Almost everything you wanted to know about Student Ratings of Instruction (SRI)...

Nira Hativa, Tel-Aviv University (**TAU**)

Are SRIs valid and reliable?
How to report results? How to design the forms?
...

BASED ON RESEARCH ON SRI

- Extensive 40-year **US-based research literature—** **Thousands of studies and publications…**
- **My analyses of TAU database:** 11 schools, about 30,000 students, 4,000 courses, 120,000 student forms per semester, during 12-years X 2 semesters, lecture courses

THE CONTENT IS PRESENTED IN MORE DETAIL IN MY TWO BOOKS ON SRIS (2014)

Book 1: Student Ratings of Instruction: A Practical Approach to Designing, Operating, and Reporting (Second Edition) (226 pp.)

Book 2: Student Ratings of Instruction: Recognizing Effective Teaching (Second Edition) (170 pp.)

AUDIENCE INTERESTS IN DECREASING ORDER

4. Do SRIs correlate with student learning?	40	52.7	92.7	55	4.42
5. Do SRIs correlate with the main behaviors of effective instruction?	46.3	44.4	90.7	54	4.31
7. Does instructor popularity/expressiveness/enthusiasm bias SRI validity?	38.2	47.3	85.5	55	4.29
8. Does perceived course difficulty or workload bias SRI validity?	43.6	40	83.6	55	4.24
28. What to report for faculty and how?	48.1	38.9	87	54	4.24
29. What to report for administrators and how?	44.4	40.7	85.1	54	4.19
25. What items should necessarily be included in the form? (those that are related to the course? Instructor? Students? Student learning?)	40.7	40.7	81.4	54	4.11
13. Does class size affect SRI?	45.5	34.5	80	55	4.09
19. What are the reliability and validity of written comments as a source for teacher evaluation?	51.9	28.8	80.7	52	4.06
6. Do SRIs correlate with other measures of effective instruction? (e.g., peer-, superior-, alumni/ retrospective evaluation)	43.6	34.5	78.1	55	4.04
9. Does expected or actual grades bias SRI validity?	40.7	37	77.7	54	4.04
31. Why and how to provide comparative data about courses?	35.2	37	72.2	54	4.04
26. Should there be different forms for different types of courses?	34	37.7	71.7	53	4
32. What is the effect of reporting end-of-term results (summative evaluation) on teaching improvement?	40.7	31.5	72.2	54	3.98
27. Should there be different forms for different disciplines?	37	35.2	72.2	54	3.96
30. What to report for students and how?	37	33.3	70.3	54	3.96
33. What is the effect of reporting midterm results (formative evaluation) on teaching improvement?	35.2	35.2	70.4	54	3.96
3. Are SRI results stable over repeated measurements?	58.8	19.6	78.4	51	3.92
12. Does student/teacher gender affect SRI?	32.7	36.4	69.1	55	3.91
1. What do SRIs measure? (Student learning? Teaching quality/effectiveness?)	56.4	21.8	78.2	55	3.89
20. Student ratings are popularity contests that measure instructor expressiveness/style/student entertainment rather than the substance of teaching?	33.3	35.2	68.5	54	3.89
35. Do online ratings encourage higher proportion of participation by dissatisfied students than with satisfied students?	42.6	29.6	72.2	54	3.85
36. Do the expected higher proportion of dissatisfied students in online ratings result in lower rating for the instructor or in higher proportion of negative written comments?	42.6	29.6	72.2	54	3.85
34. Do online ratings lower the response rate?	40.7	25.9	66.6	54	3.8
11. Does student ability (GPA) affect SRI?	53.7	16.7	70.4	53	3.77
15. Does motivation to take the course (required vs. elective) affect SRI?	44.4	22.2	66.6	54	3.76
22. Instructors who lower the level of course difficulty, assign low workload, or give high grades are rewarded by higher ratings?	37	27.8	64.8	54	3.74
10. Can students be manipulated to give faculty higher ratings?	27.8	31.5	59.3	54	3.7
2. Are SRI results consistent for students of the same class?	52.7	18.2	70.9	55	3.69
17. Why do instructors in "hard" disciplines rate substantially lower than "soft" disciplines?	34.5	23.6	58.1	55	3.69
14. Does degree level affect SRI?	33.9	23.2	57.1	56	3.63
16. Does discipline affect SRI?	42.9	16.1	59	56	3.59
18. What is the content of written comments?	37	16.7	53.7	54	3.54
23. Students cannot appreciate good teaching by the end of the course, until later years of study or after graduation?	42.6	14.8	57.4	54	3.39
21. Students cannot make consistent judgments about instructors and teaching effectiveness due to their immaturity, inexperience, and capriciousness?	24.1	22.2	46.3	54	3.31
24. Better teachers receive worse student evaluation?	20.4	14.8	35.2	54	3.11

OUTLINE

BASED ON YOUR REQUEST IN THE SURVEY, IN ORDER OF PREFERENCE

A. SRI validity
B. Factors that bias SRI validity: Myths and refuting them
C. Written comments
D. Effects of SRI reports on improving teaching
E. Faculty concerns about online ratings
F. SRI forms
G. Reporting SRI results

A. SRI validity

A. SRI validity

SRI validity—definition

What main issues do SRIs measure or correlate with/relate to?

1. Effective teaching?

2. Student learning?

Student satisfaction from teaching and the course?

What other factors do SRIs correlate with/relate to?

3. Course difficulty or workload?

4. Expected or actual grades?

5. Instructor expressiveness/enthusiasm?

6. Instructor popularity?

7. Class size?

8. Academic domain?

9. Additional extraneous factors?

SRI validity: Definition

Validity is the degree to which a measurement tool measures whatever it purports to measure

To define **SRI Validity** we need to first define **what do SRIs purport to measure**

Teaching effectiveness? Student learning?

SRIs cannot measure **directly** anything regarding teaching or learning!

However, we'll see that it measures **indirectly** major aspects of **teaching effectiveness**!

What do SRIs measure (directly)?

Whatever SRIs measure is based on **student input:**

Student thinking/opinions/conceptions/perceptions

(regarding instruction and the course)

Or in a measurable term:

SRIs measure **student satisfaction** **(with instruction and the course)**

For statistical purposes**, SRI is measured by the course mean ratings on the "global/overall" teaching item,** e.g., **"Overall teaching performance"**

SRI validity: Definition

SRI Validity is the degree to which it measures student satisfaction with instruction and the course

We'll see that it highly correlates with important aspects of teaching effectiveness

What do SRIs correlate with/related to?

① Effective teaching?
② Student learning?
③ Course difficulty/workload?
④ Expected or actual grades?
⑤ Instructor expressiveness/enthusiasm?
⑥ Instructor popularity
⑦ Class size?
⑧ Discipline?
⑨ Other extraneous factors: Instructor gender/race/physical attractiveness/academic degree, course components?

① Do SRIs relate to effective teaching?

The most agreed upon definition for effective teaching is:

- The effectiveness with which **the teacher is producing the desired student outcomes** for the given student population (Carl Wieman, Nov. 2015)
- Teaching that **achieves its goals** (mine and others' definition)

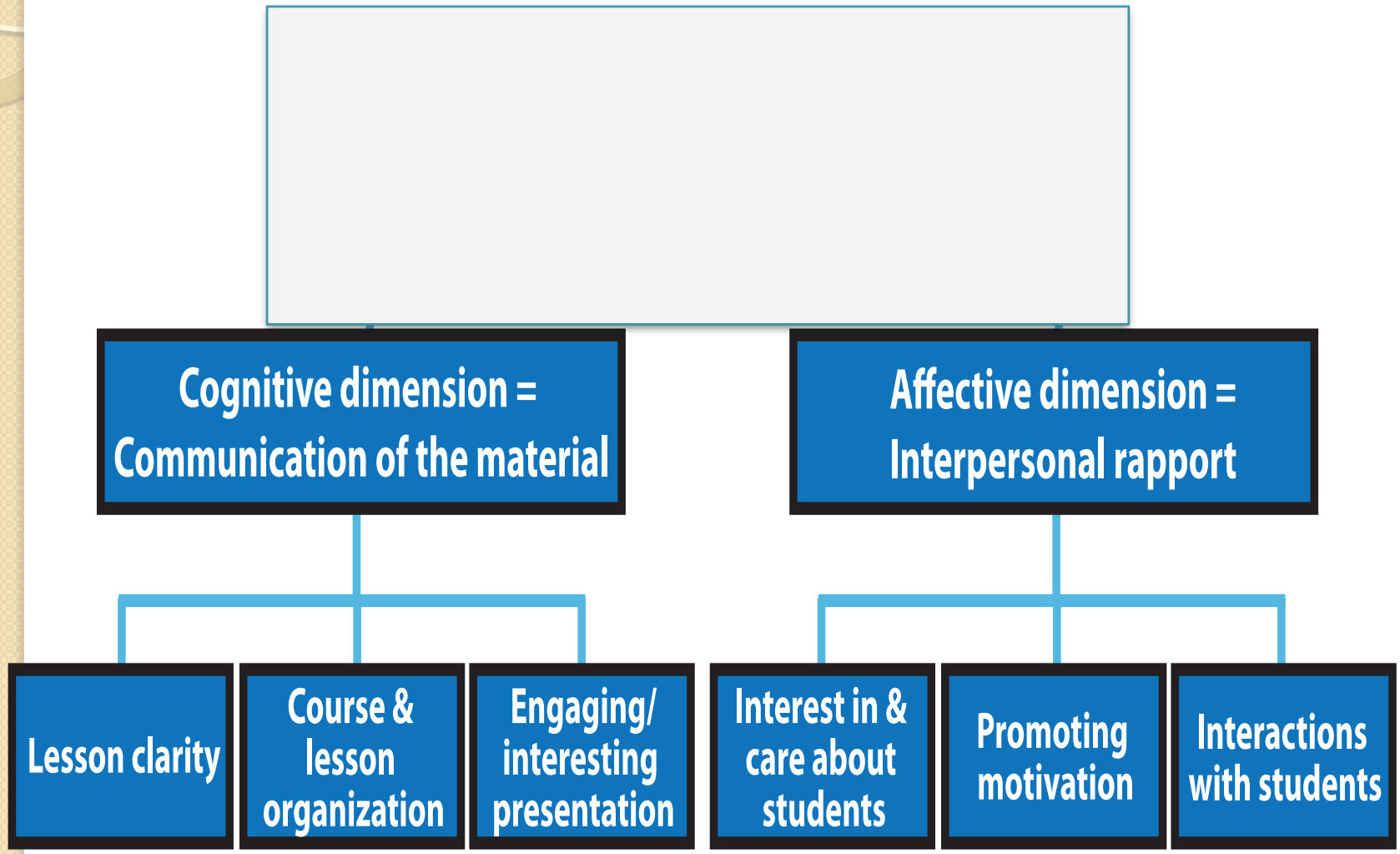
How to measure it?

Proxies for effective teaching:

- Main classroom behaviors of effective teaching

- Acceptable measures of effective teaching other than SRIs

What are the main general classroom behaviors of effective teaching?



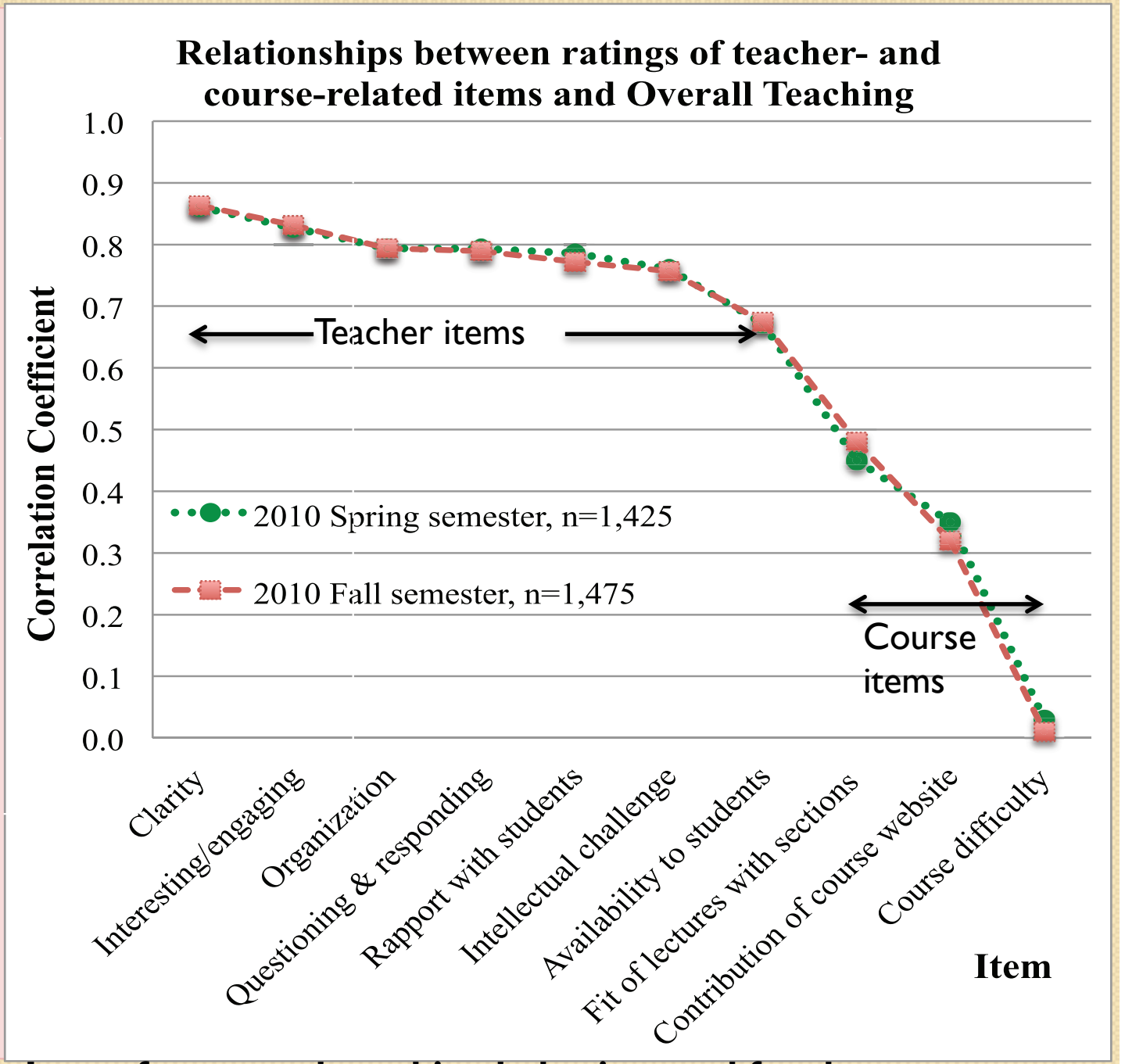
Do SRIs relate to effective teaching?

Are there correlations between SRIs and main general classroom behaviors of effective teaching?

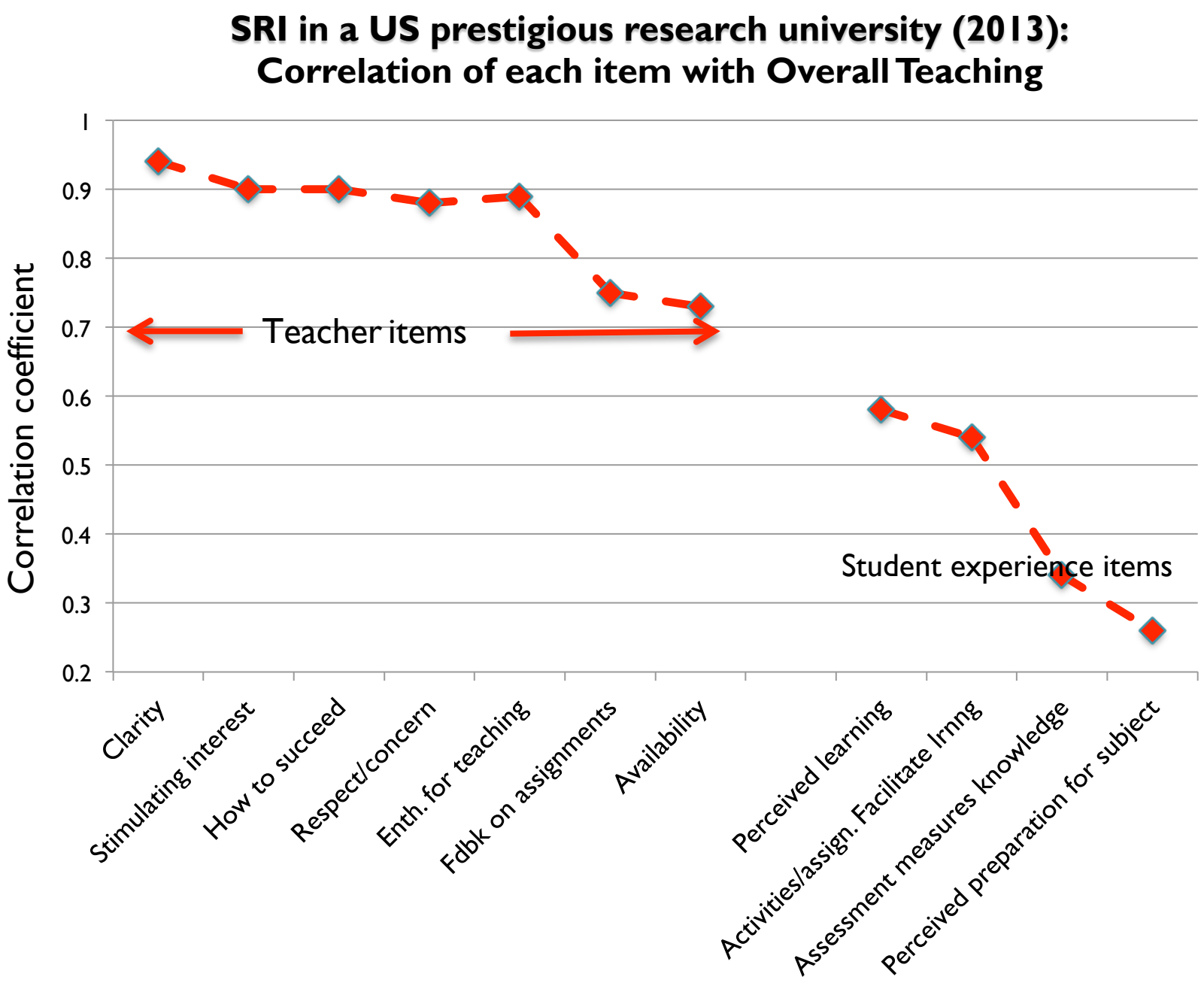
TAU,
Undergrad lecture courses

Conclusions:
1. **teacher-behavior items** are related higher to **teacher overall ratings** than course items

2. The congruent lines = stability of measurement = **high SRI reliability**



Items for general teaching behaviors and for the course


SRI in a US prestigious research university (2013):
Correlation of each item with Overall Teaching
Correlation coefficient
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
Teacher items
Student experience items
Clarity
Stimulating interest
How to succeed
Respect/concern
Enth. for teaching
Fdbk on assignments
Availability
Perceived learning
Activities/assign. Facilitate lrnng
Assessment measures knowledge
Perceived preparation for subject

A study at TAU to identify:

Relationships between teacher main classroom behaviors and SRI

Factor analysis for all undergraduate lecture courses, Schools of Humanities (n = 230) and Engineering (n = 110), Fall Semester 2009.

The 13 TAU items were reduced/converged into 4 factors

Factor	SRI Item/Factor #:	1		2		3		4	
		Hum.	Eng.	Hum.	Eng.	Hum.	Eng.	Hum.	Eng.
1. Pedagogical skills	Interesting/ engaging	.87	.87	.23	.24	.17	.21	.07	.04
	Intellectual challenge	.82	.82	.20	.23	.17	.23	.24	.21
	Clarity	.80	.76	.35	.43	.17	.23	-.19	-.16
	Organization	.75	.66	.36	.50	.42	.23	-.04	-.04
	Contribution of reading materials	.55		.22		.42		.13	
2. Instructor interactions/ rapport with students	Rapport with students	.37	.31	.85	.87	.17	.20	-.01	-.01
	Availability to students	.23	.26	.78	.75	.22	.29	.08	.13
3. Course components	Contribution of discussion/ practice sections	.12	.04	.09	.12	.87	.86	.00	.00
	Fit of lectures with sections	.07	.12	.18	.27	.82	.77	-.01	-.02
	Contribution of homework	.34	.25	.17	.17	.67	.77	-.04	.07
	Contribution of course website	.32	.24	.18	.14	.43	.71	.24	.11
	Contribution of reading materials		.38		.15		.59		.12
4. Course difficulty	Course difficulty, compared to other courses in the department	.05	.04	-.01	.03	.07	.11	.96	.97

Factors related the strongest to the SRI global item—Overall Teaching

Regression analysis: the 4 factors as independent variables, Overall Teaching as the dependent variable

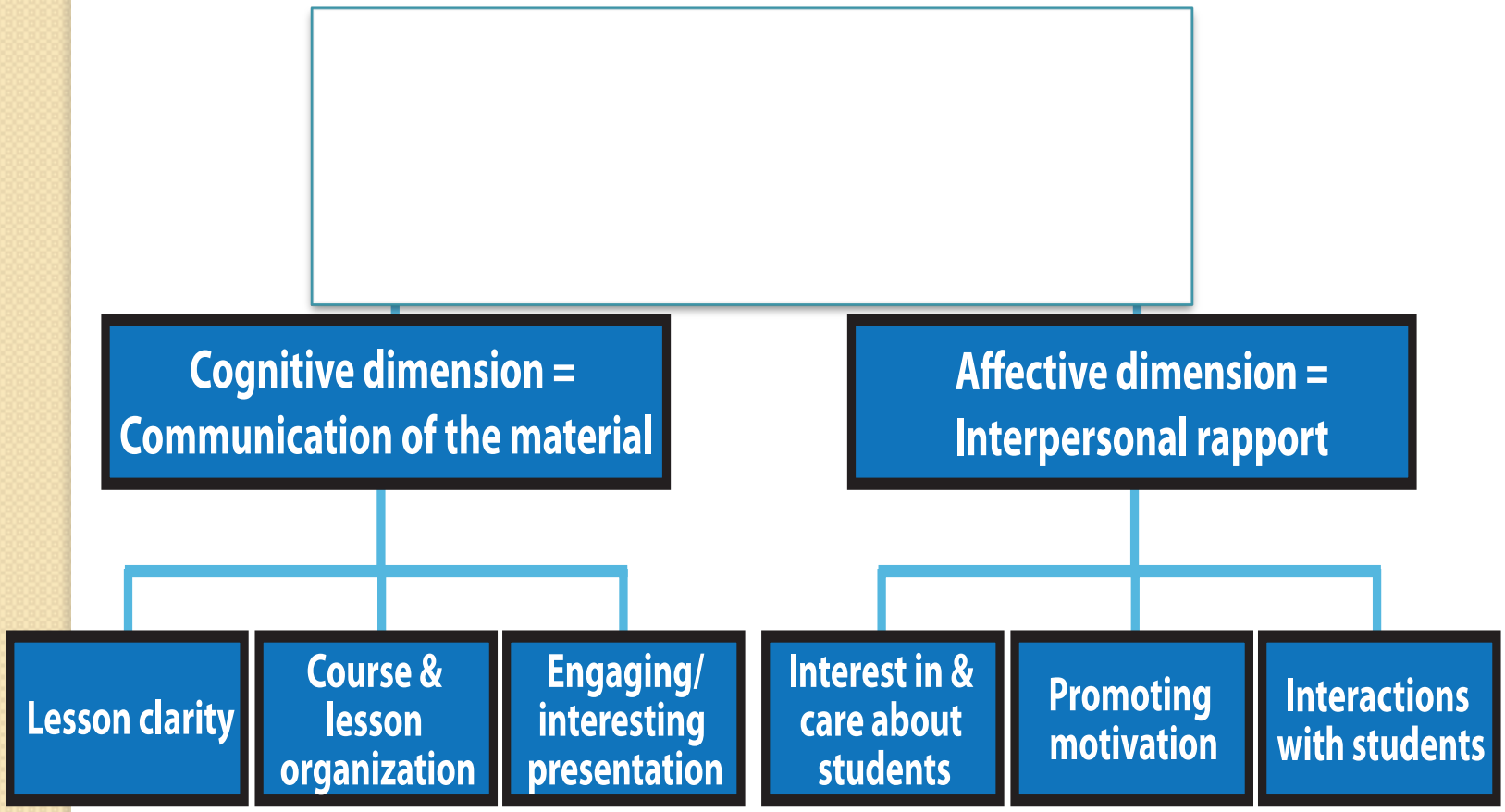
Fall Semester 2009

Factors	Beta coefficient		R		R^2 = Explained variance		Standard error	
	Hum.	Eng.	Hum.	Eng.	Hum.	Eng.	Hum.	Eng.
1. Pedagogical skills	**.70**	**.66**	.86	.89	.73	.78	.53	.77
2. Instructor rapport with students	**.46**	**.54**						
3. Course components	.17	.22						
4. Course difficulty	.04	-.02						

Conclusions:

1. **SRI** is strongly correlated with **the main effective general classroom teaching behaviors!**

2.**Students rate their courses first and foremost on aspects of good teaching**!!! Still, the contribution is not perfect and other issues apply (those related to the course, students' un-agreeable intentions). However, their effects are of a much lower effect.



Do SRIs correlate with **other acceptable measures for effective teaching**?

Research shows:

Trained classroom observers

Results of SRIs are comparable to those made by trained observers (Murray, 2007)

Alumni (retrospective ratings)

High correlations were found between the retrospective ratings by former students and by those of currently enrolled students (Benton, 2014)

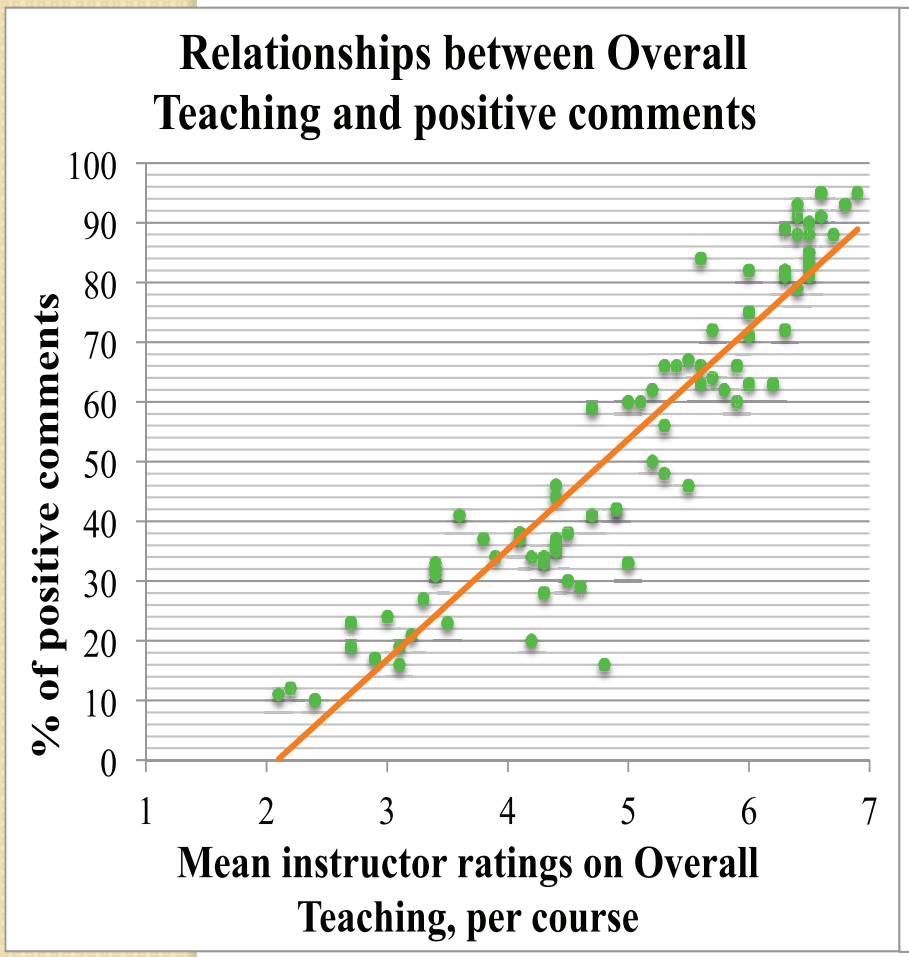
Written comments (Section C)

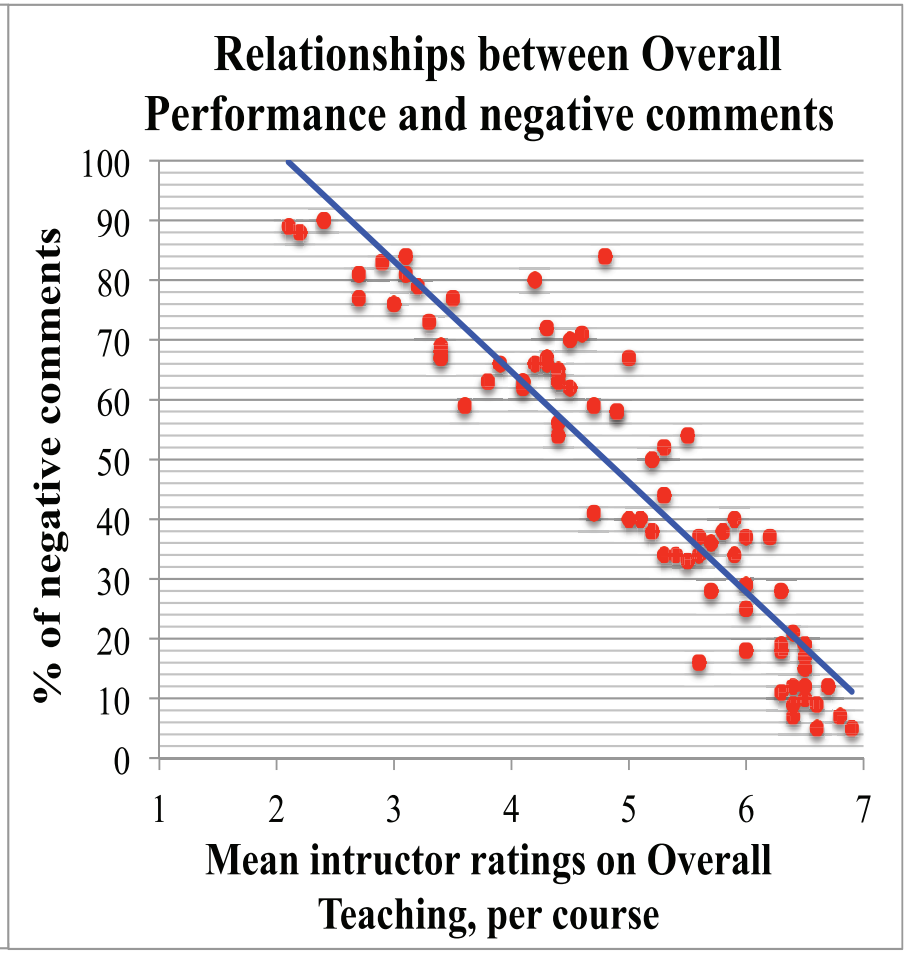
Are they related to SRIs?

C. Written comments

What do we know about the reliability and validity of written comments? Can they serve as a tool for teacher evaluation?

My study at TAU: The higher rated teachers on SRIs get more positive comments and less negatives





C. Written comments

- % of positive comments highly correlate with SRIs
- The content of written comments is highly congruent with that of teacher's numerical ratings on main behaviors (Alhija & Fresko 2009; Benton, 2014)
- There may be some idiosincretic comments that do not agree with the numerical comments

Conclusion: Written comments are soundly valid but should not be solely used in personnel decision-making. No research evidence for reliability. They may highly contribute to faculty for improving instruction and for faculty developers for helping facuty improve

① Sum-up conclusion: SRIs do measure indirectly effective teaching

This measurement is through proxies:

- Main behaviors of effective teaching
- Other acceptable measures of effective teaching

② Do SRIs relate to student learning?

We should even not expect for SRIs to measure student learning, because:

a. Learning consists of a variety of aspects, some of which are non-measurable

b. No agreed upon good method to measure even content learning (exams?)

c. Students learn the course content also from sources other than the teacher

② Do SRIs relate to student learning?

b. No agreed upon good method to measure student learning

Proxies:

- scores on the final exam
- students' perceptions of their learning

②Do SRIs relate to scores on the final exam?

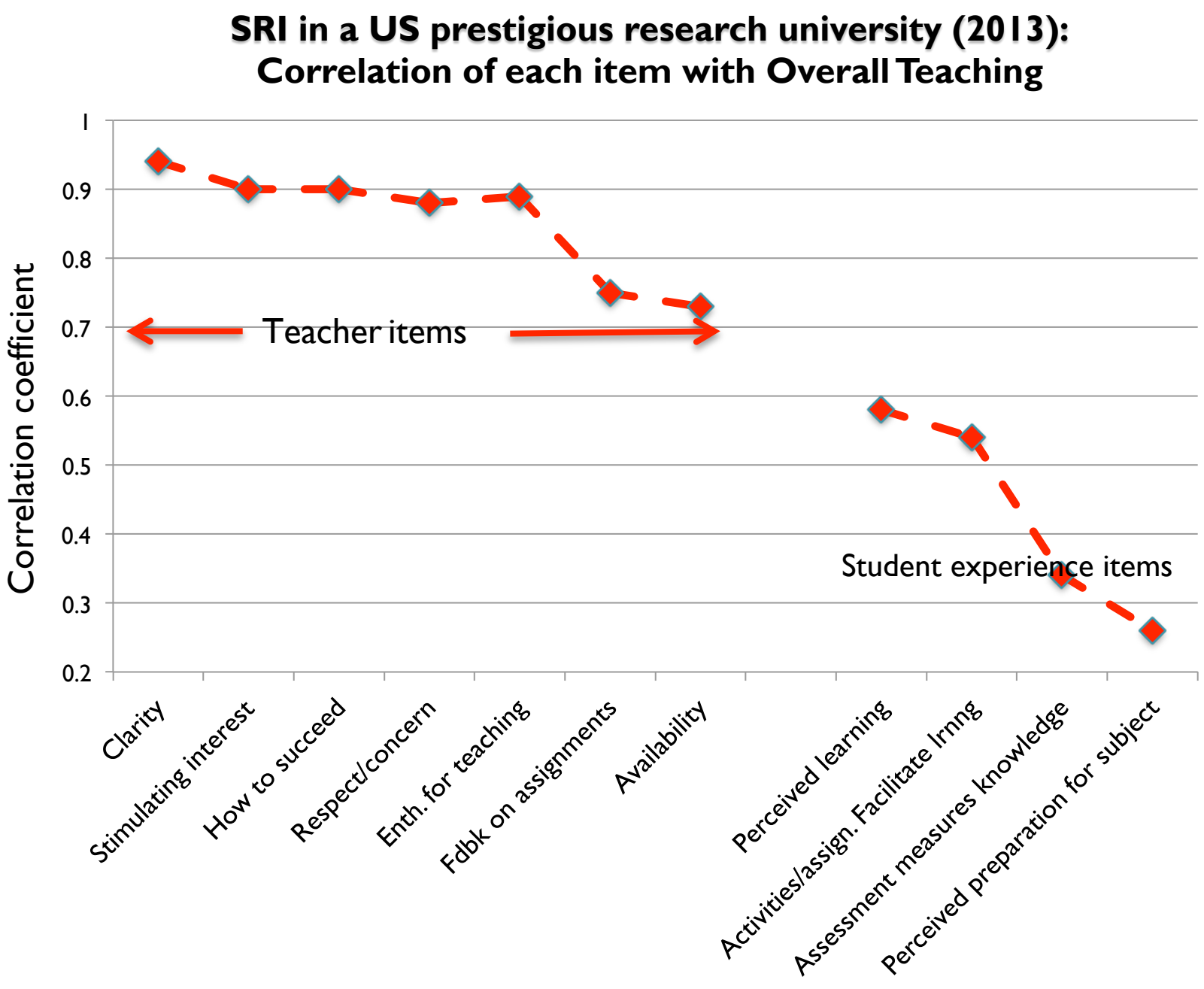
Meta-analysis of multi-section studies—low-to-moderate relationships between SRIs and scores on the final exam

Cohen (1980)

②Do SRIs relate to students' perceptions of their learning?

- Research evidence: Only medium-to-low correlations between SRI and students' self ratings of the amount they have learned in the course

Benton (2014)


SRI in a US prestigious research university (2013):
Correlation of each item with Overall Teaching
Correlation coefficient
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
Teacher items
Student experience items
Clarity
Stimulating interest
How to succeed
Respect/concern
Enth. for teaching
Fdbk on assignments
Availability
Perceived learning
Activities/assign. Facilitate lrnng
Assessment measures knowledge
Perceived preparation for subject

② Conclusion:

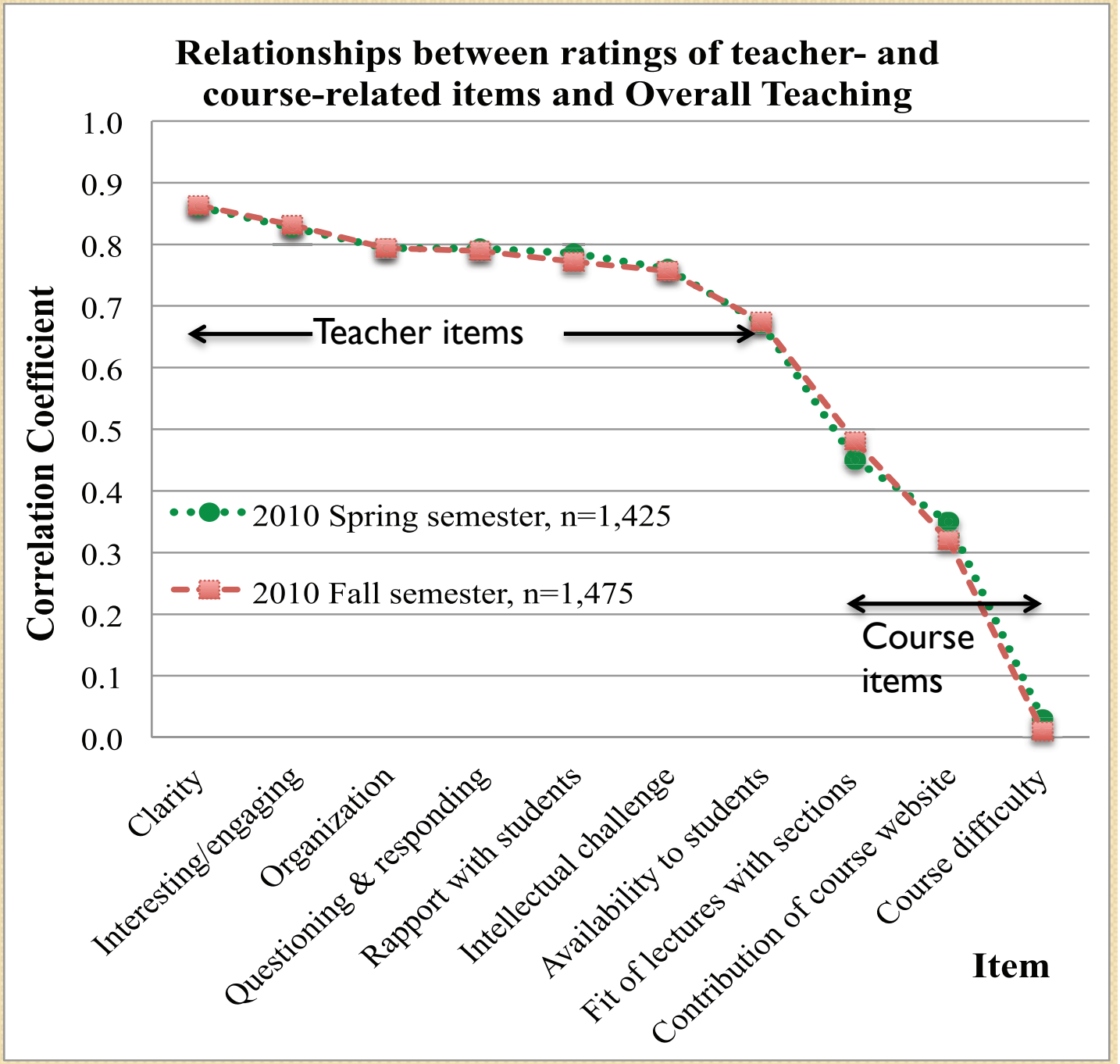
Evidence for only non-substantial relationships between **SRIs** and **student learning**

This is what should be!

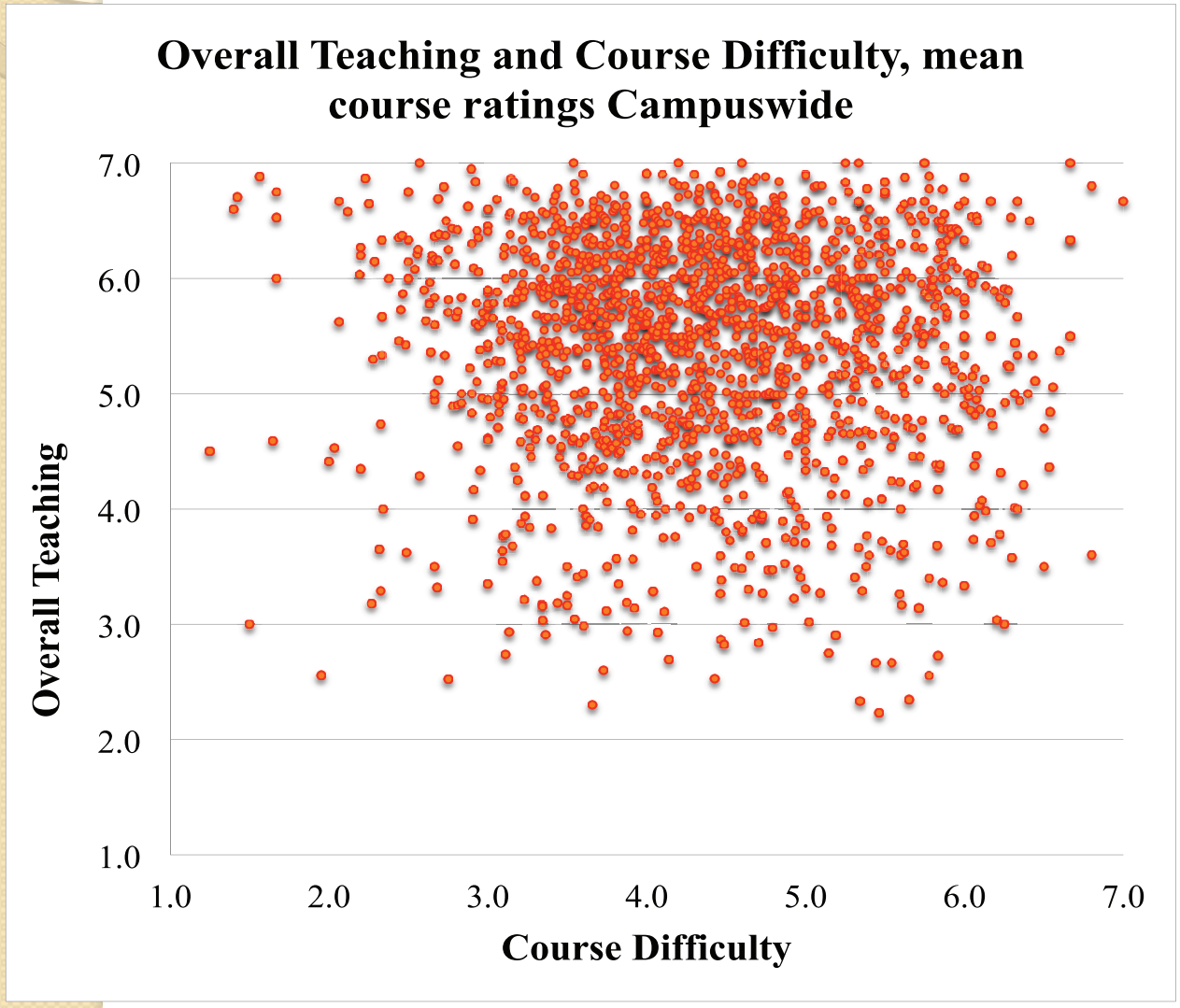
Do SRIs relate to the following factors:

③ Course difficulty/workload?
④ Expected/actual grades?
⑤ Instructor expressiveness/enthusiasm?
⑥ Instructor popularity
⑦ Class size?
⑧ Discipline?
⑨ Other extraneous factors: Instructor gender/race/ physical attractiveness/academic degree, course components, perceived learning?

③ Do SRIs relate to course difficulty/ workload?



③ Do SRIs relate to course difficulty/ workload?



General research: no/low relationships

TAU: All TAU undergraduate lecture courses (n = 1,770), Fall Semester 2007, **r =-0.02**

No relatioships

③ Conclusion: SRIs are almost not related to course difficulty/ workload

No relationships between perceived course difficulty and global teacher ratings!

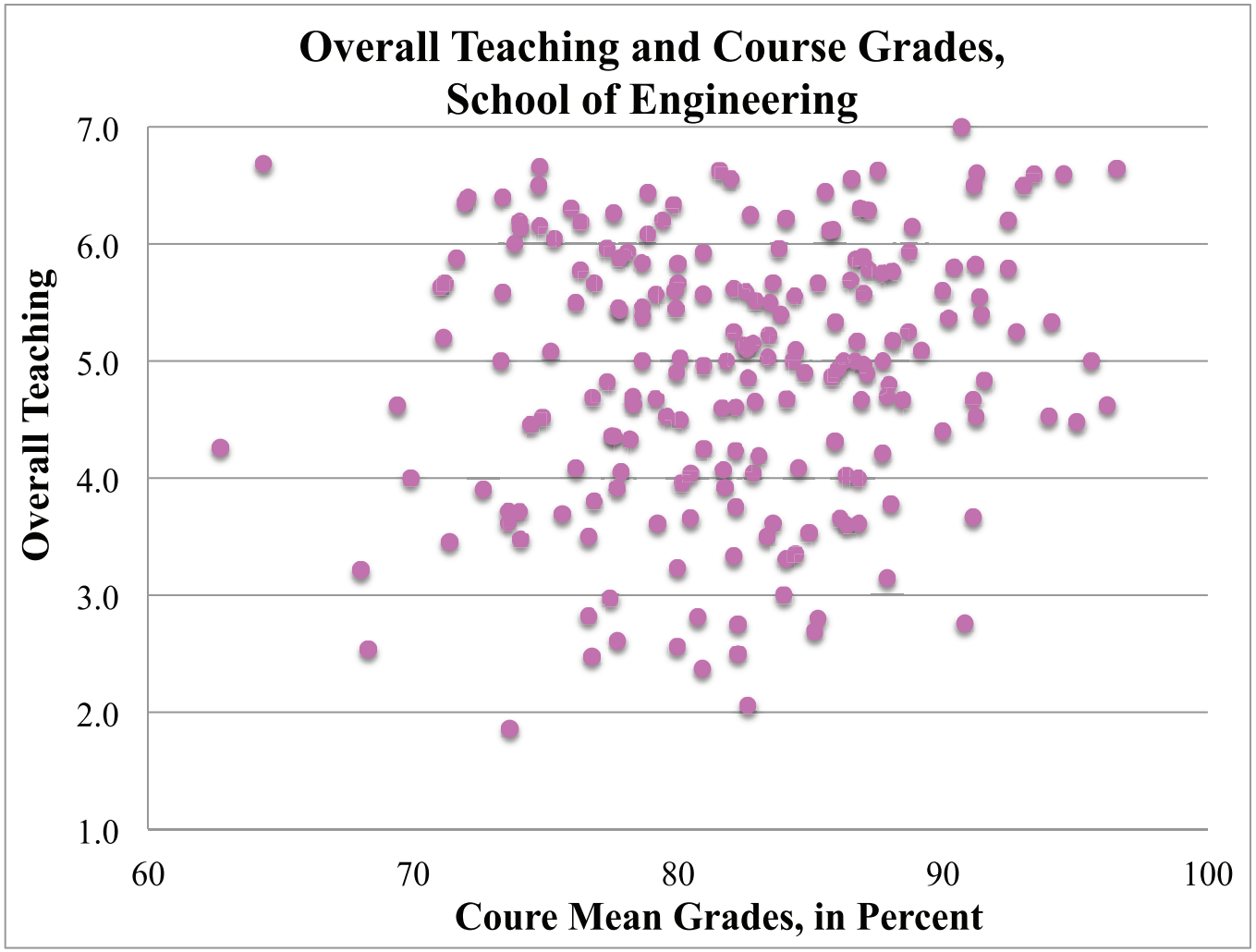
④Do SRIs relate to expected/actual grades?

General research: Almost no relationships

Course actual (final) grades & global item

School of Engineering, Fall Semester 2005, 205 courses, **r=0.13**.

No relationships for grade with instructor global ratings!



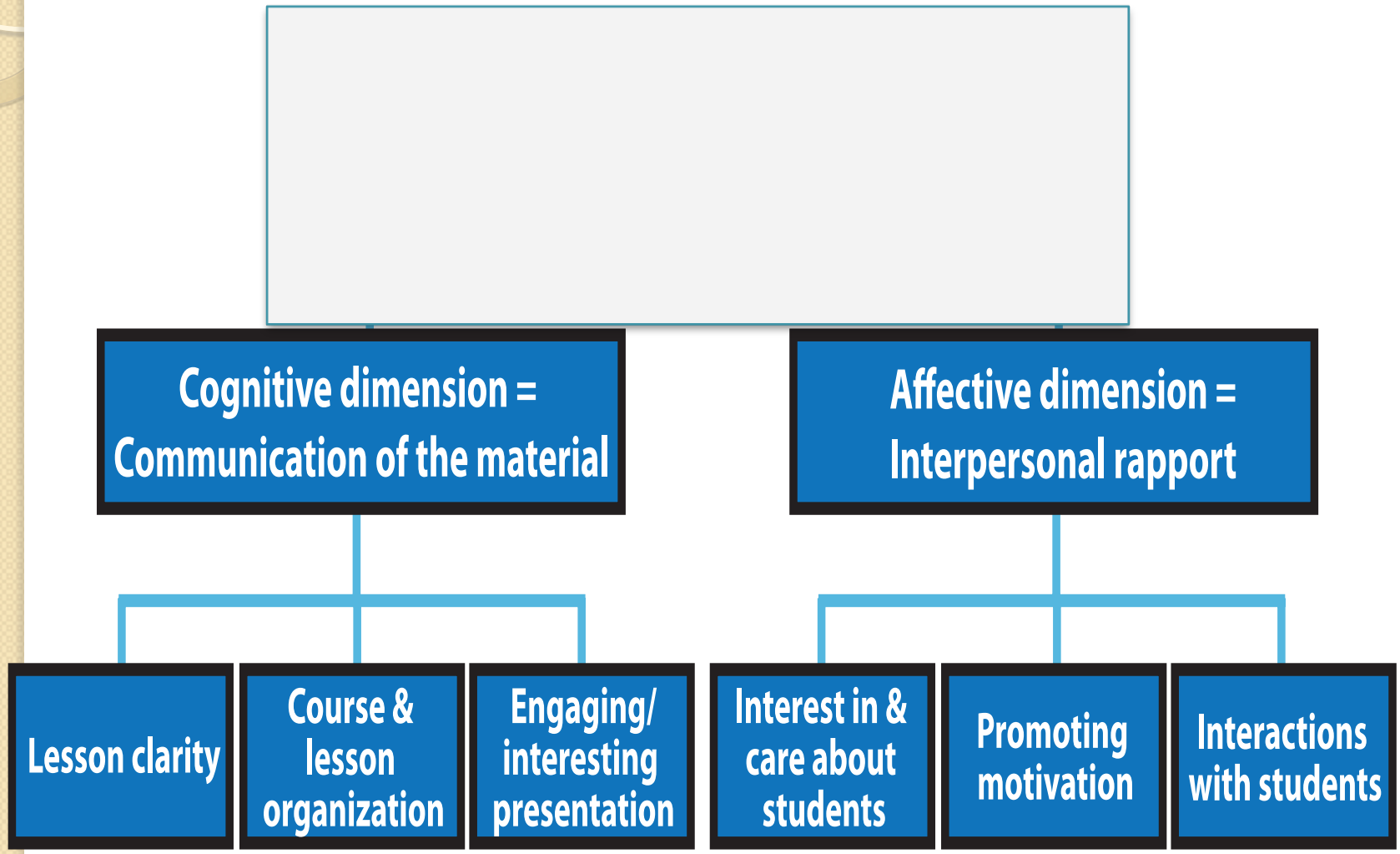
④ Conclusion: SRIs are not related to expected/actual grades

⑤Do SRIs relate to instructor expressiveness/enthusiasm?

Instructor expressiveness/enthusiasm is a major component of one of the main teaching effectiveness behaviors:"engaging/interesting presentation", that is, it does contribute to teaching effectiveness.

Conclusion: **SRIs are related to instructor expressiveness/ enthusiasm**

What are the main classroom behaviors of effective teaching?



⑥Do SRIs relate to instructor popularity?

What is instructor popularity?

Gaming? Entertaining students and fulfilling their wishes beyond what is needed to promote learning(?)

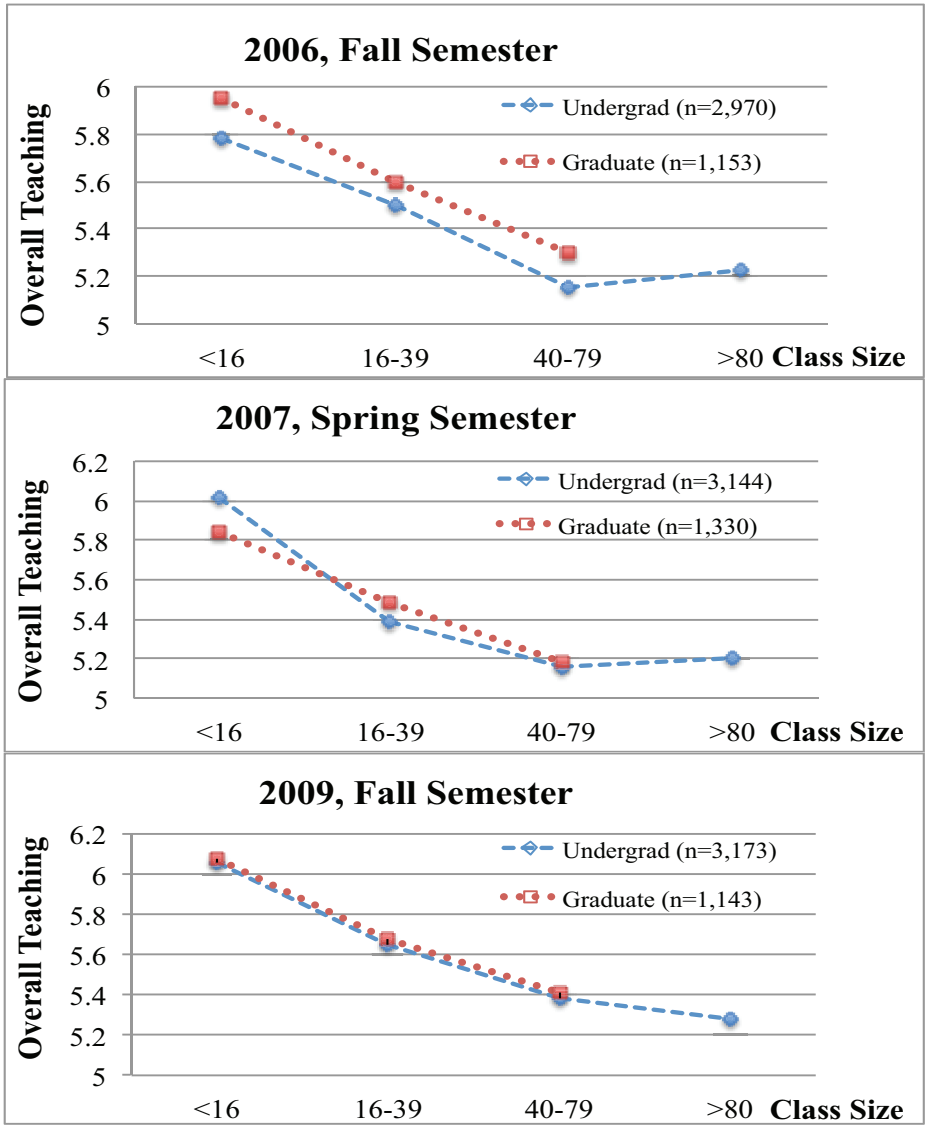
Haven't found any research evidence (except of Dr. Fox studies) that **SRIs are related to instructor popularity**

⑦Do SRIs relate to class size?

General research: Unclear mean results with tendency to inverse relationships

Inverse relationships

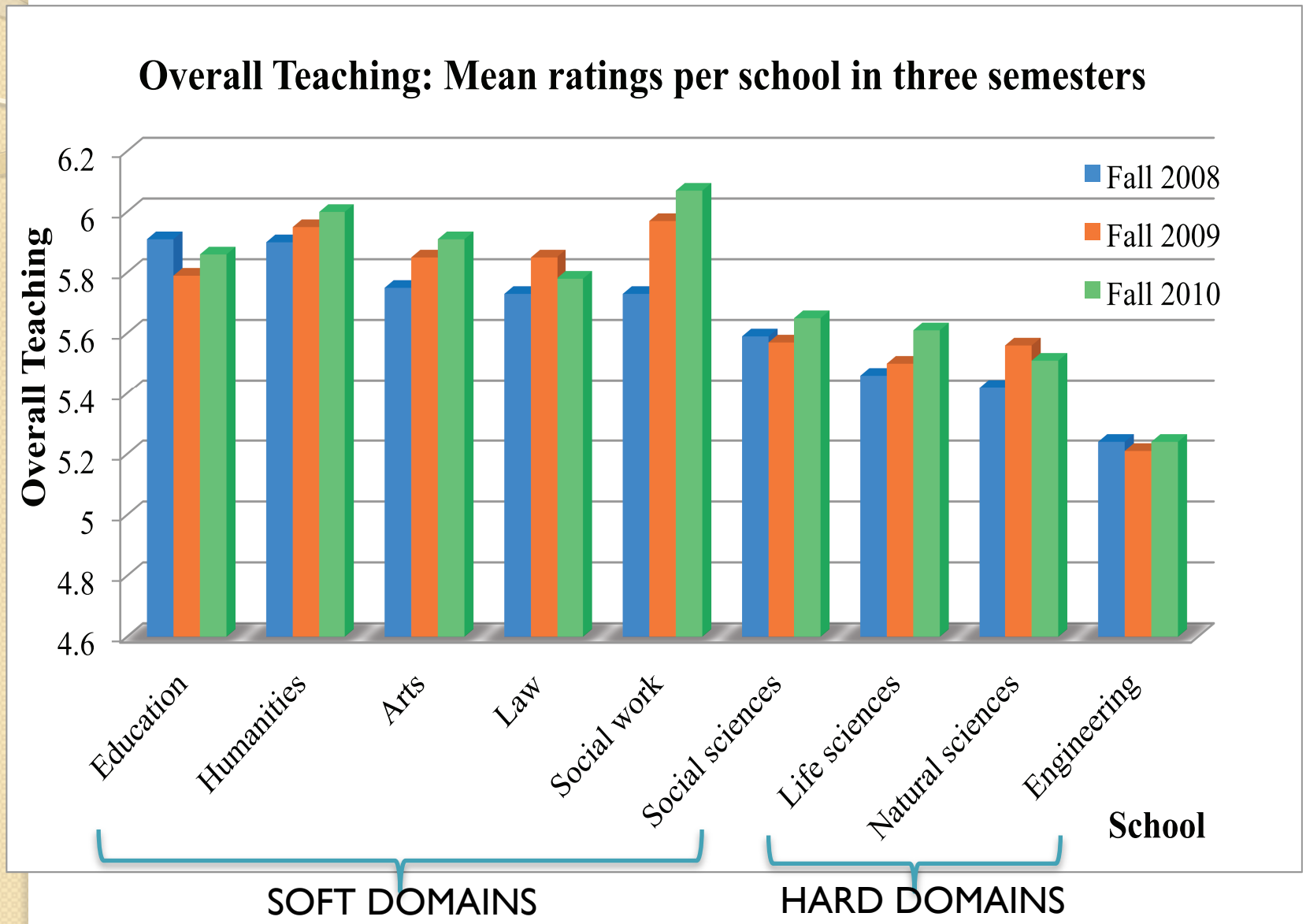
except for very large classes



⑦ Conclusion: SRIs are inversely related to class size

up to a certain (large) class size

⑧ Do SRIs relate to discipline?



⑧SRIs do relate to discipline—
hard disciplines (mostly STEM) are rated lower than soft disciplines

⑨ Do SRIs relate to a variety of extraneous factors?

Extraneous factors: Those that **do not relate to quality of teaching or to student learning**, and still are believed to affect SRIs and thus bias their validity

Faculty: academic rank, age, gender, years of teaching experience, personal characteristics (other than enthusiasm or caring), physical attractiveness, with Asian accent, research productivity

Students: age, gender, personality

Course: the time of day it is offered, length of class meetings, number of rows in the classroom

⑨ Do SRIs relate to a variety of extraneous factors?

General research: All these factors were found in most studies to be almost unrelated or only low related to SRIs

None has been established as related to/biasing SRIs!

Summary: Do SRIs relate to/correlate with?

① Effective teaching? **YES**

② Student learning? **NO EVIDENCE (NO EVIDENCE IS POSSIBLE)**

③ Course difficulty/workload? **NO**

④ Expected or actual grades? **NO**

⑤ Instructor expressiveness/enthusiasm? **YES**

⑥ Instructor popularity? **NO EVIDENCE**

⑦ Class size? **YES—INVERSE RELATIONSHIPS UP TO A CERTAIN CLASS SIZE**

⑧ Discipline? **YES—SOFT DISCIPLINES ARE RATED HIGHER THAN HARD DISCIPLINES**

⑨ Other extraneous factors: Instructor gender/race/physical attractiveness/ academic degree, course components, perceived learning? **NO GOOD EVIDENCE, AND IF YES, ONLY WITH A MINOR EFFECT**

B. SRI validity—faculty concerns and myths, and refuting them?

Major claims for biasing factors:

Source: A heated discussion titled "Do student evaluations measure teaching effectiveness?" during first months of 2014 over the Linkedin group "Higher Education Teaching and Learning".

Participants: international faculty and administrators, mostly from the US.

Students are not capable/eligible for rating their teachers, rating can be valid only in retrospect

- Students cannot make consistent judgments about instructors, instruction, and teaching effectiveness due to their immaturity, inexperience, and capriciousness.
- Students cannot recognize effective teaching by the end of the course, until called upon to apply course materials in further coursework or after graduation; SRIs of low-rated teachers dramatically improve when students are resurveyed years later and realize these teachers' contribution to their learning.

Ratings are unrelated to good teaching or student learning

- There is no direct correlation between student satisfaction and teaching quality. This is evident from the fact that students show the highest satisfaction with courses that support a surface learning approach—what the student "needs to know" to get good grades.
- SRIs measure consumer/customer satisfaction, instructor popularity and gaming—courses that are entertaining/showy, enjoyable and fun are rated the highest
- Student evaluations reflect customer satisfaction and class enjoyment. They do not consider other important elements such as teaching quality.
- Student ratings is not a reliable method to evaluate the quality of teaching, because it depends upon students' capacity and willingness to learn, as well as their level of intelligence and study habits and abilities

Difficulty/workload, grading

- Student ratings are popularity contests that measure instructor expressiveness or style rather than the substance of teaching.
- Instructors who teach more difficult courses are punished by lower ratings. Instructors who lower the level of difficulty or assign a low workload are rewarded by higher ratings.
- Students are more happy with easy courses and lenient demands that make them feel successful in their learning, than by challenging courses that force them to really think and work hard and learn. In reality, most students want showy and easy courses and this is reflected in their ratings.
- Instructors who assign high grades are rewarded by higher ratings.

SRIs damage good teaching

- Student ratings tempt teachers to 'dumb down' the content and give high undeserved grades in order to be liked and be evaluated positively
- SRIs promote grade inflation and faculty gaming, and they lower the level of teaching and of demands from students. All in all, they lead to the further decline of the higher education system.
- Some ineffective teachers manage to get good ratings by playing along with students' intellectual laziness.
- It is fairly easy for professors to manipulate student ratings. Just before the time for student evaluations, some faculty show movies in class and provide food, bonuses, and other incentives to students.

Written comments

- Anonymity encourages students to write unfounded and irresponsible comments, even untruths.
- Written comments are inconsistent and therefore unreliable. Often students in the same class write contradictory comments. This is very frustrating since there is no way to satisfy everyone.
- Students mostly write negative comments

The use of SRIs in decision-making may ruin faculty careers

The erroneous practice of using SRIs to evaluate teachers has ruined the careers of many faculty members. It's a real shame!

Why are these named "Myths" and why do they persist?

MYTH TITLE: Despite the impressive research support for SRI reliability and validity, unsubstantiated claims of potential bias continue to flourish and to win faculty trust. These claims that are refuted by research, are titled "Myths" in the domain literature

LONG EXISTANCE: These claims exist for almost 100 years:

Aleamoni, 1999). Student rating myths versus research facts from 1924 to 1998

WHY THEY PERSIST? The lack of faculty familiarity with SRI research has often generated substantial objection to the use of SRI these beliefs and myths come from a lack of understanding or misinterpretation of students' learning needs, and from generalizations of negative personal experiences

Contemporary arguments for diminished reliability and validity of SRI

a) Arguments claiming to have identified extraneous factors that bias SRI results.

b) Arguments claiming that contemporary/millennium students are very different from students of even one or two decades ago, so that SRI validity and reliability studies of the past are no longer relevant.

c) Arguments claiming that contemporary institutional/administrators' response to market pressures has changed faculty incentives and hence teaching behavior, and consequently has reduced the reliability and validity of student ratings.

What does it mean "potentially biasing factors"?

A factor is of potential bias to SRI vaidity if:

- It **does** relate to/correlate with SRIs
- It **does not** contribute to effective teaching or to student learning (and therefore should not be related to SRIs—to student satisfaction from teaching…)

What can we conclude about SRI validity?

A biasing factor:	Correlate with SRI YES	Contribute to effective teaching student learning NO
• Course difficulty/workload	NO	
• Expected/actual grades	NO	
• Instructor expressiveness/ enthusiasm	YES	YES
• Instructor popularity	No evidence	NO
• Class size	YES	YES
• Discipline	YES	YES
• Extraneous factors	No evidence	NO

Conclusion

No sound evidence for the listed factors to be biasing factors

Danger!!! Teachers who believe that lowering level of course difficulty/workload and that grading leniently will increase their ratings may try to "bribe" students by lowering the level of the course.

To conclude: Are SRIs valid? Citation by the leading SRI researcher (Marsh 2007, p. 319)

- Ratings are multidimensional, reliable and stable;
- Primarily a function of the instructor who teaches a course rather than the course that is taught;
- Relatively valid against a variety of indicators of effective teaching;
- Relatively unaffected by a variety of variables hypothesized as potential biases;
- Seen to be useful by faculty as feedback about their teaching, by students for use in course selection, and by administrators for use in personnel decisions

Marsh, H. W. (2007). Students' evaluations of university teaching: dimensionality, reliability, validity, potential biases and usefulness. In R. P. Perry & J. C. Smart (Eds.), The scholarship of teaching and learning in higher education: An evidence-based perspective (pp. 319-383). Dordrecht, The Netherlands: Springer.