

# Academic Course Evaluation

## Theoretical and Empirical Distinctions Between Self-Rated Gain in Competences and Satisfaction with Teaching Behavior

Edith Braun<sup>1</sup> and Bernhard Leidner<sup>2</sup>

<sup>1</sup>Freie Universität Berlin, Germany, <sup>2</sup>New School for Social Research, New York, NY, USA

**Abstract.** This article contributes to the conceptual and empirical distinction between (the assessment of) appraisals of teaching behavior and (the assessment of) self-reported competence acquirement within academic course evaluation. The Bologna Process, the current higher-education reform in Europe, emphasizes education aimed toward vocationally oriented competences and demands the certification of acquired competences. Currently available evaluation questionnaires measure the students' satisfaction with a lecturer's behavior, whereas the "Evaluation in Higher Education: Self-Assessed Competences" (HEsaCom) measures the students' personal benefit in terms of competences. In a sample of 1403 German students, we administered a scale of satisfaction with teaching behavior and the German version of the HEsaCom at the same time. Using confirmatory factor analysis, the estimated correlations between the various scales of self-rated competences and teaching behavior appraisals were moderate to strong, yet the constructs were shown to be empirically distinct. We conclude that the self-rated gains in competences are distinct from satisfaction with course and instructor. In line with the higher education reform, self-reported gains in competences are an important aspect of academic course evaluation, which should be taken into account in the future and might be able to restructure the view of "quality of higher education." The English version of the HEsaCom is presented in the Appendix.

**Keywords:** European higher-education reform, learning outcomes, questionnaire of academic course evaluation, self-assessment of competence-based learning, HEsaCom

### Introduction

The Bologna Process, with its planned reform of higher education throughout Europe, is a key element of higher educational policy. In order to accomplish the demanded comparability of degrees, their expected outcomes have to be clarified in terms of mediated and taught knowledge and competences (van der Wende, 2000). Various Bologna-related documents refer extensively to learning outcomes (Brennan, 2001, p. 4; cp. also Adam, 2004).

The aim of higher education is the facilitation of personal development and employment opportunities by training students in more competence-orientated ways. A full list of these competences has been formulated in a "framework of qualifications" (Bologna Working Group on Qualifications Frameworks, 2005). In the future, programs of study will be evaluated with respect to the competences a graduate is skilled in.

In light of the above, the goal of the present contribution is two-fold: (1) a theoretical and empirical distinction of the lecturer's behavior that is rather indirectly important for learning outcomes, and self-reported acquisition of competences; the "ultimate" criterion demanded by the Bologna Process, and (2), by showing that satisfaction with teaching behavior and competence acquirements can be assessed in

a distinct way, we advocate a new understanding of quality of higher education in terms of (outcome-oriented) competence acquisition rather than (process-oriented) satisfaction with teaching behavior.

### Currently Available Academic Course Evaluation Questionnaires

Up to now, currently available evaluation questionnaires that meet the criteria of classical test theory measure the design of academic courses and students' satisfaction with teaching (Aleamoni & Spencer, 1973; Burdsal & Bardo, 1986; Heckert, Latier, Ringwald-Burton, & Drazen, 2006; Marsh, 1984; Rindermann, 2001). Kolitch and Dean (1999) criticize typical student evaluation instruments that focus solely on teaching abilities; as a typical item shows: "The teacher was well-prepared for class" (Westermann, Spies, Heise, & Wollburg-Claar, 1998). So far, course evaluation instruments tend to ask the students for their opinion about the way a specific course was taught and their responses can be characterized as "numerical representations of satisfaction with teaching."

Consequently the HEsaCom (Evaluation in Higher Edu-

cation: Self-Assessment of Competences – in German: “Berlin Evaluation Instrument for Self-Reported Student Competences,” BEvaKomp) – has been developed (Braun, Gusy, Leidner, & Hannover, 2008). In the following, we will use the English name HEsaCom, although we actually used the German version for data collection. It measures the personal acquisition of competences from a specific course as assessed by the students themselves. It is the students’ benefits from the course that is important, rather than how much they liked the lecturer. It is likely that both liking the teacher and benefit are interrelated – but also different from each other.

## Using Self-Report Questionnaires

The purpose of course evaluation is to gain comparable results for different kinds of courses. The measures have to be applicable in different study programs as well as different types of courses. The economy of the surveys has to be extremely high, since many courses have to be evaluated within a short time period. Consequently, almost all course evaluations use self-report questionnaires. So does HEsaCom: Students are asked to rate the impact of each course on their gain in competences, and these ratings – on an aggregated level – are used to evaluate each course.

Certainly, there are deficits in using self-reports, as, because of their potential biases, they do not allow any inferences related to hard criteria the way standardized, objective tests do (OECD, 1999). However, past research has shown some substantial value for the use of student ratings. Particularly in the case of the assessment of key competences, which are often neglected and cannot be measured by standardized writing tests such as the ones developed to compare school children’s skills of main subjects (OECD, 1999), Klieme, Artelt, and Stanat (2002) recommend self-reports.

Self-report measures of teaching behavior are often distorted by the students’ liking for the lecturer, the grades they earned (Greenwald, 1997; Rindermann, 2001; for a strong counter position: Marsh & Roche, 2000), and so on. This can lead to inaccurate course evaluations, and comparisons between courses will be in favor of courses that are more appreciated by the majority of students (e.g., clinical psychology as compared to statistics). However, self-reported competence acquisition is less likely to be correlated with such variables. Even if we dislike the lecturer or our grades turn out to be worse than we desire, self-reported gains in competences are likely to be less distorted than self-reported course satisfaction, as self-esteem and self-appraisals. They obviously play a role when rating our own competences and counteract such a bias as well as blunting instrumentalization of the evaluation questionnaire to “take revenge” on the lecturer.

## Validity of Student Ratings

It has been discussed whether students are at all able to judge the quality of an academic course (Greenwald & Gillmore, 1998; L’Hommedieu, Menges, & Brinko, 1990; Marsh & Roche, 1997; Roche & Marsh, 1998). Some authors have argued that the validity of ratings is influenced by the amount of students’ work or by the strength of the lecturer (Greenwald & Gillmore, 1998). However, Marsh and Roche (1997; Roche & Marsh, 1998) draw a more optimistic picture. Marsh and Roche (1997) state that the validity of students’ ratings is hard to prove since there are no single criterion of effective learning. Furthermore, in their opinion, the utility of multiple-choice tests to validate course-evaluation instruments is limited; it would be more appropriate to show an instrument’s construct validity using several indicators of successful learning. Marsh and Roche (1997) explicitly mention the assessment of self-reports of competences.

If studies exploring the validity of self-ratings of personality characteristics are taken into account, empirical support for Marsh and Roche’s (1997) claims has been found. Gosling, John, Craik, and Robins (1998), as well as Spain, Eaton, and Funder (2000) showed that, although the accuracy of self-ratings depends on the person and context, self-ratings of personality characteristics or behavioral dispositions are correlated with alternative measurements of the same constructs. Lucas and Baird (2006) conclude that errors in self-report measures do not severely limit their validity. According to these findings we will regard self-ratings as valid indicators of gains in competences.

## Teaching Behavior vs. Acquisition of Competences

The HEsaCom was developed on the hypothesis that the construct *self-reported acquisition of competences* could be sufficiently distinguished from that of *appraisals of teaching behavior*, which currently available course-evaluation questionnaires are based on. If HEsaCom is, indeed, a new instrument that measures the self-reported acquisition of competences, then it will have to (1) be sufficiently delineated from other questionnaires and (2) selectively measure self-reported gains in competences as distinct from teaching behavior.

## Teaching Behavior

In our study, the “Fragebogen zur Evaluation der Lehre” (questionnaire for course evaluation) from Westermann et al. (1998) served as a representative of the currently available instruments mentioned above. Westermann and colleagues explicitly measure the students’ satisfaction with the instructor (Westermann et al., 1998, p. 135). The au-

thors distinguish satisfaction with the instructor from both the success and the quality of academic teaching (Westermann, Heise, Spies, & Trautwein, 1996). Satisfaction with studies is defined in analogy to Locke's (1976) concept of job satisfaction. Thus, general satisfaction with studies is conceptualized as the individual's attitudes toward a specific academic course. In line with this definition, Westermann et al. (1996) developed a scale that measures the satisfaction with an academic course and included this scale in their subsequent questionnaire (Westermann et al., 1998). All additional questions had been incorporated into the questionnaire only when the correlation with the previously developed satisfaction scale was high enough ( $r > = .30$ ; Westermann et al., 1998, p. 146). The resulting questionnaire of satisfaction is strongly oriented toward the persona of the instructor; many items contain appraisals of the instructor with respect to his or her behavior.

## Competences

HEsaCom claims to assess the personal acquisition of competences by the students. Competence can be understood as a complex arrangement of a human being's skills that are called into play in a variety of situations. In fact, competence "carries the dual meaning that there is a track record of such achievement (competent performance) and also that the individual has the capability to perform well in the future. It refers to good adaptation and not necessarily to superb achievement" (Coatsworth & Masten, 1998, p. 206). Weinert's (2001) definition is similar, and encompasses more than expert knowledge about a given field; systematic (e.g., problem-solving skills), personal (e.g., motivation), and social (e.g., cooperation skills) aspects are of importance, too. Thus,

HEsaCom maps the students' subjective reports of their acquisition of six domains of competences (see below). A detailed description of these competences can be found in Braun, Soellner, and Hannover (2006).

At this point it should be obvious that the two instruments – HEsaCom and Westermann et al.'s questionnaire – are assessing conceptually distinguishable constructs. However, it remains to be seen whether they are distinguishable empirically. This was tested by fitting structural equation models for each self-reported gain in six competence domains, and modeling the interrelationship of the given domain and satisfaction.

## Materials and Methods

### Participants

We evaluated 54 academic courses in 17 disciplines at nine German institutions. Paper and pencil questionnaires were mailed to the teachers, distributed by them to all students in a class at the end of the semester but before the final exams, and returned to us. We asked the lecturers to administer and collect the questionnaires during the lecture. Table 1 shows all institutions and participating disciplines, and the number of questionnaires sent out, based on the teachers' estimates of how many students attend their courses. In total, 1403 out of 2981 (47.06%) questionnaire copies provided to the lecturers were returned. As we could not obtain an exact number of the target population but only an estimate based on the teachers' estimates, it is unclear to what extent the 52.94% of questionnaires that were not returned reflect inaccurate teacher estimates or students who refused to fill out or return their questionnaires.

Table 1. Sample of universities

University	Discipline	N questionnaires
Bergische Universität Wuppertal	German language and literature studies	21
Fachhochschule Wiesbaden	Business studies	55
Freie Universität Berlin	Informatics	89
	Mathematics	11
	Philosophy and humanities	88
	Physics	47
	Political science	42
	Social science	220
	Economics	207
Humboldt Universität zu Berlin	Philosophy and humanities	11
Technische Universität Darmstadt	Informatics	14
Technische Universität Freiberg	Business studies	18
Universität Hamburg	Educational science	116
	Social science	48
	Economics	259
Universität Lüneburg	Economics	107
Universität Rostock	German language and literature studies	50
<b>Total</b>		<b>1403</b>

772 (55%) of the participating students whose questionnaires were returned were female, 515 (37%) male, and 116 (8%) students did not choose any of the gender-categories. As missing values were estimated by *multiple imputation techniques*, no cases had to be deleted, thus, all 1403 cases are included in subsequent analyses.

*Scale-dependent subsamples.* The scales for knowledge processing, systematic, communication, and personal competence were analyzed using all respondents ( $N = 1403$ ), whereas the scales for presentational and cooperation competence were analyzed using specific subsamples that solely consist of those students who answered the screening question positively (oral presentations:  $N = 524$ ; worked in groups:  $N = 676$ ). Thus, we tested the distinction between the self-rated acquisition of each domain of competence on the one hand and academic course satisfaction on the other hand, resulting in six model comparisons (see models in Appendix 1).

## Procedure

As a test of the hypothesized distinction of the constructs self-reported acquisition of competences and appraisals of academic course satisfaction, confirmatory factor analyses (CFA) have been conducted via Mplus 3.13. The course level was taken into account by specifying the course number as a cluster variable. Based on the comparison of different estimation techniques and the critical analysis of the normality assumption by Fouladi (2000), maximum-likelihood with robust standard error (MLR) was selected to estimate the model parameters in the CFAs.

In order to test whether the two different theoretical conceptions (appraisals of satisfaction vs. self-reported gains in competences) can also be empirically distinguished, two models were specified for each of the six competence domains, resulting in 12 models altogether. In both models of a given comparison, the construct of the self-reported acquisition in a particular competence domain (represented by its corresponding scale in HESaCom) was correlated with the construct academic course satisfaction (Westermann et al., 1996). The factor loading of each item was a free model parameter, as well as the corresponding variances and error terms, with one exception: A single loading for each factor was fixed to 1.0 (Jöreskog, 1993). The factor variances of the latent constructs were free parameters and the constructs were allowed to covary, as expected theoretically. All models in our study were congeneric measurement models because neither factor loadings nor error variances were equalized. According to the null hypothesis that self-reported acquisition of competences and academic course satisfaction are not empirically distinct, the correlation between both constructs was fixed to 1.0 in the null model. In the alternative model, the correlation was freely estimated, according to the alternative hypothesis that self-reported acquisition of competences and academic course satisfac-

tion are empirically distinguishable (for another example of this method see Federico, Golec, & Dial, 2005; for a methodological overview see van der Sluis, Dolan, & Stoel, 2005). Both models were compared by a likelihood-ratio ( $LR$ ) test differing at one degree of freedom in order to determine which, if any, model explained the data significantly better. Since the  $\chi^2$  statistics are very sensitive to sample size (Marsh, Balla, & McDonald, 1988), the fit of the more viable models was not only evaluated on the basis of the  $\chi^2$  significance test, but also on the basis of the ratio of  $\chi^2$ , the degrees of freedom ( $df$ ), and so-called close-fit indices: CFI, RMSEA, and SRMR. In addition to the examination of these global fit indices, we also examined the adequacy of the parameter estimates by means of factor loadings and error covariances, as recommended by Jöreskog (1993).

## Data Imputation

An extensive body of literature reports that missing values are problematic and it is far better to use modern statistical techniques to estimate those missing rather than using traditional listwise or pairwise deletion techniques (Schafer & Graham, 2002). Therefore, the data were preprocessed via the software package NORM 2.03 (Schafer, 1997) in order to estimate values for the 9% of the data that had missing values by means of multiple imputation (MI). To this end, we conducted an expectation-maximization (EM)-estimation, thereby obtaining proper starting values for the subsequent MI run. In line with Schafer and Graham (2002) we imputed 10 data sets, using the tenth set for the subsequent analyses. For each imputation cycle the remaining questionnaire items were introduced as *auxiliary variables* (year of birth, gender, course number, and course type). The use of these variables renders the assumption of randomly missing values (missing at random: MAR) more likely and lessens potential biases (Collins, Schafer, & Kam, 2001; Enders & Peugh, 2004; Enders, 2006). The goodness-of-fit of the imputation was confirmed by autocorrelations that were, for a lag of 50 in the time-series plot made available by NORM, close to zero.

## Evaluation in Higher Education: Self-Assessment of Competences (HESaCom)

The questions in HESaCom and its competence domains were derived from theoretical definitions. The English questions are presented in the Appendix. HESaCom (Braun et al., 2008) contains 26 items, each belonging to one of the following six competence scales:

- 1) knowledge processing (6 items, Cronbach's  $\alpha = .90$ )
- 2) systematic competence (3 items, Cronbach's  $\alpha = .83$ )
- 3) presentational competence (2 items, Cronbach's  $\alpha = .84$ )

- 4) communication competence (5 items, Cronbach's  $\alpha = .92$ )
- 5) cooperation competence (5 items, Cronbach's  $\alpha = .87$ )
- 6) personal competence (5 items, Cronbach's  $\alpha = .85$ .)

*Knowledge processing* refers to students' skills in applying their knowledge, and their skills in analyzing their applications of what they have learned (Dochy & Alexander, 1995; Bloom, 1956). *Systematic competence* constitutes an individual's skill in effectively planning his or her work as well as the mastery of relevant techniques. *Presentational competence* refers to students learning to speak in front of others, and consequently improve their oral presentation skills. Social competence is defined as the achievement of a balance between the development of positive outcomes for oneself and adherence to context-specific expectations of others (Wentzel, 1999). Some authors (Erpenbeck, 2003) differentiate social competence in *communication* and *cooperation competence*. A cooperative-competent person is capable of working with others effectively. Communication competence is describing the ability to verbally assert one's own opinion and to productively discuss it with others. In comparison, presentation competence is a technique; standing in front of a group and presenting facts in a comprehensible way. *Personal competence* refers to an individual's positive attitude toward learning and development of the self. It shows similarity to the construct "interest," and is meant to be adaptable. Students will develop a reflective and optimistic learning attitude in consequence of attending a course. Therefore, it is not a stable characteristic that already exists beforehand, but the outcome of high quality education (Braun et al., 2006).

It is known that not every course is aimed at the acquisition of all competences and HESaCom takes this into account by asking the students whether they gave oral presentations, for example. If so, and only then, are the students asked the questions regarding presentation competence. If they did not give oral presentations, it does not make sense to evaluate the acquisition of presentation competence. This is also true in the case of cooperation competence.

### Satisfaction Scale

The satisfaction scale included in the "Fragebogen zur Evaluation der Lehre" (course evaluation questionnaire; Westermann et al., 1998) was used as a measure for academic course satisfaction (four items, such as "Overall, I am satisfied with this course"; Cronbach's  $\alpha = .80$ ).

Participants' responses in both instruments were given on a five-point Likert scale.

## Results

### Models

In all model comparisons, for  $\chi^2$  difference values at 1 *df*, the *LR* tests were significantly ( $\chi^2_{\text{diff1}} = 588.117$ ;  $\chi^2_{\text{diff2}} = 779.460$ ;  $\chi^2_{\text{diff3}} = 305.779$ ;  $\chi^2_{\text{diff4}} = 1345.092$ ;  $\chi^2_{\text{diff5}} = 429.440$ ;  $\chi^2_{\text{diff6}} = 274.433$ ;  $p < .001$ ) in favor of the alternative models whose  $\chi^2$  values were lower than the ones of the null models. Thus, the null hypothesis, stating a lack of distinction between any competence domain and satisfaction, should be rejected. We can safely assume that all six competence domains are empirically distinct from satisfaction with teaching behavior. The factor loadings and fit indices for the alternative models are shown in the Appendix, as well as graphical representations of the models.

The amount of covariance between any given competence domain and academic course satisfaction was of secondary interest. In the alternative models, the correlations were estimated to be  $r_1 = .601$  for knowledge processing,  $r_2 = .310$  for systematic competence,  $r_3 = .255$  for presentational competence,  $r_4 = .279$  for communication competence,  $r_5 = .277$  for cooperation competence, and  $r_6 = .746$  for personal competence. These correlation coefficients indicate strong associations between knowledge processing and personal competence on the one hand and course satisfaction on the other hand. The other correlations are weak to moderate.

The common variances shared by the competence constructs and satisfaction amount to 36.12%, 9.61%, 6.50%, 7.78%, 7.67%, and 55.65% for knowledge processing, systematic, presentational, communication, cooperation, and personal competence, respectively. We conclude that self-reported acquisition of competences and satisfaction with teaching behavior are related, but empirically distinct.

### Conclusion

As theoretically expected, self-reported acquisition of competences and satisfaction with teaching behavior covary, but their covariation does not threaten the autonomy of either construct. For each model, the null hypothesis of a perfect construct can be rejected – even if the empirical correlation is fairly high between knowledge processing/ personal competence and rating of teaching behavior.

## Discussion

The Bologna Process requires academic course evaluations to aim at competences and their acquisition by students apart from the students' ratings of teaching behavior. Thus, the question of whether the assessment of the self-rated

acquisition of competences can successfully be delineated from ratings of satisfaction with teaching behavior can also be seen as a test of HESaCom's discriminatory validity. Overall, the assumed structure of the six competence domains' association with academic course satisfaction is confirmed by the global and local fit indices of the CFAs (Beauducel & Wittmann, 2005; Marsh, Hau & Wen, 2004): Competence acquisition and satisfaction with teaching behavior are related but distinct constructs.

Nevertheless, knowledge processing and personal competence were highly correlated with teaching-satisfaction. It can be argued that personal competence is conceptually close to satisfaction with the teacher. It seems to be plausible that students whose interest has been increased will be more satisfied with an academic course. While these findings correspond with other research, attention has to be paid to the question of causality. According to Cashin (1995), "instructors are more likely to receive higher ratings in classes where students had a prior interest in the subject matter" (p. 6). Therefore, further research on this issue is warranted. The concept of knowledge processing is also closely related to academic self-concept (e.g., Self-Description Questionnaire, SDQII; Marsh, 1994). In addition, it is well known that self-concept and interest are highly correlated (Hannover, 1998). Personal competences and knowledge processing are correlated ( $r = .76$ ; see Braun et al., 2008). Consequently, knowledge processing and satisfaction with teaching behavior covary as well.

Additionally to the results already shown, we conducted a mediation analysis, checking whether the relationship between personal competence and satisfaction is mediated by knowledge processing. According to Baron and Kenny (1986), we ran three regression analyses. Conducting a Sobel test (Sobel, 1982) indicated that knowledge processing is partially mediating the effect of personal competences on satisfaction with teaching behavior (Sobel = 10.24;  $p < .0001$ ); 27.34% of the total effects are being mediated, resulting in a decrease of the unmediated correlation, when the mediator is controlled for (results are available from the authors). This gives preliminary support to the assumption given above.

Apart from psychological explanations as to why some competences and teaching satisfaction can be highly correlated, it is important to emphasize that, as long as self-rated gains in competences and teaching satisfaction are theoretically and empirically distinguishable, high correlations are not particularly worrisome. Their distinction indicates that, despite even considerable overlaps, they entail something different and, even more important, have different implications. Certainly, as a criterion of quality of higher education, competence acquirements are preferable over teaching behavior. What we learn is what counts, not how much we like or enjoy it, even if both aspects of education are connected. Moreover, course instructors should focus more on the learning outcomes of their students than on their own teaching behavior and how this can satisfy the students' needs and desires. It is outcome-oriented academic course

evaluation in terms of gains in competences that can bring about this change in teaching orientation, not process-oriented evaluation targeting teaching behavior.

It has often been proposed that a standardized competence test be developed to measure gains in competences, administering such a test at the beginning and the end of the term. With respect to the fact that the Bologna Process explicitly demands basing the evaluation of programs of study on gains in competences, it might be difficult to develop an objective test measuring competence domains such as social or personal competence. However, even if such a test could be developed, gains in competences measured by it would not necessarily reflect the success of a specific course; rather, gains in competences measured in such a way might reflect diverse learning experiences both outside and inside university that affected students during the term. In this situation, we decided to assess gains in competences by the students' self-reports as an indicator of the course quality. Still, self-report measures are controversial and many authors have discussed whether self-reports can be considered as valid judgments of the educational success of a class (e.g., Greenwald & Gillmore, 1998; L'Hommedieu et al., 1990; Marsh & Roche, 1997; Roche & Marsh, 1998). Some authors have assumed that the validity of student self-reports might suffer from biases such as the perceived workload of a course or the strictness of its instructor (Greenwald & Gillmore, 1998). Marsh and Roche (1997; Roche & Marsh, 1998) criticize the typical approach of using multiple-choice-based performance tests as single validation criteria rather than construct-validating evaluation instruments. Additionally, Marsh and Roche (1997) report a synopsis of 41 studies typically finding correlations from .30 to .50 between performance scores and course evaluation scores. These findings support the validity of students' self-reports about the quality of academic courses.

Furthermore, studies focusing on the validity of self-reports of personality traits provide empirical support for the validity of self-reports (Gosling et al., 1998; Spain et al., 2000). The findings are even stronger in light of the variability of the accuracy of self-reports across individuals, contexts, and the assessed construct. In a review on the validity of self-reports, Lucas & Baird (2006) state: "Although errors surely do occur, they often do not severely limit the validity of the measures" (p. 41).

Existing evaluation instruments have often been criticized for their lack of theory. "Usually, a strictly empirical approach is applied in which the production of items is based on the designer's intuitive notions" (Schmidt, Dolmans, Gijsselaers, & Des Marchais, 1995, p. 83). Because of the commitment to several theoretical competence domains, HESaCom represents a theory-based inventory. Moreover, an adaptive strategy of "screening questioning" leads to a good approximation of a course's event. The development of HESaCom is regarded as a first step toward an outcome- and competence-oriented academic course evaluation. One of the main goals of implementing such an

orientation is to invite lecturers to think more about students' outcomes than about their own behavior. It offers the opportunity to implement a culture of outcome orientation. This parallels recent changes in European educational policies advocated by politics. Surely though, self-rated competences are subject to limitations and should not be used as an equivalent to "hard achievement measurements."

## Acknowledgments

We would like to thank Alan Woodley and anonymous reviewers for their very helpful comments on earlier drafts of this article.

## References

- Adam, S. (2004). *Using learning outcomes. A consideration of the nature, role, application, and implications for European education of employing "learning outcomes" at the local, national, and international levels*. Edinburgh: UK Bologna Seminar.
- Aleamoni, L.M., & Spencer, R.E. (1973). The Illinois Course Evaluation Questionnaire: A description of its development and a report of some of its results. *Educational and Psychological Measurement*, 33, 669–684.
- Baron, R.M., & Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173–1182.
- Beauducel, A., & Wittmann, W.W. (2005). Simulation study on fit indexes in CFA based on data with slightly distorted simple structure. *Structural Equation Modeling*, 12, 41–75.
- Bloom, B.S. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1*. New York: McKay.
- Bologna Working Group on Qualifications Frameworks. (2005). *A framework for qualifications of the European Higher Education Area*. Greena/Copenhagen: Grefta Tryk A/S. Retrieved on August 22, 2009, from [http://www.bologna-bergen2005.no/Docs/00-Main\\_doc/050218\\_QF\\_EHEA.pdf](http://www.bologna-bergen2005.no/Docs/00-Main_doc/050218_QF_EHEA.pdf)
- Braun, E., Gusy, B., Leidner, B., & Hannover, B. (2008). Das Berliner Evaluationsinstrument für selbstgeschätzte studentische Kompetenzen (BEvaKomp) [Competence-based course evaluation: Berlin Evaluation Instrument for students' competences]. *Diagnostica*, 54, 30–42.
- Braun, E., Soellner, R., & Hannover, B. (2006). Ergebnisorientierte Lehrveranstaltungsevaluation [Outcome-oriented course evaluation]. In Hochschulrektorenkonferenz (HRK) (Ed.), *Qualitätsentwicklung an Hochschulen – Erfahrungen und Lehren aus 10 Jahren Evaluation* [Quality development in higher education institutions – 10 years of experience] (pp. 60–67). Bonn: HRK.
- Brennan, J. (2001). *Lifelong learning for equity and social cohesion a new challenge to higher education*. Strasbourg: Final Conference Report.
- Burdsal, C.A., & Bardo, J.W. (1986). Measuring student's perceptions of teaching: Dimensions of evaluation. *Educational and Psychological Measurement*, 46, 63–9.
- Cashin, W.E. (1995). *Student ratings of teaching: The research revisited*. IDEA Paper No. 32. Manhattan: Kansas State University, Center for Faculty Evaluation and Development.
- Coatsworth, J.D., & Masten, A. (1998). The development of competence in favorable and unfavorable environments: Lessons from research on successful children. *American Psychologist*, 53, 205–220.
- Collins, L.M., Schafer, J.L., & Kam, C.K. (2001). A comparison of inclusive and restrictive strategies in modern missing-data procedures. *Psychological Methods*, 6, 330–351.
- Dochy, F.J.R.C., & Alexander, P.A. (1995). Mapping prior knowledge: A framework for discussion among researchers. *European Journal of Psychology of Education*, 10, 225–242.
- Erpenbeck, J. (2003). Einführung [Introduction]. In J. Erpenbeck & L. von Rosenstiel (Eds.), *Handbuch Kompetenzmessung. Erkennen, Verstehen und Bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis* [Handbook of competence assessment. Identification, comprehension, and evaluation of competences in job-related, pedagogical, and psychological practice] (pp. 9–19). Stuttgart: Schaeffer-Poeschel.
- Enders, C.K. (2006). Analyzing structural equation models with missing data. In G.R. Hancock & R.O. Mueller (Eds.), *Analyzing structural equation models with missing data* (pp. 313–342). Greenwich, CT: Information Age.
- Enders, C.K., & Peugh, J.L. (2004). Using an em covariance matrix to estimate structural equation models with missing data: Choosing an adjusted sample size to improve the accuracy of inferences. *Structural Equation Modeling*, 11, 1–19.
- Federico, C.M., Golec, A., & Dial, J.L. (2005). The relationship between the need for closure and support for military action against Iraq: Moderating effects of national attachment. *Personality and Social Psychology Bulletin*, 31, 621–632.
- Fouladi, R.T. (2000). Performance of modified test statistics in covariance and correlation structure analysis under conditions of multivariate nonnormality. *Structural Equation Modeling*, 7, 356–410.
- Gosling, S.D., John, O.P., Craik, K.H., & Robins, R.W. (1998). Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers. *Journal of Personality and Social Psychology*, 74, 1337–1349.
- Greenwald, A., & Gillmore, G. (1998). How useful are student ratings? *American Psychologist*, 53, 1228–1229.
- Greenwald, A.G. (1997). Validity concerns and usefulness of student ratings of instruction. *American Psychologist*, 52, 1182–1186.
- Hannover, B. (1998). The development of self-concept and interests. In L. Hoffmann, A. Krapp, K.A. Renninger, & J. Baumert (Eds.), *Interest and learning* (pp. 105–125). Kiel: IPN.
- Heckert, T.M., Latier, A., Ringwald-Burton, A., & Drazen, C. (2006). Relations among student effort, perceived class difficulty appropriateness, and student evaluations of teaching: Is it possible to "buy" better evaluations through lenient grading? *College Student Journal*, 40, 588–596.
- Jöreskog, K.G. (1993). Testing structural equation models. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 294–317). Newbury Park: Sage.
- Klieme, E., Artelt, C., & Stanat, P. (2002). Fächerübergreifende Kompetenzen [Interdisciplinary competences]. In F.E. Weinert

- (Ed.), *Leistungsmessungen in der Schule* [Measurements of achievement in schools] (pp. 203–218). Weinheim: Beltz.
- Kolitch, E., & Dean, A.V. (1999). Student ratings of instruction in the USA: Hidden assumptions and missing conceptions about “good” teaching. *Studies in Higher Education, 24*, 27–42.
- L’Hommedieu, R., Menges, R.J., & Brinko, K.T. (1990). Methodological explanations for the modest effect of feedback from student ratings. *Journal of Educational Psychology, 82*, 232–241.
- Locke, E.A. (1976). The nature and causes of job satisfaction. In M.D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 1297–1349). Chicago: Rand McNally.
- Lucas, R.E., & Baird, B.M. (2006). Global self-assessment. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 29–42). Washington, DC: American Psychological Association.
- Marsh, H.W. (1984). Students’ evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology, 76*, 707–754.
- Marsh, H.W. (1994). Using the national longitudinal study of 1988 to evaluate theoretical models of self-concept: The self-description questionnaire. *Journal of Educational Psychology, 86*, 439–456.
- Marsh, H.W., Balla, J.R., & McDonald, R.P. (1988). Goodness of fit in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103*, 391–410.
- Marsh, H.W., Hau, K.T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler’s (1999) findings. *Structural Equation Modeling, 11*, 320–341.
- Marsh, H.W., & Roche, L.A. (1997). Making students’ evaluations of teaching effectiveness effective. *American Psychologist, 52*, 1187–1197.
- Marsh, H.W. & Roche, L.A. (2000). Effects of grading leniency and low workload on students’ evaluations of teaching: Popular myth, bias, validity, or innocent bystanders? *Journal of Educational Psychology, 92*, 202–228.
- OECD. (1999). *Measuring student knowledge and skills. A new framework for assessment*. Paris: OECD Publishing.
- Rindermann, H. (2001). *Lehrevaluation. Einführung und Überblick zu Forschung und Praxis der Lehrevaluation an Hochschulen mit einem Beitrag zur Evaluation computerbasierter Unterrichts* [Course evaluation: Introduction and overview on research into and practice of course evaluation in higher education, including a contribution on the evaluation of computer-based teaching]. Landau: Verlag Empirische Pädagogik.
- Roche, L.A., & Marsh, H.W. (1998). Workload, grades, and students’ evaluations of teaching: Clear understanding sometimes requires more patient explanations. *American Psychologist, 53*, 1230–1231.
- Schafer, J.L. (1997). *Analysis of incomplete multivariate data*. London: Chapman and Hall/CRC Press.
- Schafer, J.L., & Graham, J. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147–177.
- Schmidt, H.G., Dolmans, D.H.J.M., Gijsselaers, W.H., & Des Marchais, J.E. (1995). Theory-guided design of a rating scale for course evaluation in problem-based curricula. *Teaching and Learning in Medicine, 7*, 82–91.
- Sobel, M.E. (1982). Asymptotic intervals for indirect effects in structural equations models. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 290–312). San Francisco: Jossey-Bass.
- Spain, J.S., Eaton, L.G., & Funder, D.C. (2000). Perspective on personality: The relative accuracy of self versus others for the prediction of emotion and behavior. *Journal of Personality, 68*, 837–867.
- Sluis, S. van der, Dolan, C., & Stoel, R.D. (2005). A note on testing perfect correlations in SEM. *Structural Equation Modeling, 12*, 551–577.
- Wende, M.C. van der (2000). The Bologna Declaration: Enhancing the transparency and competitiveness of European higher education. *Higher Education in Europe, 25*, 205–310.
- Wentzel, K.R. (1999). Social-motivational processes and interpersonal relationships: Implications for understanding students’ academic success. *Journal of Educational Psychology, 91*, 76–97.
- Weinert, F.E. (2001). Concept of competence: A conceptual clarification. In D. Rychen & L. Salganik (Eds.), *Defining and selecting key competences* (pp. 17–31). Göttingen: Hogrefe & Huber.
- Westermann, R., Heise, E., Spies, K., & Trautwein, U. (1996). Identifikation und Erfassung von Komponenten der Studienzufriedenheit [Identification and measurement of elements of students’ satisfaction]. *Psychologie in Unterricht und Erziehung, 43*, 1–22.
- Westermann, R., Spies, K., Heise, E., & Wollburg-Claar, S. (1998). Bewertung von Lehrveranstaltungen und Studienbedingungen durch Studierende: Theorieorientierte Entwicklung von Fragebögen [Students’ evaluation of courses and study conditions: Theory-oriented development of questionnaires]. *Empirische Pädagogik, 12*, 133–166.

---

#### About the authors

Edith Braun, Ph.D., is a senior researcher at the Department Schul- und Unterrichtsforschung of the Freie Universität Berlin, where she is responsible for the competence-based course evaluation. Her research interests are quality assurance, learning processes, and learning environments in higher education.

Bernhard Leidner is a Postdoctoral Fellow in the Department of Psychology at The New School for Social Research in New York City. His main research interests are intergroup relations and conflicts, (inter)group processes, moral disengagement, and social justice.

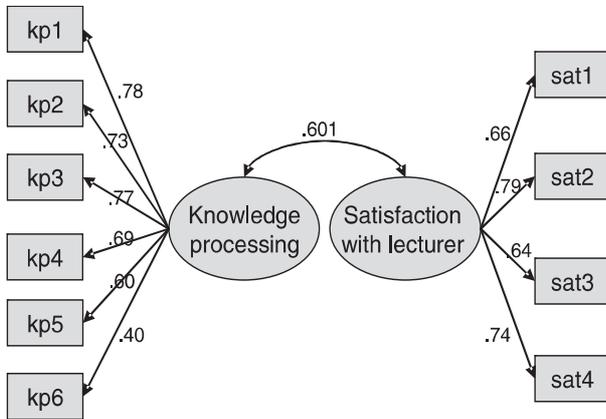
---

#### Edith Braun

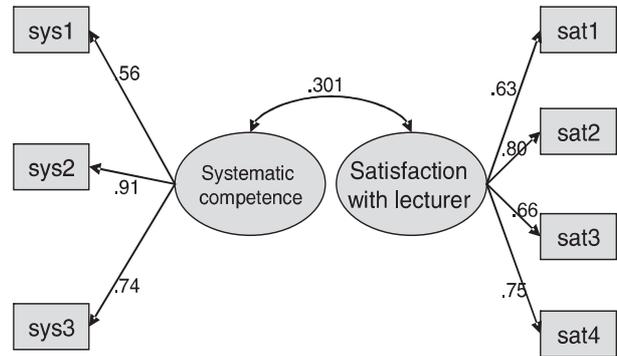
Schul- und Unterrichtsforschung  
 Freie Universität Berlin  
 Habelschwerdter Allee 45  
 D-14195 Berlin  
 Germany  
 Tel. +49 30 8-385-6952  
 E-mail edith.braun@fu-berlin.de

## Appendix 1

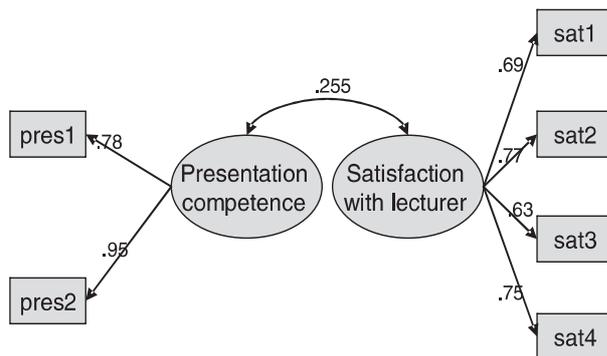
Alternative models: Factor loadings and model fit of six confirmatory factor analyses (competences-satisfaction correlation being freely estimated)



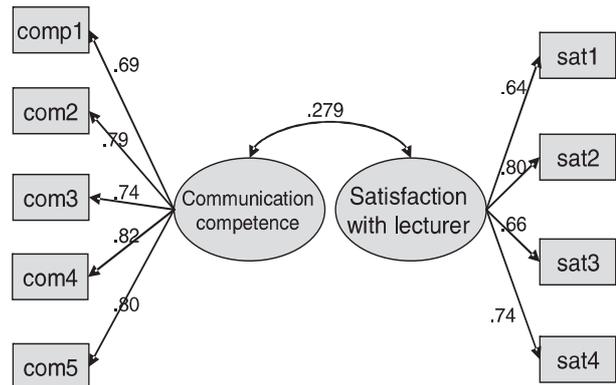
1) Knowledge Processing - Model fit:  $\chi^2=254.598$ ,  $df=34$ ,  $p$ -value<0.001,  $\chi^2/df=7.48$ , RMSEA=0.068, SRMR=0.044, CFI=0.961 (N=1403)



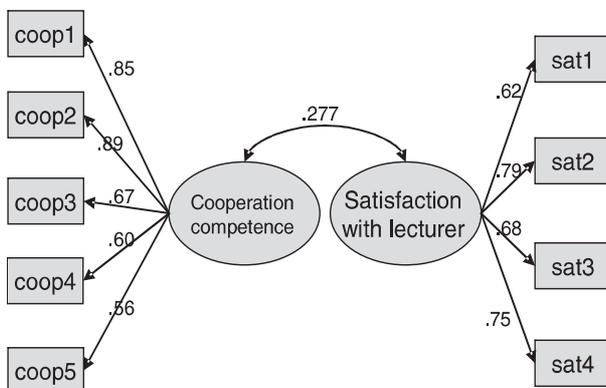
2) Systematic Competence - Model fit  $\chi^2=136.447$ ,  $df=13$ ,  $p$ -value<0.001,  $\chi^2/df=10.50$ , RMSEA=0.082, SRMR=0.047, CFI=0.953 (N=1403)



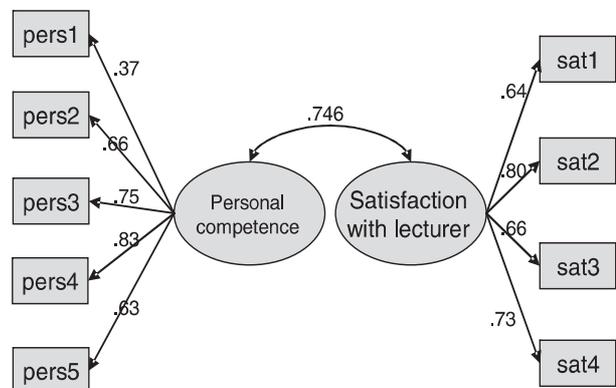
3) Presentational Competence - Model fit:  $\chi^2=28.741$ ,  $df=8$ ,  $p$ -value<0.001,  $\chi^2/df=3.60$ , RMSEA=0.070, SRMR=0.030, CFI=0.975 (N=524)



4) Communication Competence - Model fit :  $\chi^2=255.887$ ,  $df=26$ ,  $p$ -value<0.001,  $\chi^2/df=9.80$ , RMSEA=0.079, SRMR=0.039, CFI=0.941 (N=1403)



5) Cooperation Competence - Model fit:  $\chi^2=111.833$ ,  $df=26$ ,  $p$ -value<0.001,  $\chi^2/df=4.30$ , RMSEA=0.070, SRMR=0.051, CFI=0.938 (N=676)



6) Personal Competence - Model fit:  $\chi^2=203.934$ ,  $df=26$ ,  $p$ -value<0.001,  $\chi^2/df=7.80$ , RMSEA=0.070, SRMR=0.037, CFI=0.956 (N=1403)

## Appendix 2

### Items and Scales of the HEsaCom

#### Knowledge Processing

1. As a result of this course, I can remember most of the important terms and facts from this course.
2. As a result of this course, I can give an overview of the course.
3. The course has helped me improve my analysis of complex issues in this subject area.
4. This course has helped me improve my handling of typical problems in this subject area.
5. This course has helped me both to see the connections and to notice inconsistencies in this subject area.
6. This course has helped me judge the quality of academic articles in this subject area.

#### Systematic Competence

1. This course has helped me to acquire information more efficiently.
2. This course has helped me organize my work.
3. This course has helped me improve the way I work.

#### Presentation Competence

*FILTER: I gave an oral presentation as part of this course.*

1. After presenting in this course I feel I can engage better with the audience.
2. After presenting in this course I can structure my talks better.

#### Communication Competence

*FILTER: I regularly spoke in course.*

1. This course has helped me express my opinion.

2. This course has helped me to ask for clarification when I have difficulty understanding.
3. This course has helped me speak in a way that others can understand.
4. This course has helped me speak more precisely.
5. This course has helped me to improve the way I moderate discussions.

#### Cooperation Competence

*FILTER: I've worked with other students in a work group for more than 2 weeks during this course.*

1. My participation in the group work made it easier for me to help delegate tasks.
2. My participation in the work group made it easier for me to know when to hold back from contributing.
3. My participation in the work group made it easier for me to stand up for constructive team spirit.
4. My participation in the work group helped me take personal responsibility for my share of the work.
5. I feel identified with our work group's accomplishment.

#### Personal Competence

1. I have grown more interested in the subject matter as the course has progressed.
2. The course encouraged me to continue my studies.
3. The course has increased my joy of carrying out assigned tasks.
4. I feel more inspired by the topics studied in this course than at the beginning.
5. The course has inspired me to study the subject further in my own time.