*)**
IV = Grad student supervised tasks / studies	**.32***	**.28^{t}**
IV = Held conference or workshop	**.36***	-.17
IV = Held seminar or invited outside speakers	-.02	.07
IV = At least monthly face-to-face meetings	-.12	.62
IV = At least monthly phone calls or email	**.31* (.22^{t})**	.14
IV = Worked during conference or workshop	**.45****	-.05
IV = Worked during sabbatical leave	-.04	**.34* (-.27^{t})**
IV = Flew on airplane to another site to work	**.32***	.04

Note: OLS standardized estimates, ^{t}p<.10,*p<.05,**p<.01, **N=62**

Model 3 – People

	DV = Training	DV = Outreach
IV = Geographic Dispersion	-.18	-.16
IV = Faculty supervised tasks / studies	**.31***	**.30***
IV = Post-doc supervised tasks / studies	-.06	.01
IV = Grad student supervised tasks / studies	**.34***	**.37***
IV = Held conference or workshop	-.15	-.11
IV = Held seminar or invited outside speakers	**.42****	.02
IV = At least monthly face-to-face meetings	**.24*t***	.18
IV = At least monthly phone calls or email	**.22*t***	**.25*t***
IV = Worked during conference or workshop	.04	.04
IV = Worked during sabbatical leave	.13	.12
IV = Flew on airplane to another site to work	.11	.14

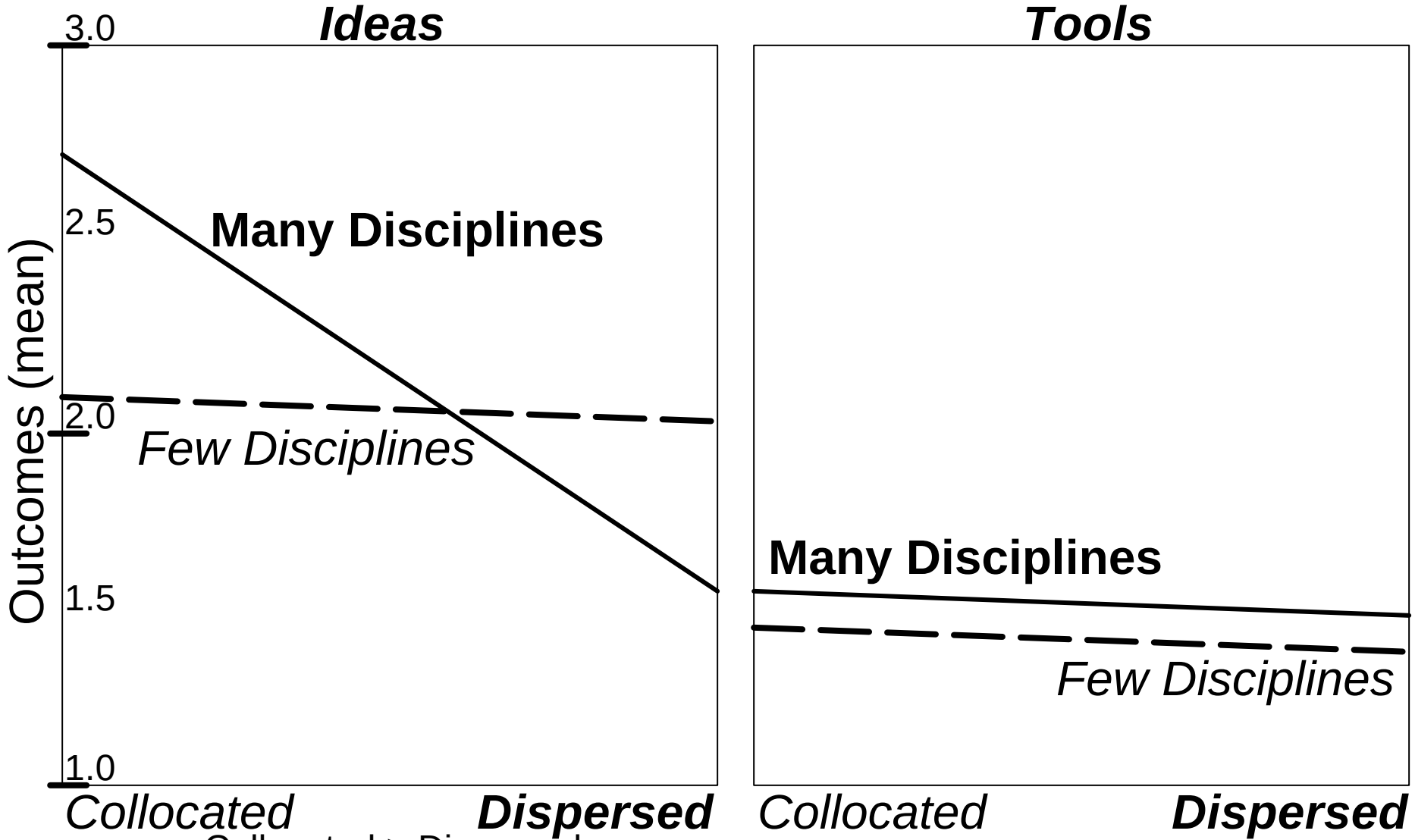
Note: OLS standardized estimates, ^{t}p<.10,*p<.05,**p<.01, **N=62**

Dispersion X Disciplines (N=62)

(marginal means adjusted for controls)

ANOVA	co-located collaborations	dispersed collaborations	
1-3 disciplines	Ideas=2.10 Tool=1.46 Training=2.18[a] Outreach=0.73 (n = 16)	Ideas=2.06 Tool=1.23 Training=1.71 Outreach=0.71[b] (n = 14)	(n=30)
4-6 disciplines	Ideas=2.72[c] Tool=1.63 Training=1.69 Outreach=1.43[b] (n = 10)	Ideas=1.61[c] Tool=1.50 Training=1.25[a] Outreach=1.01 (n = 22)	(n=32)
(unique letters indicate significant differences)	(n=26)	(n=36)	

Dispersion X Disciplines (N=62)

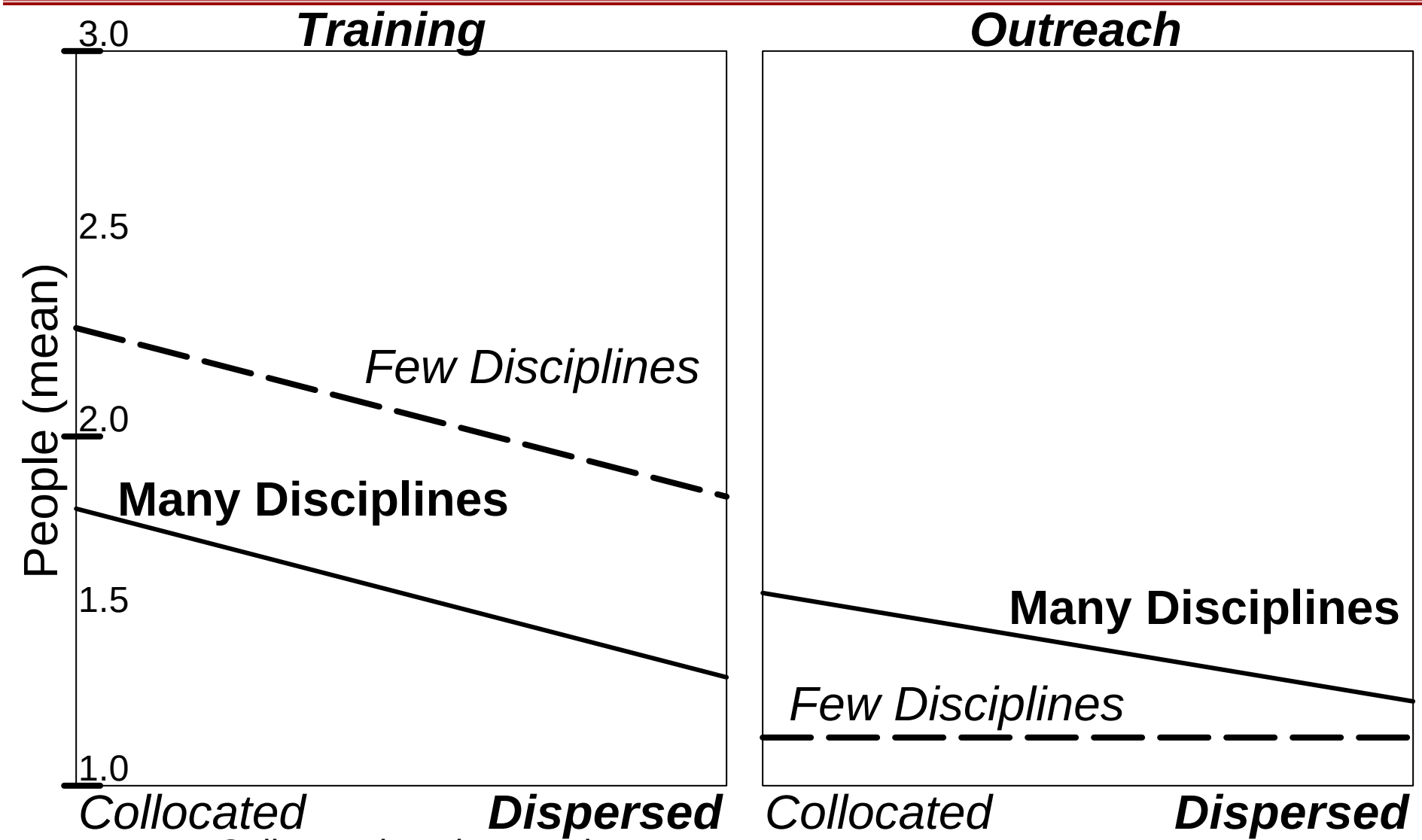


Collocated > Dispersed

Dispersed * Multidisciplinary Lower

No significant differences

Dispersion X Disciplines (N=62)



Collocated > Dispersed

Disciplinary > Multidisciplinary

Multidisciplinary > Disciplinary

Summary

- Dispersed scientific collaborations are significantly
  - *less* likely to have faculty, post-docs, and grad students supervise tasks / studies, hold seminars or invite outside speakers, and have monthly face-to-face meetings
  - *more* likely to hold a conference or workshop, and work on the project while attending a conference or workshop

Summary

- Outcome (ideas) is significantly
  - *negatively* associated with geographic dispersion
  - *positively* associated with having faculty and grad students supervise tasks / studies, holding a conference or workshop, having monthly phone calls or email, working during conference or workshop, and flying on airplane to another site to work on project
- Outcome (tools) is significantly
  - *positively* associated with having faculty and grad students supervise tasks / studies, and working during sabbatical leave

Summary

- People (training) is significantly
  - *positively* associated with having faculty and grad students supervise tasks / studies, having monthly face-to-face meetings, having monthly phone calls or email, and working during conference or workshop
- People (outreach) is significantly
  - *positively* associated with having faculty and grad students supervise tasks / studies, and having monthly phone calls or email

Alternative Explanations

- Selection bias such that dispersed projects may have been given an unfair advantage for being dispersed, thus the collocated projects were better to begin with

- Idiosyncratic features of the KDI Initiative, such as funding for three years; thus dispersed projects may not have been able to get sufficient resources for coordination

Implications for NSF

- Multidisciplinary projects are also likely to be geographically dispersed, a factor which cannot be ignored
- Dispersed projects organize differently and require nontraditional kinds of infrastructure support for coordination
- More attention should be paid to the proposal funding process given the amount of work required to propose and carry out multidisciplinary research

Implications for Evaluation

- Information about *unfunded* proposals could help us learn more about the success of multidisciplinary research projects
- Evaluations should assess short and long term project outcomes
- Annual and final reporting is an important part of the evaluation process, and should be structured to ease the impact on researchers' time