

**BUCLD
BUSINESS
MEETING
PRESENTATION**

November 1, 2002

BACKGROUND

1. BUCLD is now partially funded by conference grants from NSF and NIH. Thus, it is appropriate to be more publicly accountable than might otherwise be the case.
2. Inasmuch as BUCLD is a resource for researchers in language development, it is important that the selection methods be fair and well known to the participants.

OUTLINE

1. Making the Process Well Known
 - review and selection process
 - general statistics
2. Asking Whether the Process is Fair
 - overall acceptance rate
 - highlighting leading work
 - proportional representation by subfield
3. Questions and Discussion

**PART 1:
MAKING THE PROCESS
WELL KNOWN**

REVIEW PROCESS

1. We send each abstract to 5 reviewers.

82 REVIEWERS FOR BUCLD 2002

Nameera Akhtar
Shanley Allen
Arto Anttila
Richard Aslin
Edith Bavin
Heike Behrens
David Birdsong
Paul Bloom
Melissa Bowerman
Cynthia Brown
Joyce Bruhn de Garavito
Nancy Budwig
Susan Carey
Harald Clahsen
Peter Coopmans
Stephen Crain
Katherine Demuth
Kenneth Drozd
Catherine Echols
Richard Ely
Anne Fernald

Cynthia Fisher
Suzanne Flynn
LouAnn Gerken
Heather Goad
Susan Goldin-Meadow
Rebecca Gomez
Peter Gordon
Paul Hagstrom
Cornelia Hamann
Catherine Harris
Marco Haverkort
Lowry Hemphill
Bart Hollebrandse
Aafke Hulk
Nina Hyams
Alan Juffs
Dorit Kaufman
Judy Kegl
Deb Kelemen
Wolfgang Klein
Donna Lardiere

Beth Levin
Jeffrey Lidz
Elena Lieven
Alec Marantz
Rachel Mayberry
Richard Meier
Lise Menn
Letitia Naigles
Carol Neidle
Elissa Newport
Cathy O'Connor
Mitsuhiko Ota
Johanne Paradis
Joseph Pater
William Philip
Colin Phillips
Clifton Pye
Marnie Reed
Mabel Rice
Tom Roeper
Lynn Santelmann

Jeannette Schaeffer
Carson Schütze
Bonnie Schwartz
Ann Senghas
Yasuhiro Shirai
Jesse Snedeker
William Snyder
Rex Sprouse
Carol Stoel-Gammon
Helen Tager-Flusberg
Margaret Thomas
Michael Tomasello
Anne Vainikka
Angeliek van Hout
Jürgen Weissenborn
Lydia White
Frank Wijnen
Fei Xu
Andrea Zukowski

ASSIGNMENT OF ABSTRACTS TO REVIEWERS

Reviewers are individually assigned to abstracts by the faculty advisors, with the following conditions:

1. Ensure that the reviewer is sufficiently familiar with the content of the abstract.
2. Ensure that the reviewer is not unfriendly to the theoretical perspective of the abstract.
3. Don't assign abstracts to reviewers who are colleagues, students, advisors, close friends, or enemies of the authors (insofar as we know this).
4. Each reviewer gets a minimum of 10 abstracts, most get between 15 and 20.

REVIEW PROCESS

1. We send each abstract to 5 reviewers.
2. Reviewers rate each abstract independently on a scale of 1-10 (double-blind procedure).

RATING SCALE

1 = must reject

3 = low priority

5 = medium priority

7 = high priority

10 = must accept

RATING GUIDELINES

1. Question/issue clearly stated?
2. Significance of work clearly stated?
3. Method/analysis well-designed and appropriate?
4. Conceptual framework coherent? Theoretical analysis well-argued?
5. Strong promise of work being completed?
6. Conclusions justified relative to data/analyses?
7. Abstract clear and well-organized?
8. Topic of scientific, methodological, theoretical importance?
9. Paper relevant to current issues in the field?
10. Paper of interest to BUCLD attendees?

REVIEW PROCESS

1. We send each abstract to 5 reviewers.
2. Reviewers rate each abstract independently on a scale of 1-10 (double-blind procedure).
3. We calculate two scores for each abstract:
 - a. Mean raw score
 - b. Mean z score

RAW SCORE

Definition: Score out of 10 from a reviewer.

Assumption: Every reviewer's use of a particular score category is equivalent.

Problem: May be misleading if a reviewer is particularly lenient or stringent in their ratings.

Z SCORE

Definition: ___ Standard score indicating how far, and in what direction, a given raw score deviates from the mean of all the raw scores assigned by a given reviewer.

Assumption: Every reviewer's use of a particular score category may NOT be equivalent. Some reviewers may be more demanding or lenient than others, or may use a restricted range.

Problem: May be misleading if a reviewer receives a set of unusually excellent or unusually terrible papers. (The z score effectively forces the ratings from a given reviewer to fit a bell curve.)

REVIEW PROCESS

1. We send each abstract to 5 reviewers.
2. Reviewers rate each abstract independently on a scale of 1-10 (double-blind procedure).
3. We calculate two scores for each abstract:
 - a. Mean raw score
 - b. Mean z score
4. We rank each abstract by raw score and z score, and calculate a composite rank.

SAMPLE ABSTRACT RANKING DATABASE

ABS#	RAW SCORES			Z-SCORES		COMPOSITE
	Actual	Mean	Rank	Mean	Rank	Rank
A	7,8,10,10,10	9	8	1.043	9	7
B	7,9,9,10,10	9	8	0.801	22	16
C	8,8,8,10,10	8.8	10	1.023	10	8
D	7,8,9,10,10	8.8	10	0.679	39	24
E	5,8,10,10,10	8.6	12	1.251	1	4
F	8,8,8,9,10	8.6	12	1.190	3	6
G	7,8,9,9,10	8.6	12	0.883	15	14
H	7,7,9,10,10	8.6	12	0.705	36	23
I	7,8,9,9,10	8.6	12	0.961	12	10
J	7,7,8,10,10	8.4	17	0.857	17	17
K	5,8,9,9,10	8.2	18	0.774	27	21
L	5,8,9,9,10	8.2	18	0.397	77	45
M	6,8,8,9,10	8.2	18	1.186	5	9
N	6,8,8,9,10	8.2	18	0.760	28	22

REVIEW PROCESS

1. We send each abstract to 5 reviewers.
2. Reviewers rate each abstract independently on a scale of 1-10 (double-blind procedure).
3. We calculate two scores for each abstract:
 - a. Mean raw score
 - b. Mean z score
4. We rank each abstract by raw score and z score, and calculate a composite rank.
5. We select 90 abstracts for the program and 12 alternates.

ABSTRACT SELECTION PROCESS

1. We select the top 60 abstracts from the raw score list, and the top 60 abstracts from the z score list. This totals 75-80 abstracts (there is a lot of overlap between the two sets).
2. We create a pool of the next 40 abstracts based on composite rank.
3. We select 10-15 abstracts from the pool to complete the program of 90 papers, based as much as possible on composite rank, with the goal of forming coherent sessions.
4. We select 12 alternate abstracts from the pool, based as much as possible on composite rank, with the goal of getting a good distribution of content areas.

SOME STATISTICS ...

ABSTRACTS SUBMITTED AND ACCEPTED

	N SUBMITTED	N ACCEPTED	% ACCEPTED
2001	298	90	30%
2002	277	90	33%

NOTE: 90 accepted papers includes only papers in the program, does not include 12 alternate papers

ABSTRACT RATINGS BY MEAN RAW SCORE

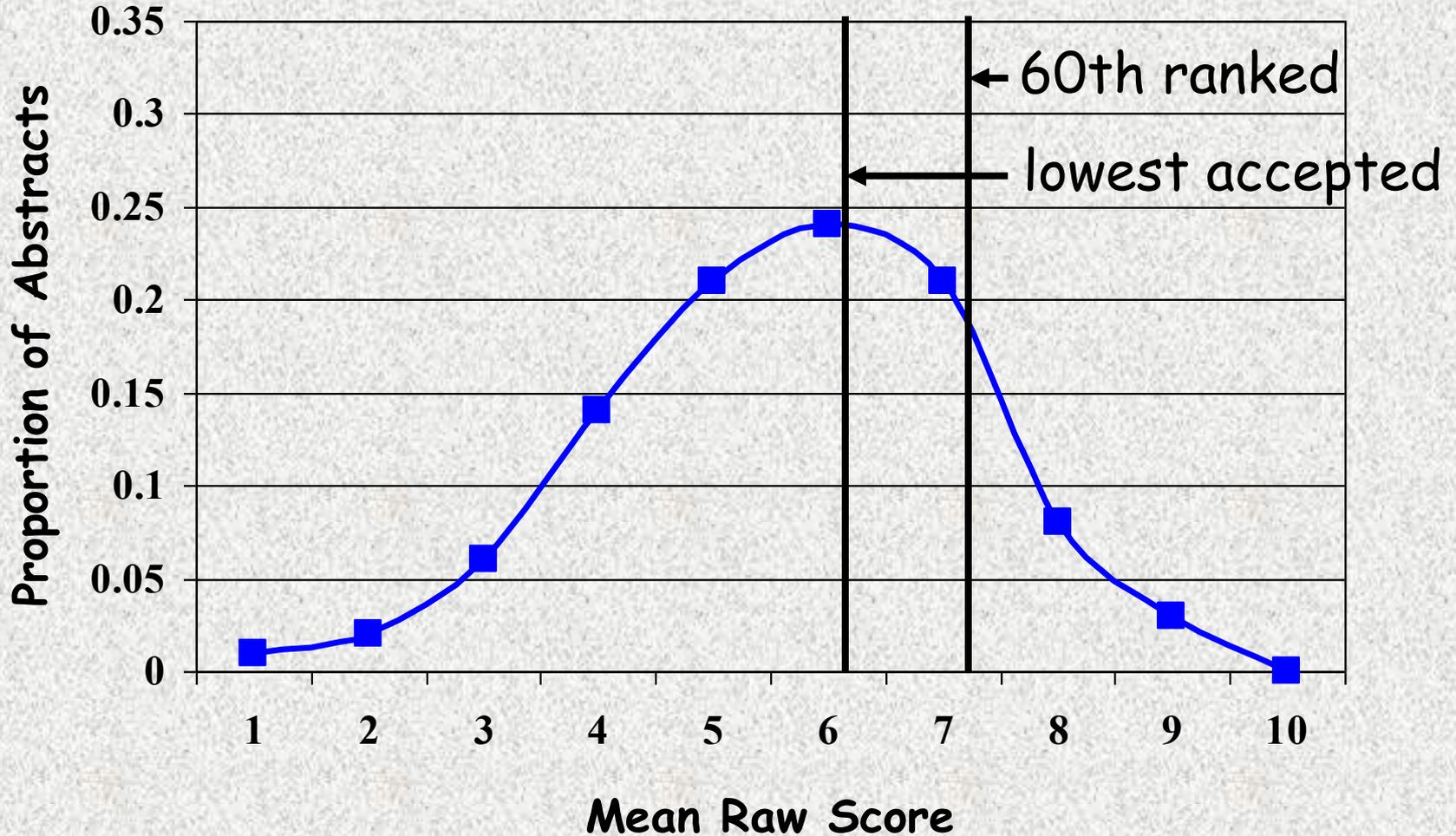
Source: Mean rating for each of the 277 abstracts
(2002 data only)

Range: 1.2 to 9.2

Mean: 6.08 (SD: 2.08)

Spread: 5.02 (average spread between the highest and
lowest of the five scores for an abstract)

DISTRIBUTION OF ABSTRACTS' MEAN RAW SCORES



ABSTRACT RATINGS BY MEAN Z SCORE

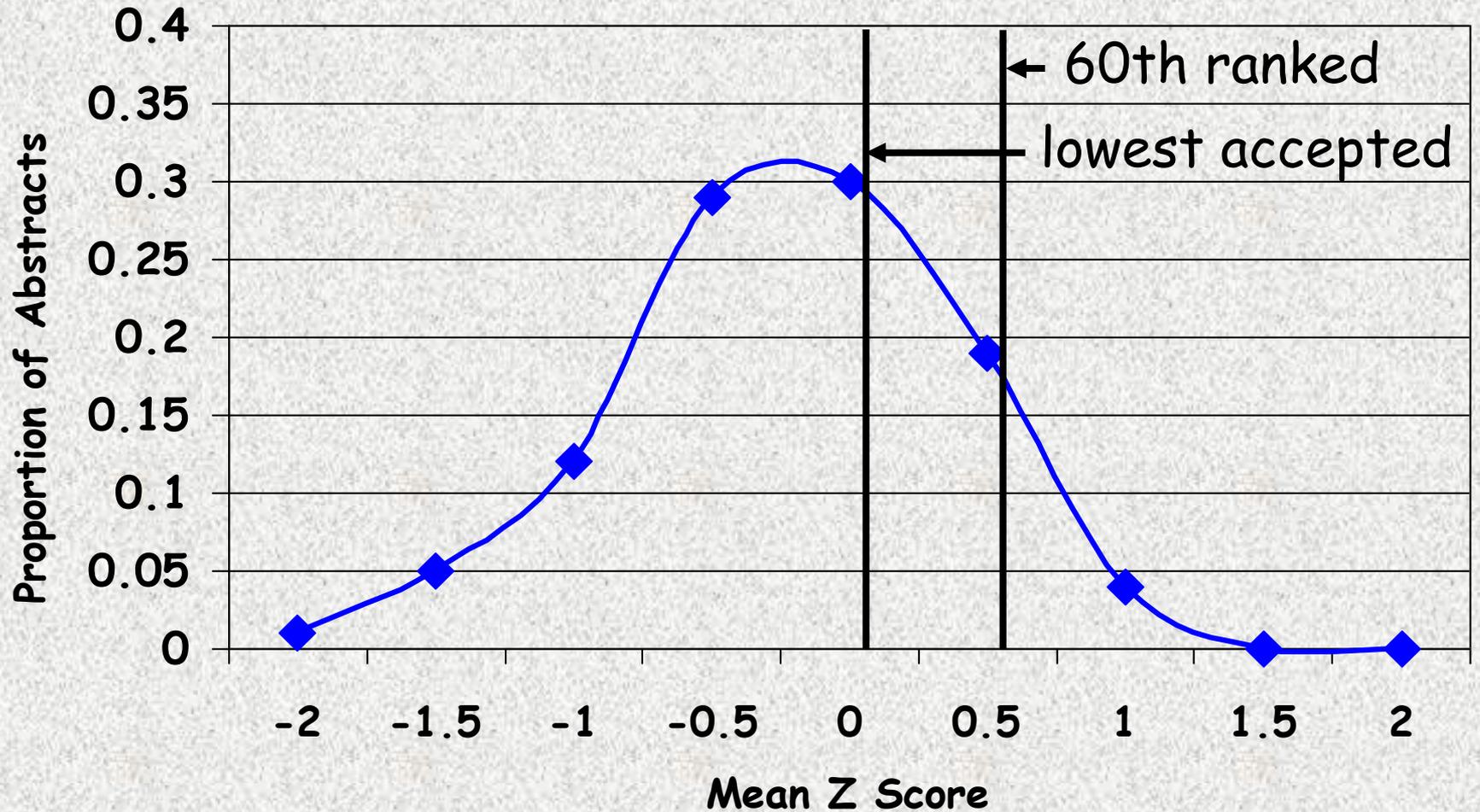
Source: Mean rating for each of the 277 abstracts
(2002 data only)

Range: -1.95 to 1.25

Mean: 0 (SD: 0.62)

Spread: 1.92 (average spread between the highest and
lowest of the five scores for an abstract)

DISTRIBUTION OF ABSTRACTS' MEAN Z-SCORES



WHY SUCH A BIG SPREAD BETWEEN THE 5 RATINGS FOR AN ABSTRACT?

	Mean Spread
Raw Score	5.0
Z Score	1.9

Presumably this shows the effect of reviewers looking for different things in an abstract.

WHY SUCH A BIG SPREAD BETWEEN THE 5 RATINGS FOR AN ABSTRACT?

	Mean Spread	Spread in Highest 60	Spread in Middle 60	Spread in Lowest 60
Raw Score	5.0	4.1	5.7	4.8
Z Score	1.9	1.8	2.0	1.9

Presumably this shows the effect of reviewers looking for different things in an abstract.

There is more variability within the middle-ranked than within the highest-ranked or lowest-ranked abstracts.

These results are consistent with moderate but not high correlation between the scores.

PART 2:
ASKING WHETHER THE
PROCESS IS FAIR

IS THE SELECTION PROCESS FAIR? DEPENDS ON THE GOAL...

Goal: Forum for presentation of the best quality work in language development broadly defined; meeting place for scholars from different perspectives to exchange ideas.

Possible structures to best serve that goal include:

1. Fully representative with maximum possible number of papers.
2. Highlight leading innovative new work and programmatic long-term research in dedicated sessions.
3. Proportional representation by content area and theoretical focus.

POSSIBLE STRUCTURE #1: MAXIMUM NUMBER OF PAPERS

ADVANTAGES

1. Many people get papers accepted.
2. The program has more selection and represents more work.

DIFFICULTIES

1. It is more difficult to attend every paper you are interested in.
2. The quality of papers may become more variable.

OPTIONS FOR RAISING THE BUCLD ACCEPTANCE RATE

1. Increase the number of parallel sessions:
 - 4 sessions instead of 3
 - would add 30 papers
 - would increase 2002 acceptance rate from 33% to 43%
2. Increase the number of days:
 - 4 days instead of 3
 - would add 36 papers
 - would increase 2002 acceptance rate from 33% to 46%
3. Add a poster session:
 - 25 posters for each of 2 days
 - would add 50 posters
 - would increase 2002 acceptance rate from 33% to 51%

POSSIBLE STRUCTURE #2: HIGHLIGHTING LEADING WORK

ADVANTAGES

1. Identify and foster leading work in the field, both newly-developing areas and long-term programmatic work.
2. A more coherent program if this work is presented in thematic sessions.

DIFFICULTIES

1. Unless these papers were added to the regular program, they would displace other abstract-based papers.
2. Who would determine what work to highlight?

ABSTRACT-BASED PROGRAM ALREADY CONTAINS LEADING WORK

LONG-TERM PROGRAMMATIC WORK

phenotype for SLI - Mabel Rice, Heather van der Lely, ...
null subjects, optional infinitives - Nina Hyams, Ken Wexler, ...
syntactic bootstrapping - Lila Gleitman, Cindy Fisher, ...
UG in L2 acquisition - Lydia White, Bonnie Schwartz, ...
infant speech perception - Peter Jusczyk, LouAnn Gerken, ...

NEW INNOVATIVE WORK

statistical modeling - Toby Mintz, Elissa Newport, ...
sentence processing - Jesse Snedeker, John Trueswell, ...
OT phonology - Heather Goad, Katherine Demuth, ...

POSSIBLE FORMATS FOR SESSIONS ON LEADING WORK

1. Abstracts solicited for parasession
(e.g. CLS, BLS)
2. Abstracts solicited for symposia in addition to papers
(e.g. IASCL, SRCO, GURT)
3. Invited symposia
(e.g. BUCLD lunch symposium)

POSSIBLE STRUCTURE #3: PROPORTIONAL REPRESENTATION?

ADVANTAGES

1. Each field gets represented in proportion to how many papers are submitted in that field.

DIFFICULTIES

1. Who determines what field a paper represents, or how the fields should be divided, especially given the increasing amount of interdisciplinary work?
2. Would poorer quality papers in a given field get accepted just to meet the predetermined quota?

PROPORTIONAL REPRESENTATION BY LEARNING CONDITION?

11 Learning Conditions Represented in 2001:

Aphasia, Attrition, Bilingual, Deaf, Down Syndrome, General Language Delay, Hearing Impaired, Specifically Language Impaired, Typical L1, Typical L2

11 Learning Conditions Represented in 2002:

Attrition, Autism, Bilingual, Deaf, Down Syndrome, General Language Delay, Hearing Impaired, Specifically Language Impaired, Typical L1, Typical L2, Williams Syndrome

STATISTICS FOR TOP 4 LEARNING CONDITIONS

	2001		2002	
	N SUB	% ACC	N SUB	% ACC
Typical L1				
Typical L2				
Bilingual				
SLI				
Overall				

STATISTICS FOR TOP 4 LEARNING CONDITIONS

	2001		2002	
	N SUB	% ACC	N SUB	% ACC
Typical L1	186	35%		
Typical L2	68	21%		
Bilingual	16	53%		
SLI	9	14%		
Overall	298	30%		

STATISTICS FOR TOP 4 LEARNING CONDITIONS

	2001		2002	
	N SUB	% ACC	N SUB	% ACC
Typical L1	186	35%	170	34%
Typical L2	68	21%	68	19%
Bilingual	16	53%	18	44%
SLI	9	14%	11	45%
Overall	298	30%	277	33%

PROPORTIONAL REPRESENTATION BY THEORETICAL FOCUS?

Included all and only abstracts on phonology, morphology, syntax, semantics, and pragmatics

Generative:

Abstract includes terminology and/or references and/or notations from Principles & Parameters, Minimalism, Optimality Theory, Formal Semantics, Formal Pragmatics

Non-Generative:

Abstract does not include terminology or references or notations from above perspectives.

STATISTICS FOR THEORETICAL FOCUS

	2001		2002	
	N SUB	% ACC	N SUB	% ACC
<i>Generative:</i>				
Morphosyntax				
Semantics / Pragmatics				
Phonology				
Total				
<i>Non-Generative:</i>				
Morphosyntax				
Semantics / Pragmatics				
Phonology				
Total				
Overall				

STATISTICS FOR THEORETICAL FOCUS

	2001		2002	
	N SUB	% ACC	N SUB	% ACC
<i>Generative:</i>				
Morphosyntax	75	17%		
Semantics / Pragmatics	13	31%		
Phonology	6	50%		
Total	94	21%		
<i>Non-Generative:</i>				
Morphosyntax	79	38%		
Semantics / Pragmatics	18	11%		
Phonology	11	27%		
Total	108	32%		
Overall	298	30%		

STATISTICS FOR THEORETICAL FOCUS

	2001		2002	
	N SUB	% ACC	N SUB	% ACC
Generative:				
Morphosyntax	75	17%	73	25%
Semantics / Pragmatics	13	31%	26	35%
Phonology	6	50%	11	27%
Total	94	21%	110	27%
Non-Generative:				
Morphosyntax	79	38%	50	30%
Semantics / Pragmatics	18	11%	10	40%
Phonology	11	27%	11	45%
Total	108	32%	71	34%
Overall	298	30%	277	33%

PROPORTIONAL REPRESENTATION BY CONTENT AREA?

Morphosyntax:	Morphology, syntax
Sound System:	Phonetics, phonology, speech perception, speech segmentation
Meaning:	Semantics, pragmatics
Words:	Word learning, structure of lexicon
Cognitive/Social:	Language & cognition, literacy & reading, narrative, discourse, sociolinguistics, child-directed speech
Models:	Sentence processing, statistical models, connectionist networks

CAVEAT FOR CONTENT CLASSIFICATION

Many abstracts can be classified as more than one content area. The analysis here is only for the one most salient area for each abstract.

STATISTICS FOR CONTENT AREAS

	2001		2002	
	N SUB	% ACC	N SUB	% ACC
Morphosyntax				
Sound System				
Meaning				
Words				
Cognitive & Social				
Models				
Overall				

STATISTICS FOR CONTENT AREAS

	2001		2002	
	N SUB	% ACC	N SUB	% ACC
Morphosyntax	154	28%		
Sound System	35	46%		
Meaning	31	19%		
Words	35	40%		
Cognitive & Social	26	27%		
Models	12	25%		
Overall	298	30%		

STATISTICS FOR CONTENT AREAS

	2001		2002	
	N SUB	% ACC	N SUB	% ACC
Morphosyntax	154	28%	123	27%
Sound System	35	46%	46	44%
Meaning	31	19%	36	36%
Words	35	40%	29	28%
Cognitive & Social	26	27%	25	36%
Models	12	25%	10	60%
Overall	298	30%	277	33%

WHY A LOWER/HIGHER ACCEPTANCE RATE? LOWER/HIGHER SCORES!

	All Abstracts	L2	Generative Morsyn.	Sound System
No. Submitted	277			
Acceptance Rate	33%			
Mean Raw Score	6.08			
Mean Z Score	0			

WHY A LOWER/HIGHER ACCEPTANCE RATE? LOWER/HIGHER SCORES!

	All Abstracts	L2	Generative Morsyn.	Sound System
No. Submitted	277	68		
Acceptance Rate	33%	19%		
Mean Raw Score	6.08	5.54		
Mean Z Score	0	-0.14		

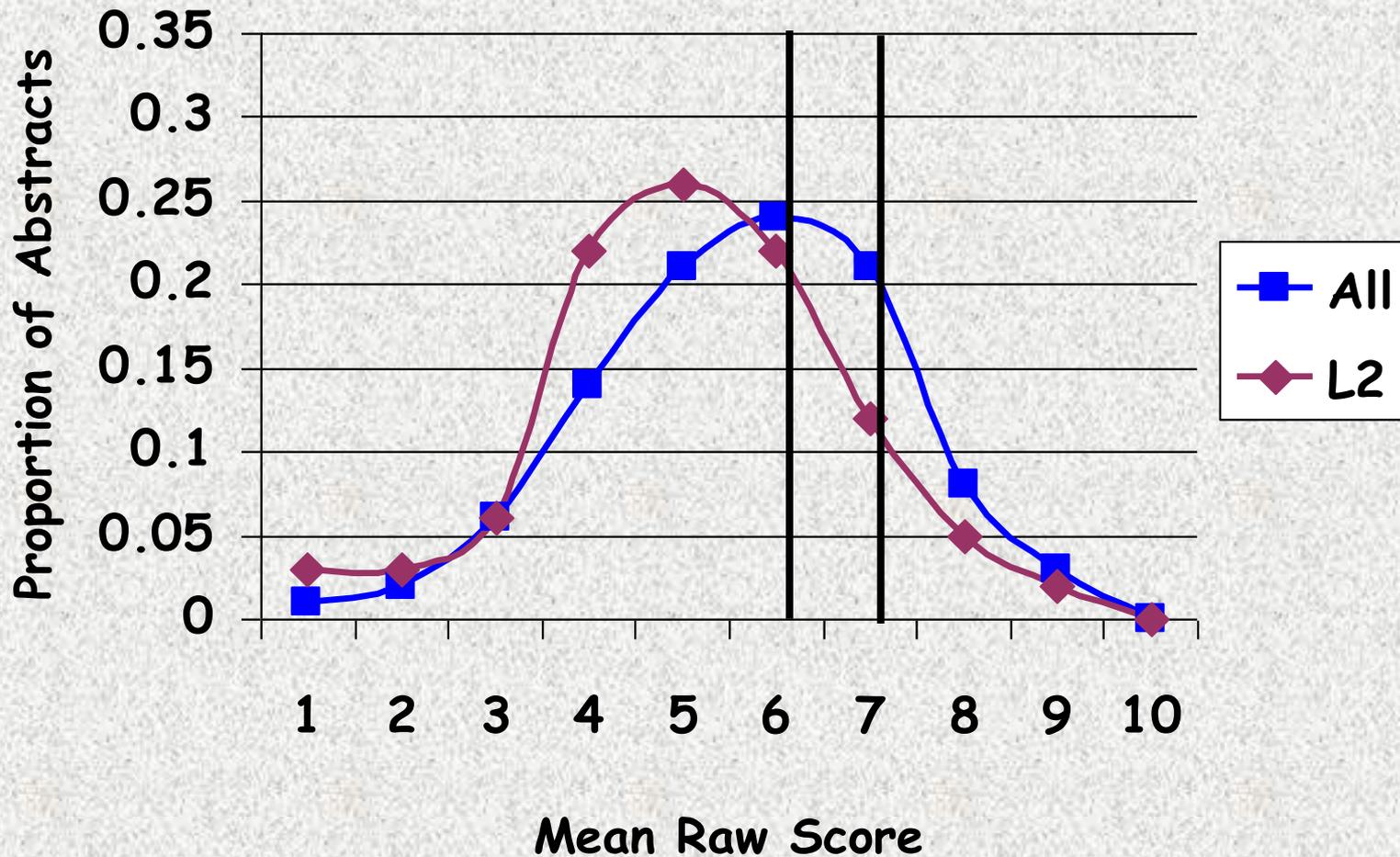
WHY A LOWER/HIGHER ACCEPTANCE RATE? LOWER/HIGHER SCORES!

	All Abstracts	L2	Generative Morsyn.	Sound System
No. Submitted	277	68	73	
Acceptance Rate	33%	19%	25%	
Mean Raw Score	6.08	5.54	5.68	
Mean Z Score	0	-0.14	-0.01	

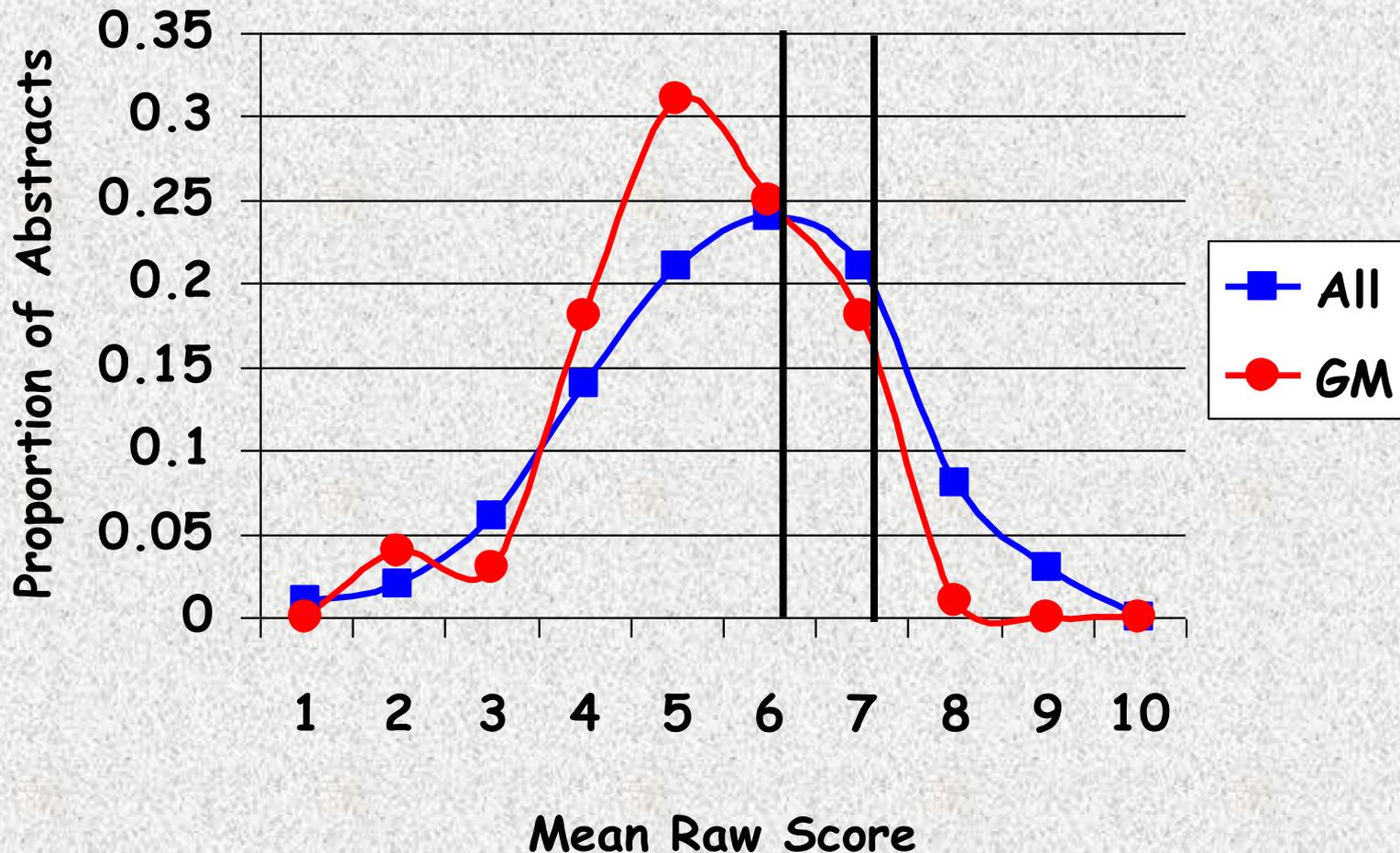
WHY A LOWER/HIGHER ACCEPTANCE RATE? LOWER/HIGHER SCORES!

	All Abstracts	L2	Generative Morsyn.	Sound System
No. Submitted	277	68	73	46
Acceptance Rate	33%	19%	25%	44%
Mean Raw Score	6.08	5.54	5.68	6.15
Mean Z Score	0	-0.14	-0.01	-0.03

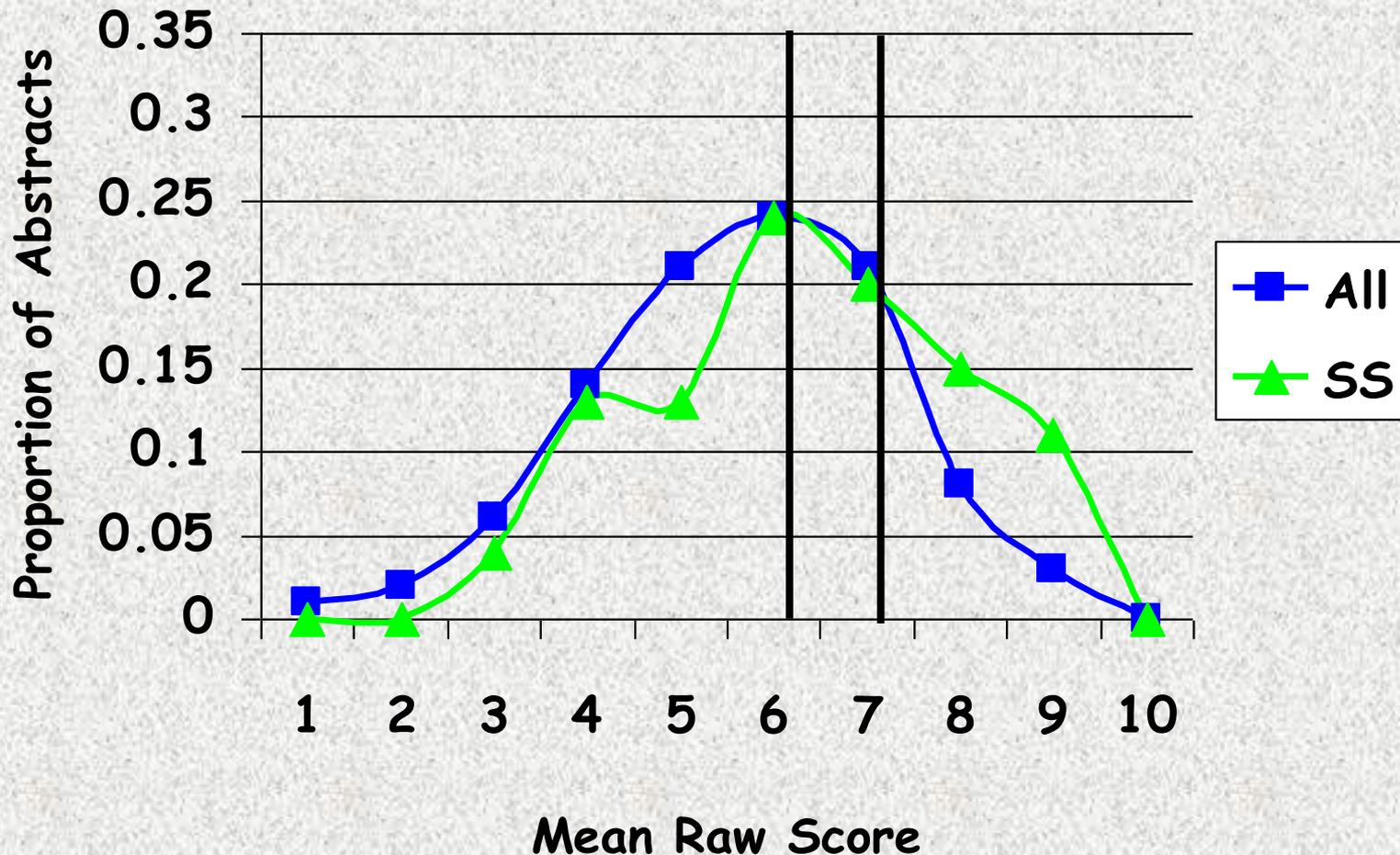
DISTRIBUTION OF MEAN RAW SCORES FOR L2 ABSTRACTS



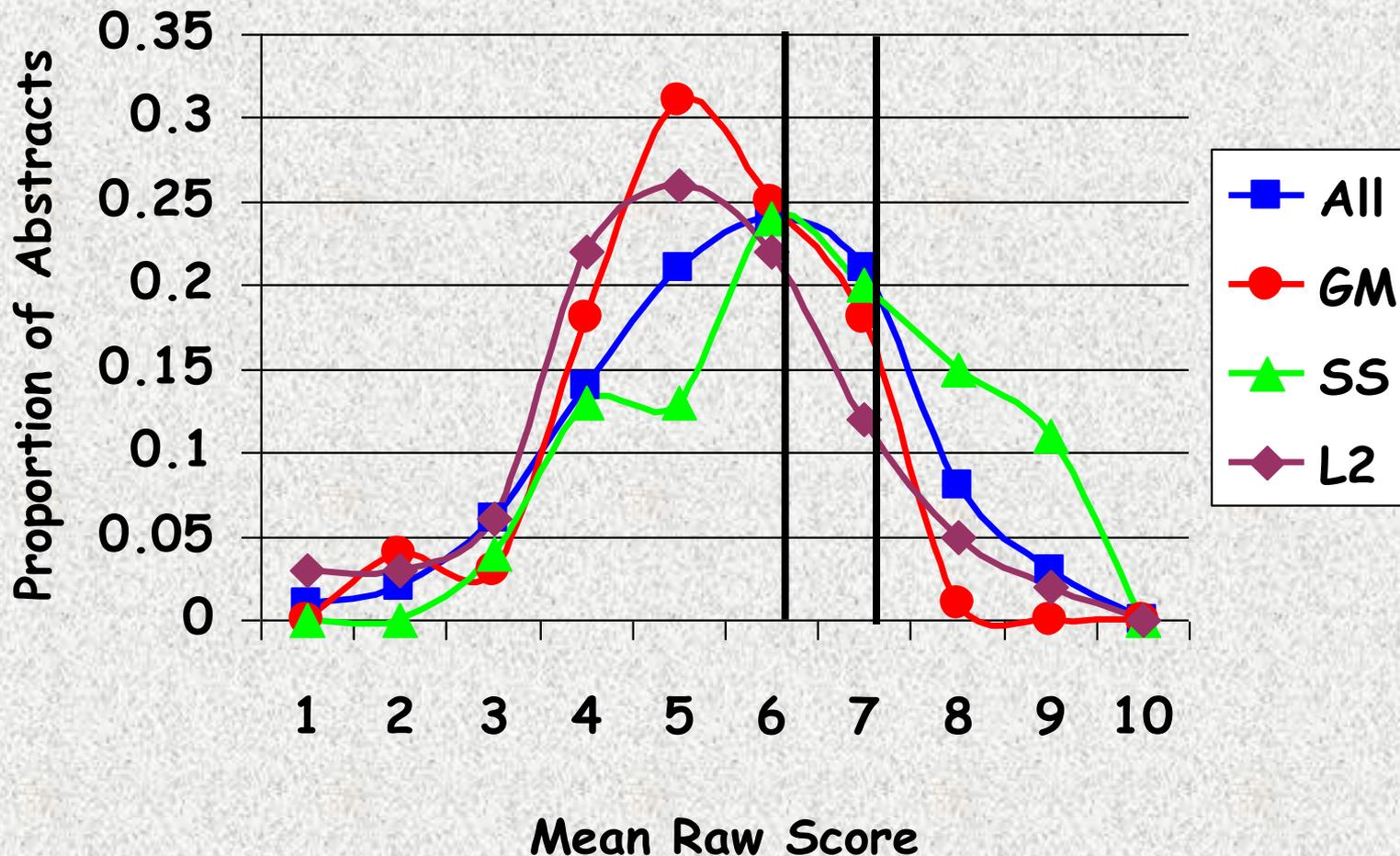
DISTRIBUTION OF MEAN RAW SCORES FOR GENERATIVE MORPHOSYNTAX ABSTRACTS



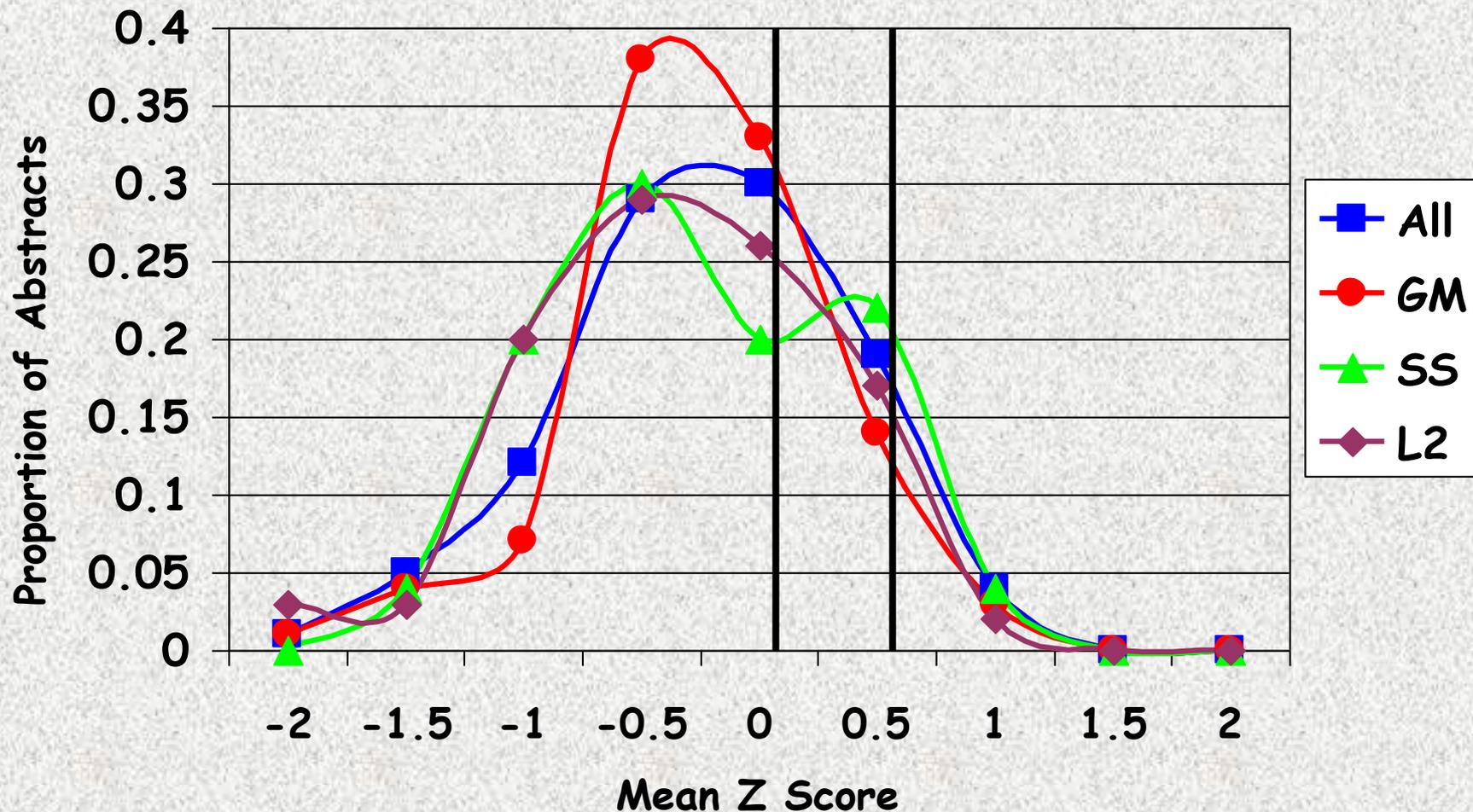
DISTRIBUTION OF MEAN RAW SCORES FOR SOUND SYSTEM ABSTRACTS



SUMMARY OF DISTRIBUTION OF MEAN RAW SCORES FOR THREE SUBFIELDS



DISTRIBUTION OF MEAN Z SCORES FOR THREE SUBFIELDS



POSSIBLE REASONS FOR LOWER/HIGHER SCORES IN CERTAIN SUBFIELDS

1. Quality of work submitted in one subfield is better/worse than that submitted in others.
2. Quality of abstract-writing in one subfield is better/worse than in others.
3. Exciting developments have recently happened in one subfield more than in others.
4. Reviewers for one subfield are more lenient/stringent than those for other subfields.

HOW ARE RATINGS DISTRIBUTED BY REVIEWER?

Source: Mean rating for each of the 82 reviewers
(2002 data only)

RAW SCORE

Range: 3.23 - 9.00

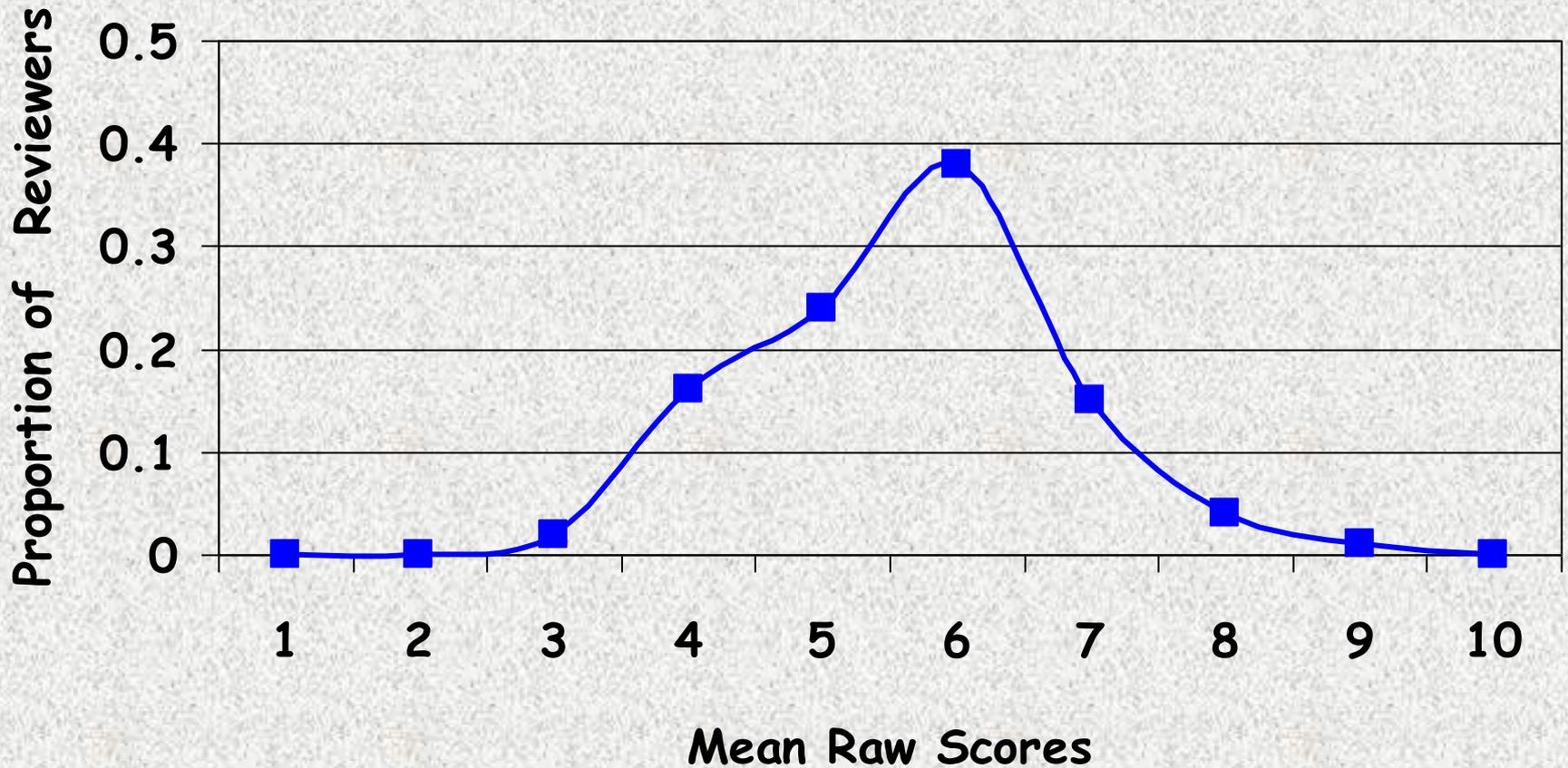
Mean: 6.06 (SD: 1.11)

Z SCORE

Range: all = 0

Mean: 0

DISTRIBUTION OF REVIEWERS' MEAN RAW SCORES



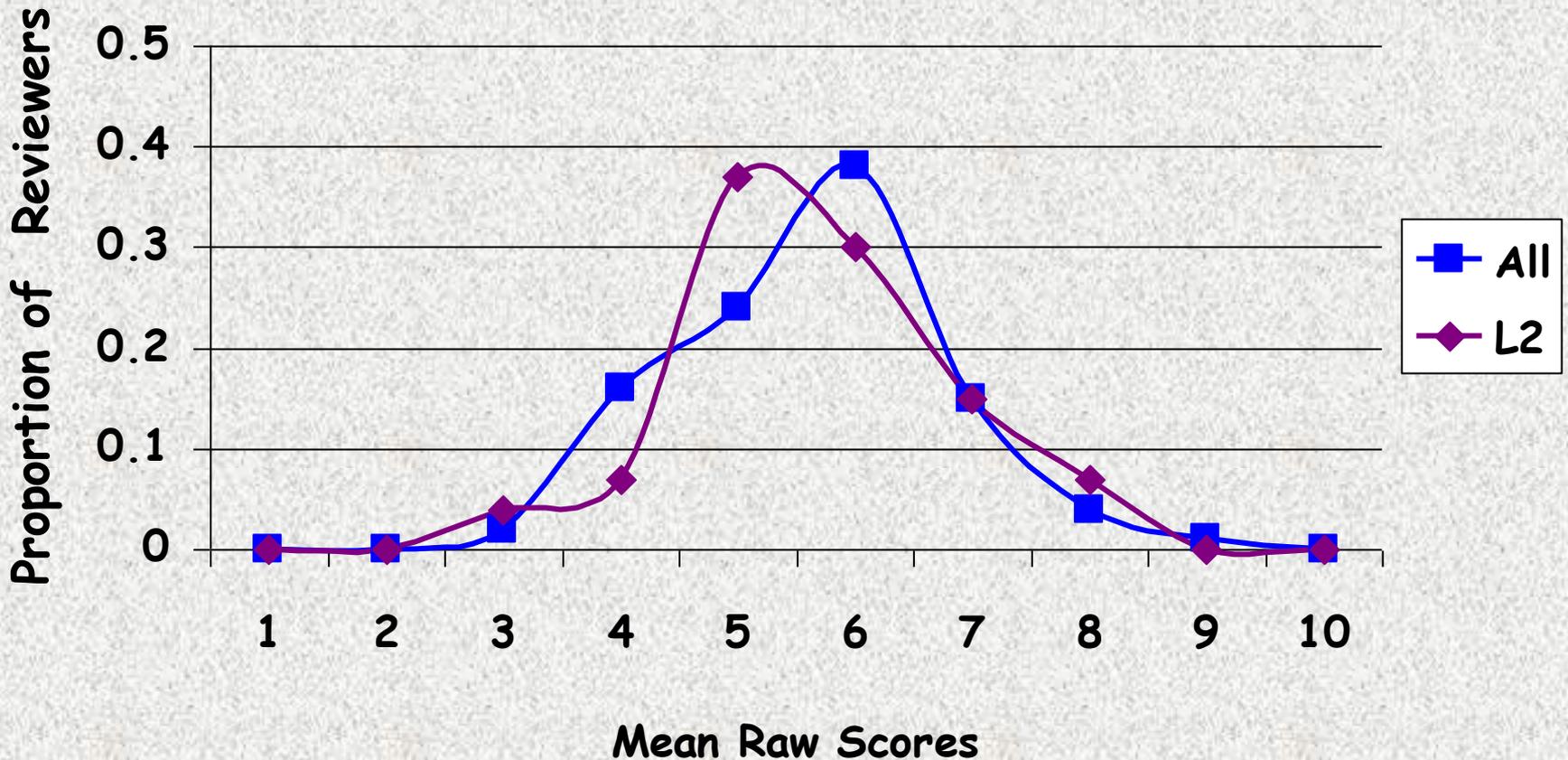
HOW DO REVIEWERS CONTRIBUTE TO LOWER/HIGHER SCORES BY SUBFIELD?

We determined which reviewers reviewed 5 or more abstracts in each subfield of interest:

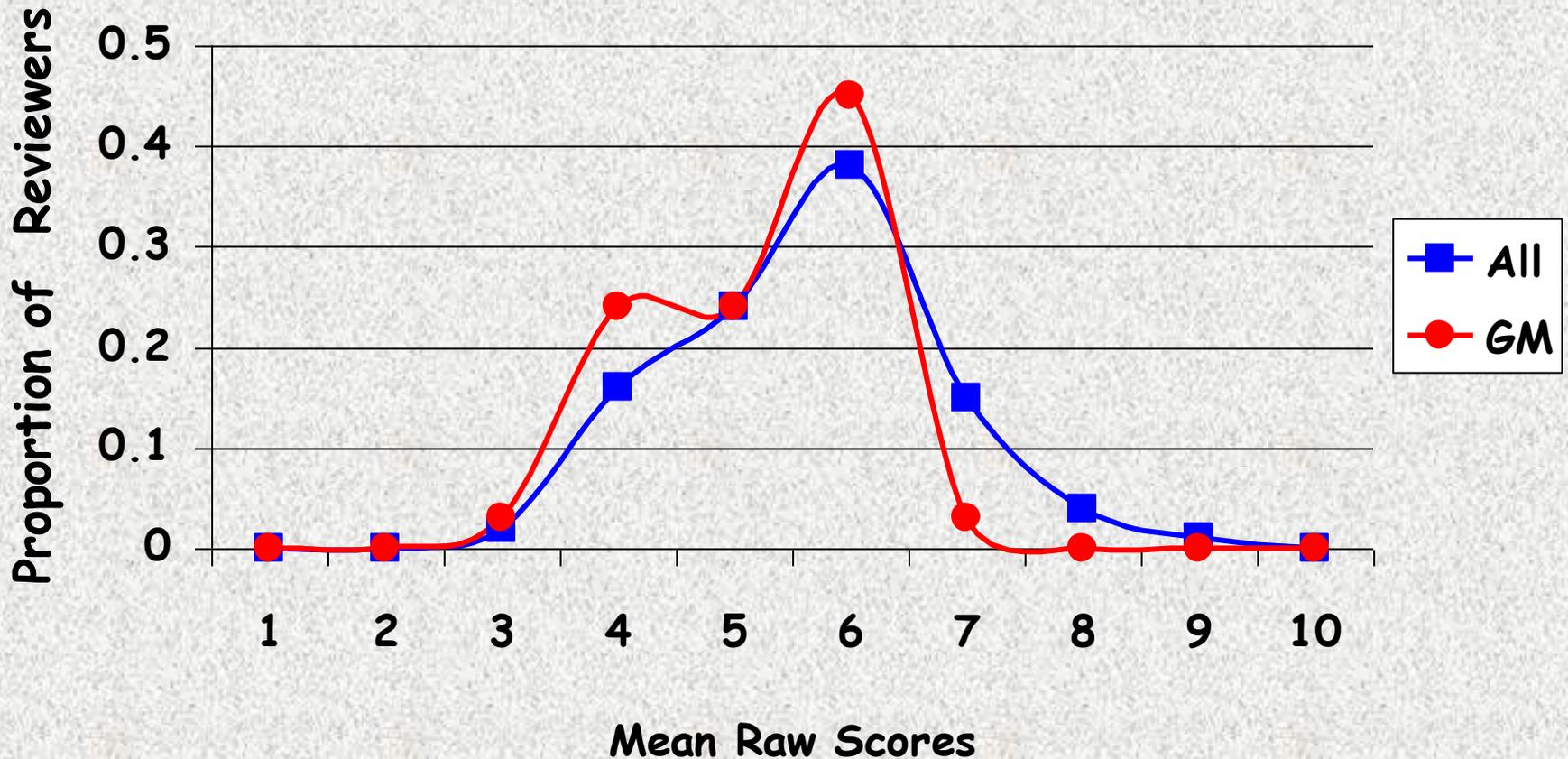
- 27 such reviewers for L2
- 33 such reviewers for *Generative Morphosyntax*
- 15 such reviewers for *Sound System*

For each subfield, we plotted the distribution of the mean raw scores for each of the relevant reviewers.

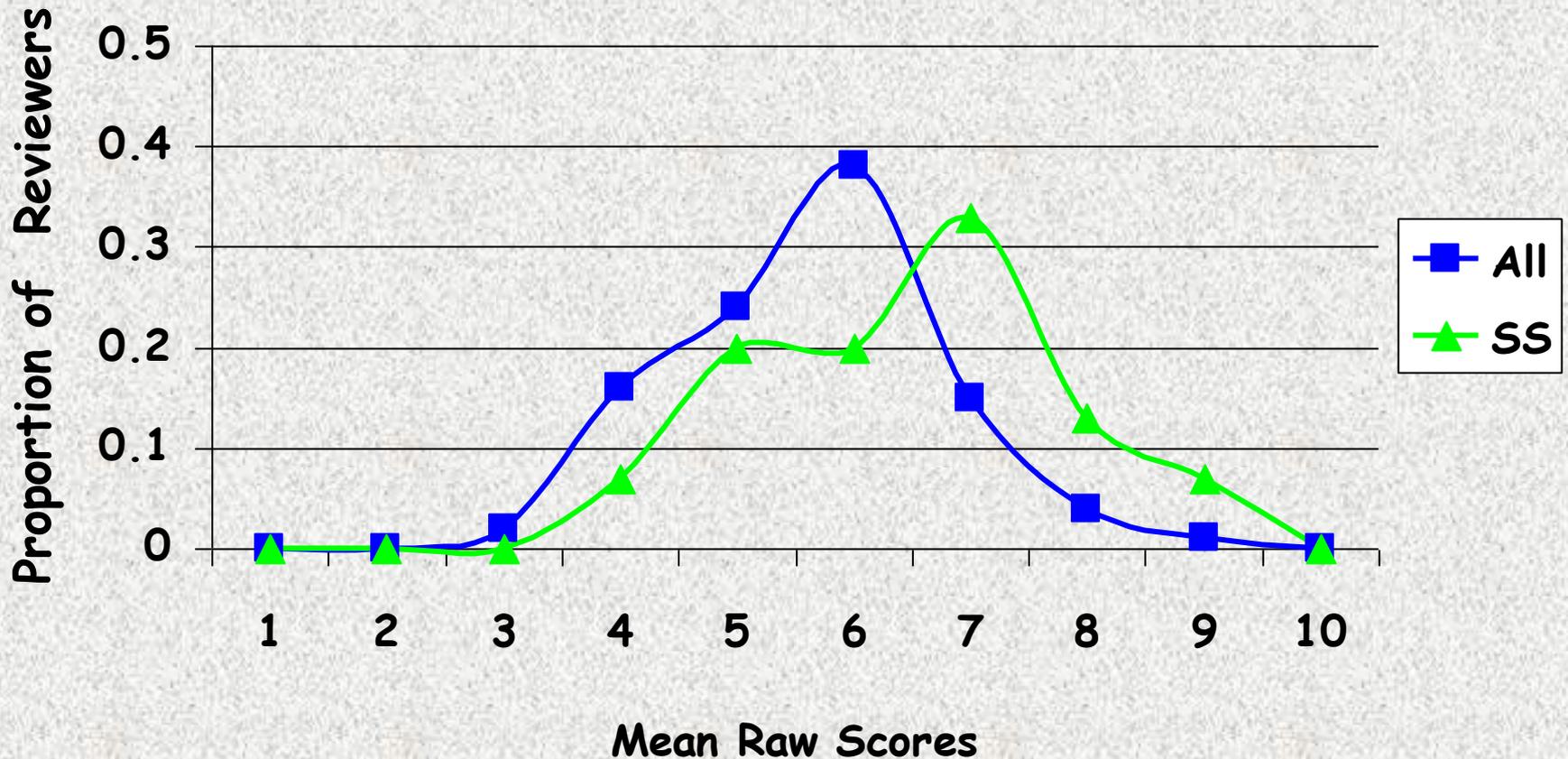
DISTRIBUTION OF L2 REVIEWERS' MEAN RAW SCORES



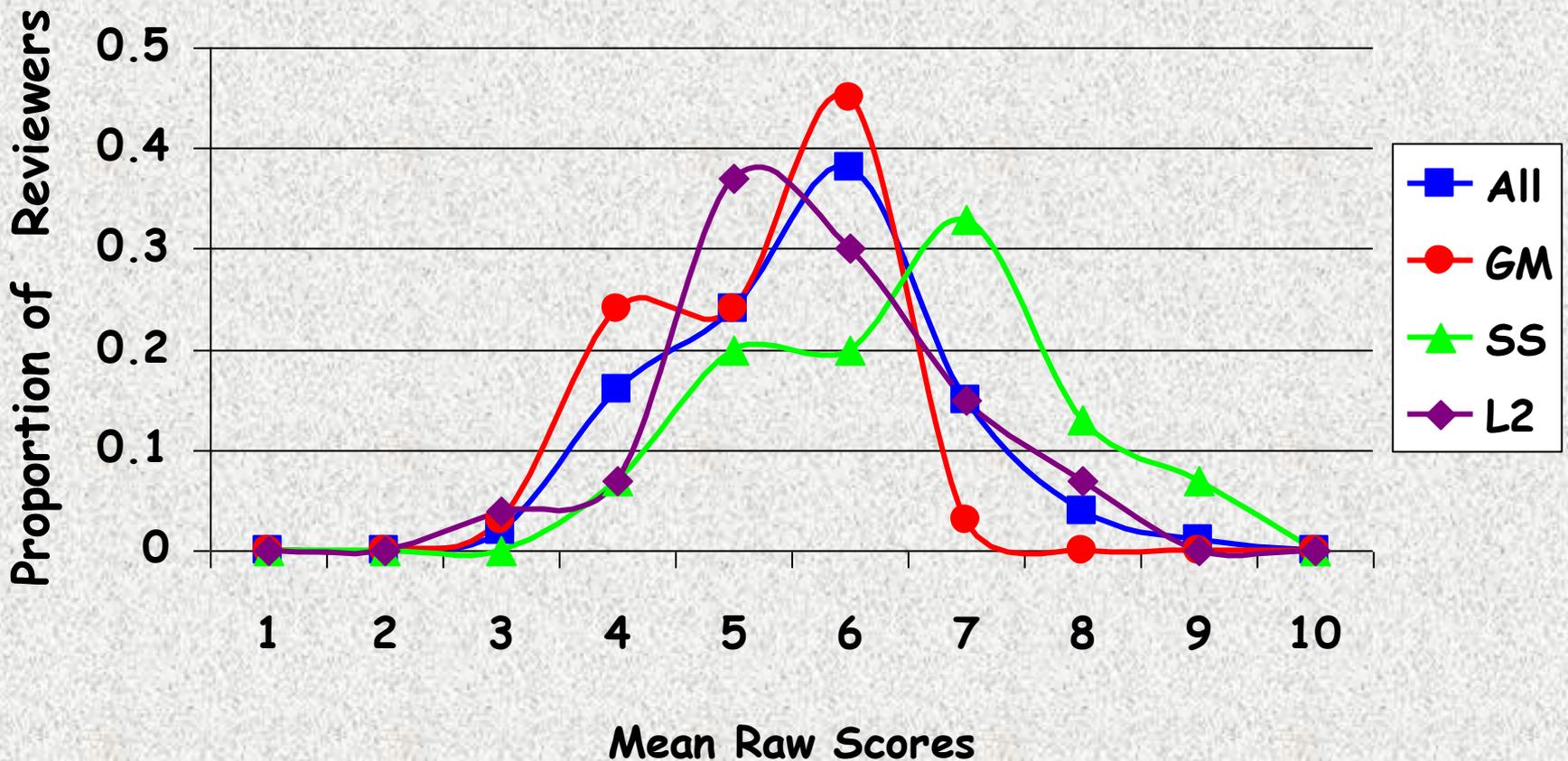
DISTRIBUTION OF GENERATIVE MORPHOSYNTAX REVIEWERS' MEAN RAW SCORES



DISTRIBUTION OF SOUND SYSTEM REVIEWERS' MEAN RAW SCORES



SUMMARY OF DISTRIBUTION OF REVIEWERS' MEAN RAW SCORES FOR THREE SUBFIELDS



SUMMARY PAGE...

**PART 3:
QUESTIONS AND
DISCUSSION**

QUESTIONS & DISCUSSION

1. Should we accept more papers for BUCLD? If so, by which of the three options?
2. Should we move towards more sessions highlighting leading work? If so, by which format(s)?
3. How should we deal with the issue of proportional representation? Is this something we want?
4. How should we deal with the variability in reviewer scoring tendencies by subfield?

FOR FURTHER COMMENTS AND DISCUSSION

Write comments on your BUCLD questionnaire and submit it at the registration desk.

Talk to, call, or e-mail one of the associated faculty:

Shanley Allen	shanley@bu.edu	(617)358-0354
Cathy O'Connor	mco@bu.edu	(617)353-3818

Ask questions right now.

THANK YOU FOR YOUR
CONTINUING SUPPORT
FOR THE
BOSTON UNIVERSITY
CONFERENCE ON
LANGUAGE DEVELOPMENT!

APPENDIX

REVIEWERS FOR L2 ABSTRACTS

The following 27 reviewers contributed 5 or more ratings to the total 335 ratings (68 abstracts x 5 ratings each):

Nameera Akhtar	Catherine Harris	William Philip
Edith Bavin	Marco Haverkort	Colin Phillips
David Birdsong	Alan Juffs	Marnie Reed
Cynthia Brown	Dorit Kaufman	Bonnie Schwartz
Joyce Bruhn	Wolfgang Klein	Rex Sprouse
Peter Coopmans	Donna Lardiere	Carol Stoel-Gammon
Katherine Demuth	Lise Menn	Margaret Thomas
Suzanne Flynn	Cathy O'Connor	Anne Vainikka
Cornelia Hamann	Mitsuhiko Ota	Lydia White

REVIEWERS FOR GENERATIVE MORPHOSYNTAX ABSTRACTS

The following 33 reviewers contributed 5 or more ratings to the total 365 ratings (73 abstracts x 5 ratings each):

Shanley Allen

Joyce Bruhn

Harald Clahsen

Peter Coopmans

Kenneth Drozd

Suzanne Flynn

Paul Hagstrom

Cornelia Hamann

Marco Haverkort

Bart Hollebrandse

Angeliek van Hout

Aafke Hulk

Nina Hyams

Alan Juffs

Donna Lardiere

Carol Neidle

Johanne Paradis

William Philip

Colin Phillips

Mabel Rice

Tom Roeper

Lynn Santelmann

Jeannette Schaeffer

Carson Schütze

Bonnie Schwartz

William Snyder

Rex Sprouse

Margaret Thomas

Anne Vainikka

Jürgen Weissenborn

Lydia White

Frank Wijnen

Andrea Zukowski

REVIEWERS FOR SOUND SYSTEM ABSTRACTS

The following 15 reviewers contributed 5 or more ratings to the total 230 ratings (46 abstracts x 5 ratings each):

Arto Anttila

Richard Aslin

Cynthia Brown

Katherine Demuth

Catherine Echols

Anne Fernald

LouAnn Gerken

Heather Goad

Rebecca Gomez

Richard Meier

Lise Menn

Mitsuhiko Ota

Joseph Pater

Marnie Reed

Carol Stoel-Gammon

REVIEWERS FOR PSYCHOLOGY ABSTRACTS

(Includes language-cognition interface, sentence processing, speech segmentation, speech perception and production, statistical models, and word learning)

The following 28 reviewers contributed 5 or more ratings to the total 315 ratings (63 abstracts x 5 ratings each):

Nameera Akhtar

Richard Aslin

Paul Bloom

Melissa Bowerman

Cynthia Brown

Susan Carey

Catherine Echols

Anne Fernald

Cynthia Fisher

LouAnn Gerken

Susan Goldin-Meadow

Rebecca Gomez

Peter Gordon

Catherine Harris

Deb Kelemen

Richard Meier

Lise Menn

Letitia Naigles

Elissa Newport

Marnie Reed

Mabel Rice

Carson Schütze

Ann Senghas

Jesse Snedeker

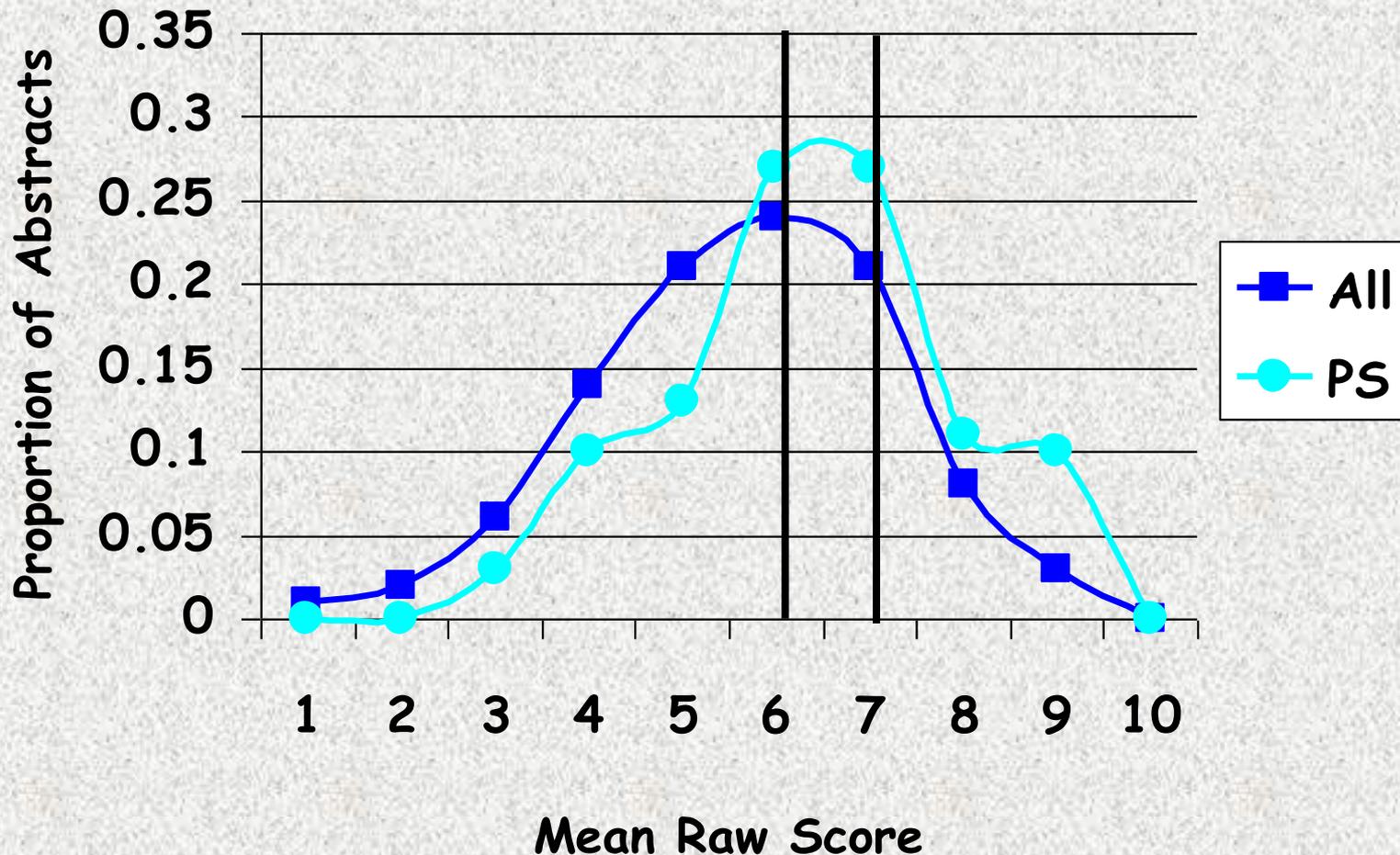
Carol Stoel-Gammon

Helen Tager-Flusberg

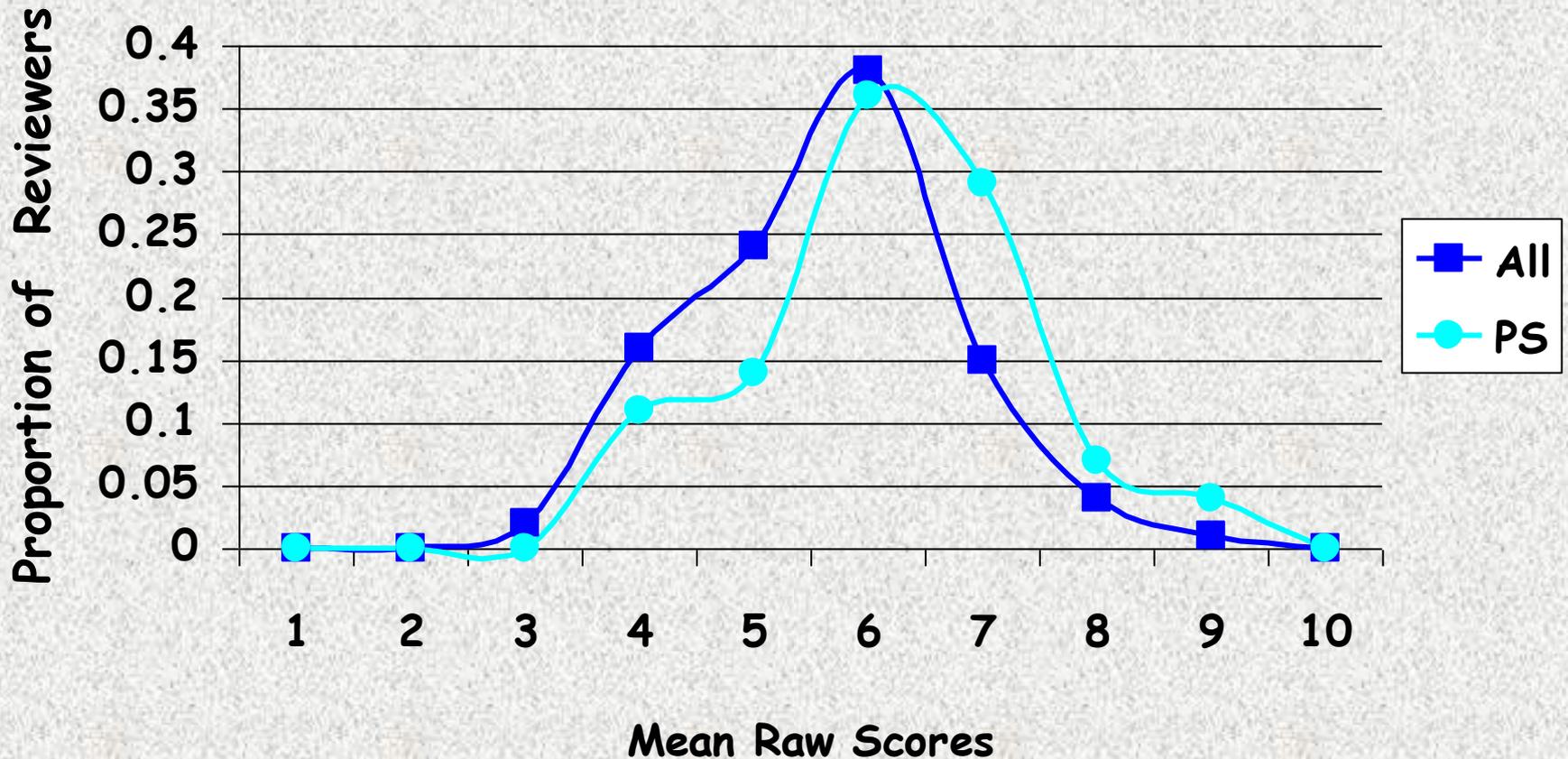
Michael Tomasello

Fei Xu

DISTRIBUTION OF MEAN RAW SCORES FOR "PSYCHOLOGY"



DISTRIBUTION OF "PSYCHOLOGY" REVIEWERS' MEAN RAW SCORES

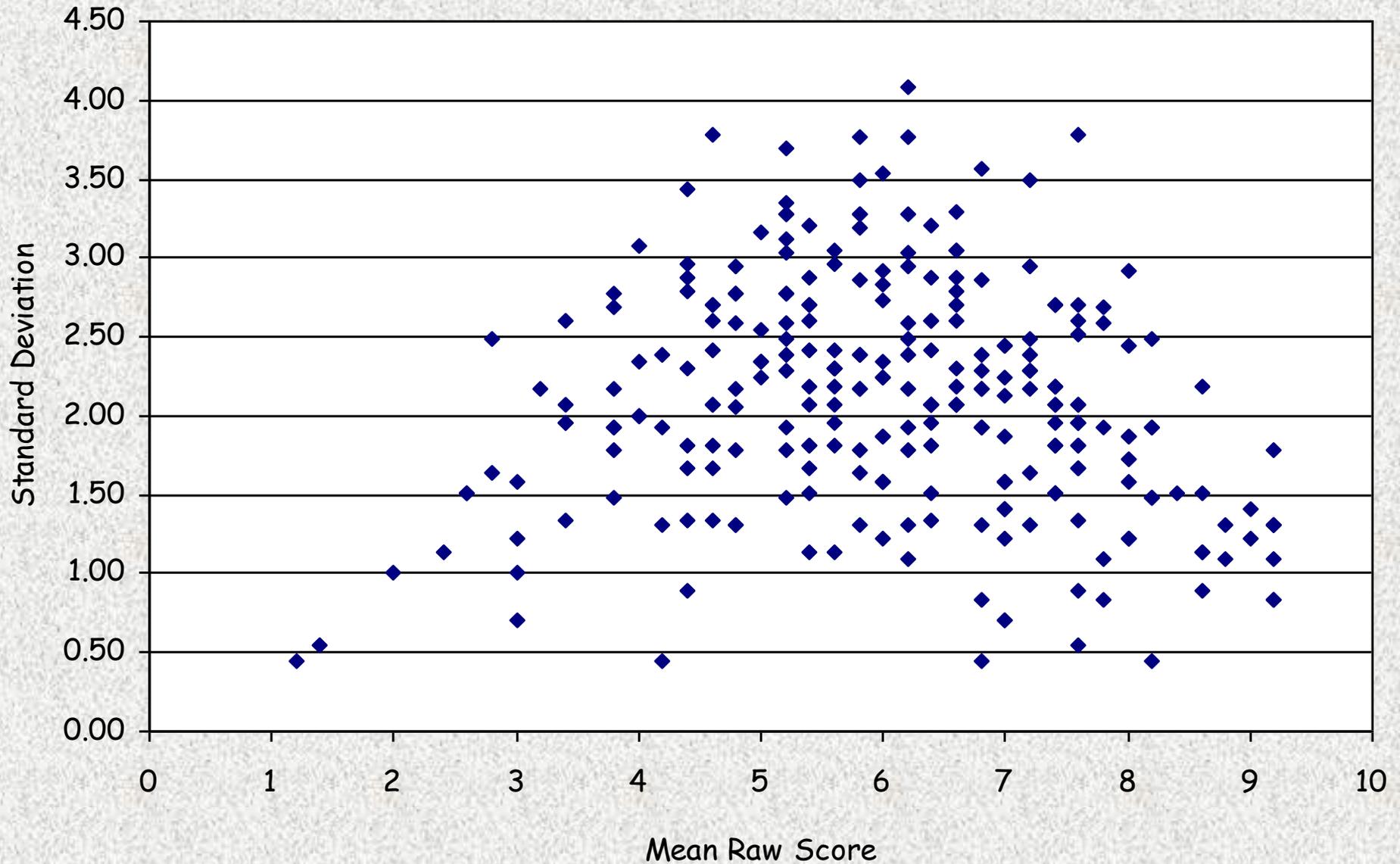


HOW ARE PLENARY AND KEYNOTE SPEAKERS SELECTED?

1. Solicit suggestions on annual BUCLD questionnaire.
2. Solicit suggestions from many people in the field.
3. Balance content areas, theoretical focus, and learning conditions.
4. No repeat speakers.

EXTRA SLIDES

RELATIONSHIP OF MEAN TO SD



IS THE SELECTION PROCESS FAIR?

1. The process we use addresses some obvious concerns as much as is reasonably possible.
2. The other statistical massaging methods we've tried (e.g. different cutoff levels, different weightings of raw vs. z scores, adding an "offset score") all result in the same top 60 or 70 papers getting accepted. Sensitivity near the threshold is inevitable with any system. The important thing is that the really good papers don't get rejected.
3. Too much statistical massaging of the ratings would undermine the work of the reviewers.
4. There is no obvious objective standard of what fairness would be, so there's bound to be some room for debate.

INVENTORY OF LANGUAGES

19 Languages Represented in 2001 (out of 33 submitted):

African American English, American Sign Language, Chinese, Dutch, English, French, German, Greek, Hebrew, Inuktitut, Italian, Japanese, Korean, Mandarin, Nicaraguan Sign Language, Sesotho, Spanish, Swahili, Turkish

22 Languages Represented in 2002 (out of 42 submitted):

Afrikaans, American Sign Language, British Sign Language, Catalan, Dutch, English, French, German, Greek, Hebrew, Icelandic, Italian, Japanese, Kannada, Korean, Norwegian, Portuguese, Russian, Spanish, Swahili, Thai, Turkish

STATISTICS FOR TOP 6 LANGUAGES

	2001		2002	
	N RCD	% ACC	N RCD	% ACC
English				
Spanish				
French				
German				
Japanese				
Dutch				

STATISTICS FOR TOP 6 LANGUAGES

	2001		2002	
	N RCD	% ACC	N RCD	% ACC
English	161	39%		
Spanish	25	28%		
French	23	26%		
German	20	40%		
Japanese	18	11%		
Dutch	9	33%		

STATISTICS FOR TOP 6 LANGUAGES

	2001		2002	
	N RCD	% ACC	N RCD	% ACC
English	161	39%	154	42%
Spanish	25	28%	24	21%
French	23	26%	23	30%
German	20	40%	21	33%
Japanese	18	11%	16	31%
Dutch	9	33%	13	23%

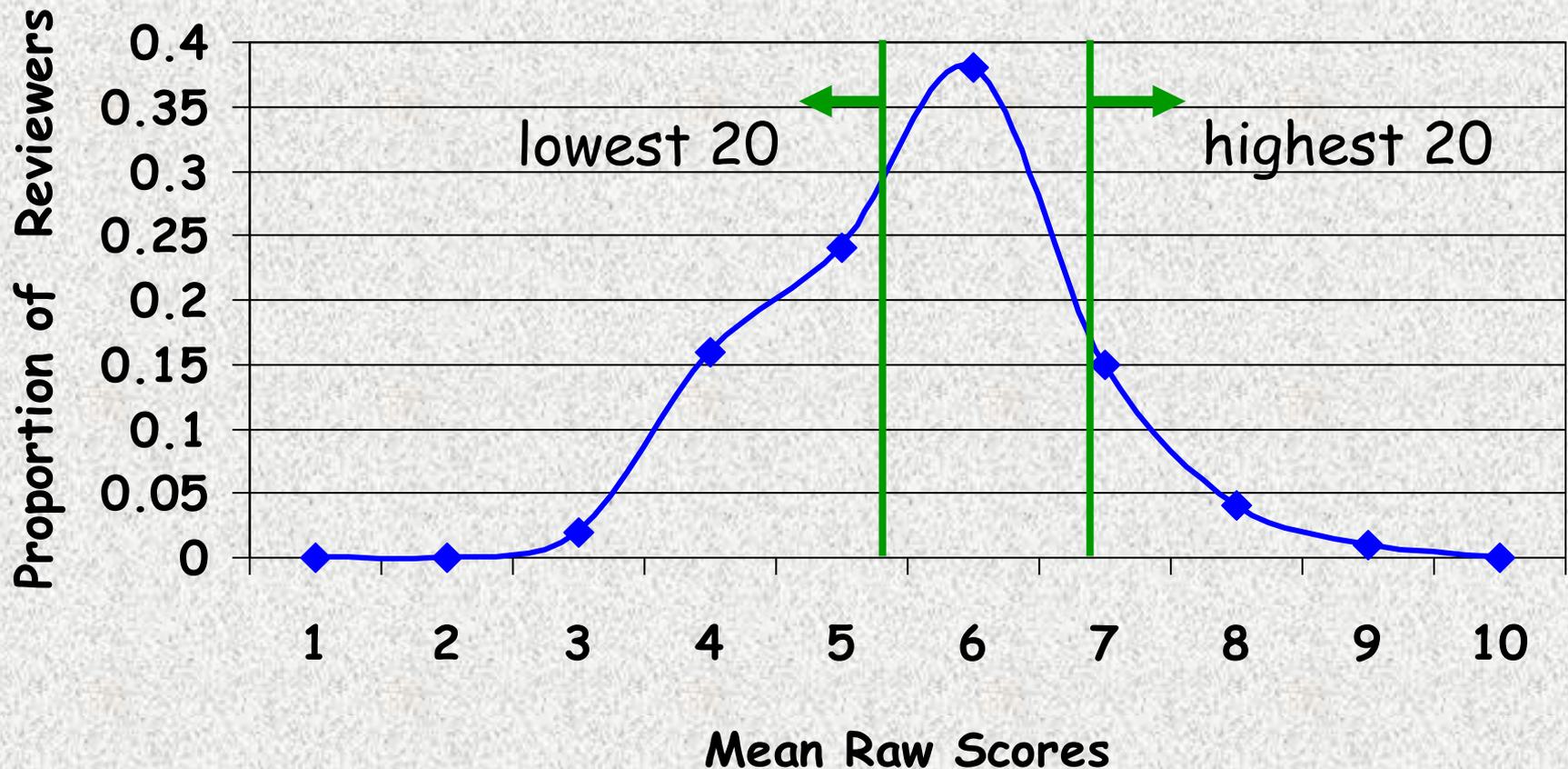
HOW DO REVIEWERS CONTRIBUTE TO LOWER/HIGHER SCORES BY SUBFIELD?

at least 5 abstracts

	All Reviewers	L2	Generative Morsyn.	Sound System
Mean Raw	6.06	6.04	5.66	6.85
Range Raw	3.23-9.00	3.23-8.94	3.23-7.06	4.73-9.00

WHICH SUBFIELDS GET MORE SCORES FROM HIGH OR LOW SCORERS?

Compare contributions to ratings in different subfields by the 20 highest-rating reviewers and the 20 lowest-rating reviewers (by mean raw score).



RATINGS FROM HIGHEST 20 AND LOWEST 20 REVIEWERS

	Total No. of Ratings	Percent Ratings from:	
		Highest 20 Reviewers	Lowest 20 Reviewers
Total	1385	25%	24%
L2	335	14%	30%
Generative Morphosyntax	365	10%	35%
Sound System	230	63%	8%

RATINGS FROM HIGHEST 20 AND LOWEST 20 REVIEWERS

	Total No. of Ratings	Percent Ratings from:	
		Highest 20 Reviewers	Lowest 20 Reviewers
Total	1385	25%	24%
L2	335	14%	30%
L1	850	27%	23%
Gen. Morphosyn.	365	10%	35%
Non-Gen. Morsyn.	250	19%	20%
Sound System	230	63%	8%
Morphosyntax	615	14%	29%