Jacob Annotated Bib

[Students' Evaluations of Instruction: A Longitudinal Study of Their Stability](https://psycnet.apa.org/record/1980-24186-001) (1980)

RQ: What is the long-term stability of students’ evaluations of instructional effectiveness?

* Investigated the correspondence between students' end-of-term and retrospective assessments of their courses and instructors.
* n = 1,374 (under)graduate business students from 100 different classes between 1974-1977.
* IVs:
  + Time taken since original SAI; in this case, 1 year after graduation or 1-3 years after having taken any given course
  + Type of course taken; in this case, 11 graduate courses and 8 undergraduate courses designated as either quantitative (i.e., accounting, economics, finance, etc.) or nonquantitative (i.e., legal, management, organizational theory, etc.)
* DVs:
  + Student ratings of instructional effectiveness were obtained at the end of the term through use of the 28-item Student Assessment of Instruction (SAI) Questionnaire
  + Retrospective instructor ratings were based on graduates' responses to an abbreviated 9-item SAI questionnaire.
* Results showed large and statistically significant correlations (median *r =* .83) between end-of-term and retrospective ratings. These findings demonstrate that students' evaluations collected at the end of a course are remarkably similar to the retrospective ratings provided by the samestudents several years later.

[The Measurement Invariance of University Students’ Ratings of Instruction](http://repository.bilkent.edu.tr/bitstream/handle/11693/53077/The_measurement_invariance_of_university_students’_ratings_of_instruction.pdf?sequence=1&isAllowed=y) (2019)

RQ: What is the invariance in the scores of student ratings of instruction (SRI) across high and low achieving classrooms?

* Investigated the invariance levels of SRIs across groups defined with respect to students’ expected grades and their end-of-semester grades through four levels of invariance requirements (configural, weak, strong, and strict). The measurement invariance was studied through the following steps: invariance of factor patterns (configural), factor loadings (weak), intercepts (strong), and variances of residuals of observed variables (strict).
* n = 9,230 Turkish (under)graduate students
* IVs:
  + Students’ perception about their learning level (based on level of agreement with the statement, “I learned a lot in this course”)
  + Students’ end-of-semester grades
* DVs:
  + Ratings on a 9-item SRI scale (7 items actually used in analyses)
* Differences in item intercepts across different groups indicate that for the same instructor or classroom practice, different achievement groups may have different perceptions – thus, the ratings of these groups are different. In other words, the same degree of instructional effectiveness is rated differently by the students from two scoring groups (low achieving students rated the same instructional performance more positively than the high achievers).

[How reliable are students’ evaluations of teaching quality? A variance components approach](https://www.tandfonline.com/doi/pdf/10.1080/02602938.2016.1261083) (2017)

RQ: Do the variance components of teachers, courses, and students differ systematically between different aspects of teaching quality?

* Investigated the inter-rater reliability of the different scales and global ratings included in the questionnaire, the variance components that contribute to reliable teacher-course scores, and the variance components that are due to students and the interactions between students and teachers.
* n = 4,224 evaluations of psychology courses conducted by 480 students between the winter semester 2011 and the summer semester 2014 at the University of Kassel, Germany
* IVs: 3 different models
  + Model 1: IV was the course taken
  + Model 2: IVs were the course and teacher taken
  + Model 3: IVs were the course and teacher taken as well as the degree of teacher-student interaction
* DVs:
  + Ratings on either a 31-item (lecture) or 34-item (seminar) evaluation questionnaire that addressed teacher planning/presentation, interaction with students, interestingness, difficulty, and general rating of teacher performance
* Results showed that teachers and courses were essential sources of variance for all four facets of teaching quality examined in this study as well as for the overall ratings of courses and teachers. A considerable proportion of variance (~20%) was also explained by students. However, in five of the six measures, the proportions of variance explained by teachers and courses together was nearly twice the proportion of variance explained by students.

[Student evaluation of instruction in higher education: exploring issues of validity and reliability](https://www.tandfonline.com/doi/pdf/10.1080/02602938.2010.523819) (2012)

RQ: Is the student evaluation of instruction (SEI) a valid and reliable instrument for evaluating faculty instruction at The Ohio State University?

* Examines the construct validity and internal consistency reliability of the SEI used at The Ohio State University after 11 years of implementation.
* n = 73,500 completed SEIs by undergraduate students in the fall semester of 2005 (may represent multiple responses from the same student for different courses)
* IVs:
  + Ratings/responses provided on a 10-item SEI
* DVs:
  + Construct validity of the SEI assessed via structural equation modeling
  + Reliability of SEI assessed via the internal-consistency reliability index
* The SEI items correlated highly with instructional effectiveness (r = .70 to .76), accounting for between 49% and 58% of the total variance in instructor effectiveness. The internal-consistency reliability for the questionnaire was high (r = .95). Hence, only about 5% of the variability in the responses can be attributed to error.

[The number of feedbacks needed for reliable evaluation. A multilevel analysis of the reliability, stability and generalisability of students’ evaluation of teaching](https://www.tandfonline.com/doi/pdf/10.1080/02602938.2011.625471) (2013)

RQ: How many student feedbacks are necessary for a reliable student evaluation of instruction?

* Partitioned the total variance of students’ ratings into that which can be attributed to teachers, courses, implementations, students, and teacher by course interaction.
* n = 11,904 feedback forms collected from 1,585 students between 2003 and 2006 at HAAGA-HELIA University of Applied Sciences
* IVs:
  + 6 models different including/excluding variance attributed to previously mentioned criteria as measured through a multilevel analysis
* DVs:
  + Sample size needed for reliable assessment of teaching
* To assess a teacher’s general teaching effectiveness, one needs to evaluate four randomly chosen methods of implementing the course. Two implementations are needed when one course is evaluated, and if one implementation is evaluated, up to 15 feedbacks are needed.

[Do University Teachers Become More Effective With Experience? A Multilevel Growth Model of Students’ Evaluations of Teaching Over 13 Years](https://www.researchgate.net/profile/Herb-Marsh/publication/232498766_Do_University_Teachers_Become_More_Effective_With_Experience_A_Multilevel_Growth_Model_of_Students%27_Evaluations_of_Teaching_Over_13_Years/links/5bdfa69292851c6b27a7927a/Do-University-Teachers-Become-More-Effective-With-Experience-A-Multilevel-Growth-Model-of-Students-Evaluations-of-Teaching-Over-13-Years.pdf) (2007)

RQ: Do university teachers, like good wine, improve with age? (ok but the real question is, “Do ratings of the same teachers collected over a 13-year period systematically increase, decrease, or remain stable over time (mean level stability)?”)

* Applied a multiple-level growth modeling approach to the long-term stability of students’ evaluations of teaching effectiveness (SETs).
* n = 195 teachers evaluated on an average of 30.9 classes per teacher
* IVs:
  + The teacher average rating, course level (undergrad/grad), the year rating was collected, and if the teacher was early in his/her career; these 3 criteria were used in 7 models processed through a multilevel analysis
* DVs:
  + Class-average rating taken from SETs based on the SEEQ instrument
* There was little evidence that teachers became either more or less effective with added experience. Whereas there were substantial individual differences between teachers in terms of their teaching effectiveness, these individual differences were also highly stable over time.

[Student evaluation of teaching in the virtual and traditional classrooms: A comparative analysis](https://www.sciencedirect.com/science/article/pii/S1096751605000795?casa_token=owvHQ6yINLsAAAAA:2ME7h6t19I5mfuSD_l07MtPrPaFmewNIlJ71pBZS77Ei7iGXR0KmIldRZgScVPHLnsCvH0FfEg) (2006)

RQ: What are the major themes of student responses to open-ended SET questions in online and face-to-face higher education courses? How do the evaluations differ between online and on-campus courses?

* Examined SET comments obtained from required research design and statistics courses in education and school counseling doctoral programs.
* n = 202 doctoral students who contributed SET responses over 4 years of the same 3 teachers
* IVs:
  + The students’ enrollment in either a traditional or fully online course
* DVs:
  + Students’ SET comments later reviewed through content analysis
  + Themes discernable from responses (i.e., organization & planning, communication & clarity, etc.) as deduced from content analysis
* Results provide evidence that doctoral students tend to evaluate online courses differently and more negatively than on-campus courses. Additionally, substantial within-class variance in evaluations was noted with some students praising one aspect of the online course and other students criticizing the very same aspect.

[Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related](https://www.sciencedirect.com/science/article/pii/S0191491X16300323?casa_token=JegULRkMyFIAAAAA:3usWXpjCGSEYlUTzCBGGa1E3N2PxgOnB228rGmFinQo7ociCMwQOm1288R98OW3-CUIqnyovsg) (2017)

RQ: Do previous studies of SET/learning correlations provide valid explanations despite their flaws/limitations?

* Re-examined the evidence for the SET/learning correlations and for the extant claims that SETs are valid measures of professors' teaching effectiveness rather than measures of student satisfaction.
* n = 51 articles observing SET and learning/achievement in college; contained 97 multisection studies
* IVs:
  + Ratings and other responses collected from three previous meta-analyses of SET responses
  + Controlled vs. uncontrolled prior learning within studies
  + A bunch of metadata info (i.e., course name, number of sections, instructor experience, etc.) as well as study quality info (i.e., means, SDs, distributions, etc.)
* DVs:
  + Average SET/learning correlations as determined from data aggregation & the random effect model
  + SET/learning correlation strength
* Findings reported in previous meta-analyses ([Clayson, 2009](https://www.sciencedirect.com/science/article/pii/S0191491X16300323?casa_token=JegULRkMyFIAAAAA:3usWXpjCGSEYlUTzCBGGa1E3N2PxgOnB228rGmFinQo7ociCMwQOm1288R98OW3-CUIqnyovsg" \l "bib0085), [Cohen, 1981](https://www.sciencedirect.com/science/article/pii/S0191491X16300323?casa_token=JegULRkMyFIAAAAA:3usWXpjCGSEYlUTzCBGGa1E3N2PxgOnB228rGmFinQo7ociCMwQOm1288R98OW3-CUIqnyovsg" \l "bib0100), [Feldman, 1989](https://www.sciencedirect.com/science/article/pii/S0191491X16300323?casa_token=JegULRkMyFIAAAAA:3usWXpjCGSEYlUTzCBGGa1E3N2PxgOnB228rGmFinQo7ociCMwQOm1288R98OW3-CUIqnyovsg" \l "bib0170)) are an artifact of poor meta-analytic methods; additionally, students do not learn more from professors with higher SETs.

[Attractiveness, easiness and other issues: student evaluations of professors on Ratemyprofessors.com](https://www.tandfonline.com/doi/abs/10.1080/02602930601122803)(2008)

RQ: Do American college students who post professor evaluations at Ratemyprofessors.com consider courses to be of high quality when the professor is attractive and the course is easy?

* Examined whether social and personal factors not directly connected to educational achievement are involved in SET; replication of earlier studies with larger database
* n = 6,852 professors from 369 institutions in the United States and Canada with at least 20 ratings on Ratemyprofessors.com
* IVs:
  + Scores from website grading course easiness, professor helpfulness, clarity, overall quality, and hotness
  + Course discipline
* DVs:
  + Correlations between individual variables
    - Subdivided by discipline as well
* Results indicated strong positive correlations between Quality and Easiness and between Quality and Hotness in addition to significant differences between departments and institutions.

[The Fair Process Effect in the Classroom: Reducing the Influence of Grades on Student Evaluations of Teachers](https://journals.sagepub.com/doi/pdf/10.1177/0273475318772618)(2019)

RQ: Does distributive justice perception mediate the relationship between grades and SET?

* Examined the role that student perceptions of fairness may play in the relationship between expected course grades and SET.
* n = 250 students from 3 different public universities
* IVs:
  + SET responses
  + Grade
  + Distributive justice perception as measured via the Chory-Assad and Paulsel classroom justice scale
  + Procedural justice perception as measured via the Chory-Assad and Paulsel classroom justice scale
  + Qualitative assessment of fairness determined via one open-ended question
  + Student/teacher demographic information
* DVs:
  + Correlations between SET responses, grades, and perceptions of distributive/procedural justice
* Distributive justice perception partially mediated the relationship between grades and SET, and procedural justice perception moderated the relationship between distributive justice perception and SET. Additionally, procedural justice perception moderated the indirect relationship between grades and SET through distributive justice perception, such that only when procedural justice perception was low did grades predict SET via distributive justice perception.

[Effects of Course and Instructor Characteristics on Student Evaluation of Teaching across a College of Engineering](https://onlinelibrary.wiley.com/doi/pdf/10.1002/jee.20013) (2013)

RQ: How and do course and instructor characteristics affect SETs?

* Examine the relationship between course and instructor characteristics and SETs.
* n = ≥137,431 student responses from 3,938 courses taught over 7 consecutive semesters
* IVs:
  + SET score
  + Class size
  + Course level
  + Course requirement/elective
  + Instructor experience
  + Instructor gender
  + Instructor academic rank
  + Actual/expected grade
* DVs:
  + Correlations between variables as determined via Fisher’s z and ANOVAs
* Results showed statistically significant positive correlations between course grades and SET. Instructors receiving poor fairness in grading scores had higher correlations between SET and average course grades.

[A Study of the Correlation of the Improvement of Teaching Evaluation Scores Based on Student Performance Grades](https://files.eric.ed.gov/fulltext/EJ1140507.pdf) (2017)

RQ: Do teaching evaluations influence teachers into giving higher grades in order to get higher teaching evaluation scores?

* Analyzed the correlations between teaching evaluation scores, student’s final grades and course fail rates, and examined whether students’ final scores and course fail rates are important predictors of teaching evaluation scores.
* n = 412,905 student responses (holy moly) from 9,147 courses at a humongous Taiwanese university
* IVs:
  + SET scores
  + Students’ final grades (including average class grades)
  + Course fail rate
* DVs:
  + Correlations between the aforementioned variables as determined via Fisher’s z and Pearson’s r
* Results showed that both student’s final grades and course fail rates are predictors of teaching evaluation scores. There was a positive correlation between teaching evaluation scores and students’ final grades and a negative correlation between teaching evaluation scores and course fail rates.

[Students’ perceptions of the evaluation of college teaching](https://www.emerald.com/insight/content/doi/10.1108/EUM0000000006158/full/html) (2001)

RQ: What variables do students use to evaluate professors?

* Surveyed student perceptions to provide evidence of inherent weaknesses in the use of SETs to measure and report teaching effectiveness accurately.
* n = 530 undergrad/grad students enrolled in accounting courses over two years at a large public university
* IVs:
  + Student gender
  + Student classification
  + Course grade expectations
    - Differences for all analyzed via an ANOVA
* DVs:
  + SET ratings
* Findings revealed 42 percent of the students would punish instructors for being asked embarrassing questions, 28 percent for being graded hard, 27 percent for pop quizzes, 20 percent for significant homework, 14.5 percent for using humor, 14.3 percent for use of untyped overheads, 11 percent for being called on, and 9 percent for merely grading homework.

[Determinants of Student Evaluations of Global Measures of Instructor and Course Value](https://journals.sagepub.com/doi/pdf/10.1177/0273475300222005?casa_token=5SdSizNwN8MAAAAA:Jz1jHNGYCg1ObF_pK8N6aoF5A3SSjpEtbyhltEv7CDNyTwnd7S1a7-9tmD_ccMp4VA1vYWwxxdrc) (2000)

RQ: Even though the questions on evaluations have face validity and a modicum of methodological support, are the students responding to the actual questions or to other contaminating variables?

* Determined the influence of contaminating factors and tested a likely sequence in the formation of an overall evaluation of the instructor and course.
* n = 700 students enrolled in business courses at a regional midwestern university
* IVs:
  + Course organization
  + Workload/difficulty
  + Expected/fairness of grading
  + Liking/concern of professor
  + Instructor knowledge
  + Perceived learning
* DVs:
  + Perceived teaching effectiveness measured via a custom evaluation instrument
* Expected/fairness of grading is posited to be positively related to either overall course value or instructor rating. Instructors who exceed the norm for workload and difficulty will have their ratings more adversely affected than those who don’t.

[Student Evaluations and Consumer Orientation of Universities](https://www.tandfonline.com/doi/abs/10.1300/J054v08n01_04) (2008)

RQ: Are student evaluations an appropriate measure of teaching effectiveness, or can elements of the interaction between the students and the instructor produce summative student ratings of instructors?

* Explored the relationship of factors involved primarily with the instructor’s interaction with students, rather than the amount of learning of the students, with the overall evaluation of instructors.
* n = 116 student surveys of four different instructors at a business school assessing opinions of various aspects about the course and instructor
* IVs:
  + Willingness to help students
  + Competency
  + Credibility
  + Degree of caring
  + Degree of friendliness
  + Effectiveness in communicating
  + Reliability
  + Fairness in grading
  + Overall appearance
    - All assessed using a five-point scale
* DVs:
  + Scores of each attribute as designated by student
* Communication, fairness, and appearance had a statistically significant positive effect on overall instructor rating. These variables by themselves explained 63% of the total variance in overall rating of instructors.