Practical Assessment, Research, and Evaluation

Volume 19 *Volume 19, 2014* Article 4

2014

Influences of mood on academic course evaluations

Joerg Zumbach

Joachim Funke

Recommended Citation
Zumbach, Joerg and Funke, Joachim (2014) "Influences of mood on academic course evaluations," *Practical Assessment, Research, and Evaluation*: Vol. 19 , Article 4.
DOI: https://doi.org/10.7275/vbcz-h361
Available at: https://scholarworks.umass.edu/pare/vol19/iss1/4

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 19, Number 4, May 2014 ISSN 1531-7714

Influences of mood on academic course evaluations

Joerg Zumbach, *University of Salzburg, Austria*
Joachim Funke, *University of Heidelberg, Germany*

In two subsequent experiments, the influence of mood on academic course evaluation is examined. By means of facial feedback, either a positive or a negative mood was induced while students were completing a course evaluation questionnaire during lectures. Results from both studies reveal that a positive mood leads to better ratings of different dimensions of lecture quality. While in Study 1 (N=109) mood was not directly controlled, Study 2 (N=64) replicates the findings of the prior study and reveals direct influences of positive and negative mood on academic course evaluation.

Course assessment and evaluation in university programs has become an everyday phenomenon in academic life. This has led to the emergence of a number of instruments which cover a broad range of dimensions using different constructs in order to describe teaching quality. Such instruments, mostly rating-scale questionnaires, not only differ in terms of the underlying dimensions of university teaching being assessed but also in the assessment methods used. Several approaches to avoiding systematic disturbance variables have been developed in this respect. These include specific assessment approaches (e.g., mid-term course evaluation in order to also capture students who might drop out before the end of the term or to compare competence ratings with course appraisals; Braun & Leidner, 2009). Systematic errors in course evaluation can also be expected to result from other sources of variance (e.g., overall popularity of a specific course or a specific lecturer, c.f. Greenwald, 1997; Greenwald & Gillmore, 1998; or actual involvement in learning tasks, cf. Reinhard & Sporer, 2010). We start with a short review of the literature on biases in academic course evaluation before analyzing specific effects of mood on classroom ratings.

Biases in course evaluations

One systematic review of biases in course evaluations is provided by Spinath and Stehle (2011), who mention potential biases as a result of (a) the number of participants in a course, (b) the personality of the teacher, (c) student expectations, and (d) course requirements.

Concerning the number of participants in a course, Feldman (1984) found a "very weak" inverse correlation between the number of students enrolled on a college course and students' overall evaluation of the course and its teacher. Based on data from 1,157 classes, Mateo and Fernández (1996) confirmed this result and the weak effect size.

With respect to teachers' personality, McPherson, Kearney, and Plax (2003) analyzed students' reactions to teachers' normative and non-normative expressions of anger in the classroom. They found that normative expressions of anger are associated with negative evaluations of the teacher and the course. In addition, effects in terms of social desirability have not been found. Spooren, Mortelmans, and Thijssen (2012) report negative evidence (based on data from 1,835 students from two different studies) for the hypothesis of a response style (in particular, acquiescence). However, as attendance was not compulsory in these samples, the results might not be representative.

With respect to students' expectations, mature work from Bock (1979) might be of interest here: data from four semesters of students in Abnormal Psychology (N=160), taught by the same instructor, indicated that those whose expectations coincided with the course content were more likely to rate the instructor as 'highly effective' or 'effective' (94%), while students who had expected a different type of course were more likely to

rate the instructor as 'little effective' or 'ineffective' (83%).

With regard to course requirements, Greenwald and Gillmore (1997) analyzed the effects of grading leniency on ratings. They concluded that a statistical correction for removing the unwanted inflation of ratings produced by lenient grading might overcome the detected bias (see also Gillmore & Greenwald, 1999; Marsh & Roche, 1999). Similar results, based on data from 220 enrolled students, were reported by Svanum and Aigner (2011), who managed to demonstrate that students assessed the same course and instructor differently depending on such factors as the degree of their own success, their motivations for taking the course, and the amount of effort invested. Feldman (1998) showed that course content also has an impact on course evaluation. College instructors teaching classes whose content was deemed more complex received better scores than professors teaching classes with less sophisticated content.

He also showed that optional courses receive better ratings than mandatory courses and that "hard" courses (e.g., science courses) are given lower ratings than "soft" courses (e.g., pedagogy courses). Another major impact on academic course evaluation is the participants' subjective interest in a class (cf. Daniel, 1994; Döring, 2005; Kromrey, 1995). According to Komrey (1996), the popularity of the content of a course and the individual interest and motivation of participants might explain up to 80% of variance between different course ratings. Although the majority of instruments used here appear to be reliable (e.g., Rindermann & Schofield, 2001), it is generally difficult to judge the validity of students' course ratings due to methodological issues such as small sample sizes and missing validity criteria (e.g., Abrami, d'Apollonia & Rosenfield, 2007). While most of the research mentioned above predominantly addresses the influence of different courses and teachers, our research focuses on variance derived from intra- and inter-individual sources, with one well-known factor impacting people's judgments is their mood.

Effects of mood on judgments

The influence of mood on people's judgments has been examined in a number of experiments (e.g., Griskevicius, Shiota & Neufeld, 2010; Huntsinger, Clore & Bar-Anan, 2010; Schwarz, Bless & Bohner, 1991). While there is a large body of evidence regarding the influence of mood on judgments and underlying processing strategies in experimental laboratory research (e.g., Elaboration Likelihood Model; cf. Petty & Cacioppo, 1986; Eagly & Chaiken, 1993), applied research studies pose the question of whether and how different moods affect academic course evaluation. There is a body of research investigating the role of mood in human information processing. For instance, mood may activate mood-congruent information in memory. Thus, it might provide a basis for evaluation processes as suggested in the Affect Infusion Model (AIM; Forgas, 1995; Forgas, Goldenberg & Unkelbach, 2009). Mood itself might serve as information if no or little information is available for judgment processes (cf. Schwarz & Clore, 2007).

We assume that a positive mood leads to a more shallow information processing, and thus, to a more positive course evaluation. Evidence for this hypothesis is provided by Fiedler (2001), who suggests a differentiation between assimilative and accommodative processing, both of which are affected by a positive or negative mood in different ways. *Assimilative information processing* refers to processes of information transformation and active elaboration of stimuli. *Accommodative information processing* mainly aims at an exact preservation of information with an emphasis on avoidance of errors rather than developing new relations or interpretations. According to Fiedler (2001), these two styles of information processing correspond to learning approaches in appetitive or aversive situations. Appetitive situations are better mastered by means of explorative and creative behavior, while in aversive situations a more careful approach (searching for information and avoiding errors) is more appropriate. As these situations are typically accompanied by positive or negative affect, a similar type of information processing as in the original situation can be triggered: positive affect activates assimilative processing while negative mood triggers accommodative processing (for different types of information processing see Gawronski & Bodenhausen, 2005; Inbar, Cone & Ginovich, 2010).

Hypothesis

We conclude that positive mood supports creative, generative information processing. However, a more conservative and careful strategy, as affected by negative mood, can also be advantageous. This is contradictory to the common understanding that negative emotions always have an inhibitive effect. With regard to course evaluation, it can be expected that negative mood might cause students to focus more on criticizing details of a course while positive mood might lead to a global and mostly positive overall impression.

In order to test this assumption, we conducted two consecutive experiments.

Experiment 1

Method

Participants

109 participants (86 female, 23 male, mean age=20.84; *SD*=3.42) were recruited for this study. 81 participants were undergraduate university students of Psychology at the University of Heidelberg (Germany) attending a lecture in General Psychology. 28 participants were undergraduate university students of Psychology at the University of Salzburg (Austria) attending an introductory lecture in Instructional Psychology. Participants were randomly assigned to one of two conditions (positive- vs. negative-mood condition). No reward was given for participation as the course evaluation was part of the academic program. Evaluations were conducted once, three weeks before the end of the term.

Material

Mood induction. In order to operationalize different mood conditions (positive vs. negative mood), we used the assumptions of the Facial Feedback Hypothesis (FFH; Strack, Martin & Stepper, 1988). The FFH suggests that different states of mood caused by specific muscle contraction can influence information processing (cf. Bodenhausen, Kramer & Süsser, 1994; Kleinke, Peterson & Rutledge, 1998; Strack & Neumann, 2000). Consequently, a change in contraction of certain facial muscles is sufficient in order to produce a more positive or negative mood. A contraction of the lips activates the *orbicularis oris* and stimulates disapproval while clenching the teeth stretches the *zygomaticus major* and produces a smile. According to recent embodied-cognition approaches, the associated mimic is not only an expression of a certain mood but simultaneously serves as information (Havas, Glenberg, Gutowski, Lucarelli & Davidson, 2010). Based on specific muscle contraction and relaxation, this information can be used again in order to adopt a certain mood. For instance, a "smiling" expression caused by a pencil kept between the teeth may provide the information that the keeper is in a positive mood. This will be adopted and will indeed lead to a positive mood. On the other hand, a pencil kept between the lips leads to a facial expression that indicates a negative mood. Correspondingly, the induction of a negative mood becomes more likely. Evidence for the success of using pencils that have to be kept between the lips or the teeth and the induction of the expected mood is provided by Strack et al. (1988). For our experiment, we also asked participants, depending on the study arm they had been assigned to, to keep a pencil or a pen either between their lips or their teeth.

Course evaluation questionnaire. Parallel to the mood induction, participants had to fill in a course evaluation questionnaire. We used a special questionnaire designed for course evaluation in lectures. The instrument itself was developed at the Department of Psychology at the University of Heidelberg. It is a further development of a scale provided by Zumbach, Spinath and Schahn (2007). Overall, the following dimensions are assessed here by means of 39 items: (1) Knowledge Acquisition (e.g., "Participation in this lecture provided me with basic expertise in this area."; *Cronbach's Alpha* in this sample = 0.74), (2) Teacher (e.g., "The teacher has structured the content of the lecture well."; *Cronbach's Alpha* = 0.75), (3) Framework Conditions (e.g., "The technical equipment is adequate for this lecture."; *Cronbach's Alpha* = 0.71), (4) Workload (e.g., "The speed of this lecture is too fast."; *Cronbach's Alpha* = 0.75), (5) Content (e.g., "The content presented relates to my prior knowledge."; *Cronbach's Alpha* = 0.47), (6) Participants (e.g., "Participants attended the lecture regularly"; *Cronbach's Alpha* = 0.48), and (7) Overall Evaluation (e.g., "I would grade the course as followed:"; *Cronbach's Alpha* = 0.60; one item was excluded here in order to increase internal consistency). All scales with the exception of "Overall Evaluation" are based on a five-point Likert scale from 1 (failed) to 5 (very good). The scale for "Overall Evaluation" is reversed from 1 (very good) to 5 (failed) following the grading system in Austria and Germany.

Design and Procedure. This study uses a one-factorial design with the independent variable "Mood" (positive vs. negative) and the mean values of the sub-scales of the course evaluation questionnaire as well the overall grade as dependent variables. The course evaluation was announced one week beforehand and took place within the scheduled lecture time. Participants were told that they would be taking part in an experiment which investigates people's ability to perform different tasks with parts of their body not normally used for those tasks, and the influence of this on a main task (here: course evaluation). They were also informed that the procedure of keeping the pencil in their mouth was an essential part of the research. They were not informed beforehand (but after participation) that this analysis focused on the influences of mood. Before they received

the questionnaires, they were asked to volunteer in having a pencil or pen either between their lips or their teeth. The exact position of the pen (either between lips or teeth) was demonstrated by the instructor who also ensured that all participants kept their pencil in the correct position according to the study arm they had been assigned while performing the course evaluation. Only participants who agreed to the study requirements were included in the analysis; a total of 18 students did not want to participate and were excluded from the analysis. The students were then randomly assigned to one of the two study arms and requested to put the pen or pencil into their mouth according to their study arm. They then received and filled in the course evaluation questionnaire inside the lecture theatre. The students were able to observe their colleagues but were mainly focused on their own questionnaires. Participants were given twenty minutes to complete the evaluation, which was sufficient for all participants to answer all questions. Afterwards the anonymous questionnaires were collected and participants (all naïve to FFH) were debriefed.

Results

All statistical analyses were performed using the PASW 18 statistical software package. Assumptions for parametric inferential tests (homogeneity of variances, sphericity, and normal distribution) were given for almost all sub-scales in Experiment 1 and Experiment 2 (four sub-scales within each sub-sample differed marginally in significance from normal distribution). Due to the robustness of Analysis of Variance we decided to employ parametric inferential testing using a MANOVA (cf. Buehner, 2010). This MANOVA revealed a significant overall effect of "Mood" on the evaluation scales ($F(7, 99)=2.31$, $p=0.032$; $\eta^2=0.14$). Analysis of between-subject effects within the MANOVA showed significant differences regarding the "Overall Evaluation" (F (1, 105)=7.00, $p=0.009$; $\eta^2=0.06$) and "Participants" (F (1, 105)=7.06, $p=0.009$; $\eta^2=0.06$) sub-scales. There were no significant group differences with regard to "Knowledge Acquisition" (F (1, 105)=1.57, $p=0.21$; $\eta^2=0.06$),"Teacher" (F (1, 105)=2.65, $p=0.11$; $\eta^2=0.03$), "Framework Conditions" (F (1, 105)=0.02, $p=0.90$; $\eta^2<0.00$), "Workload" (F (1, 105)=0.01, $p=0.92$; $\eta^2<0.00$) , and "Content" (F (1, 105)=0.54, $p=0.47$; $\eta^2=0.01$).

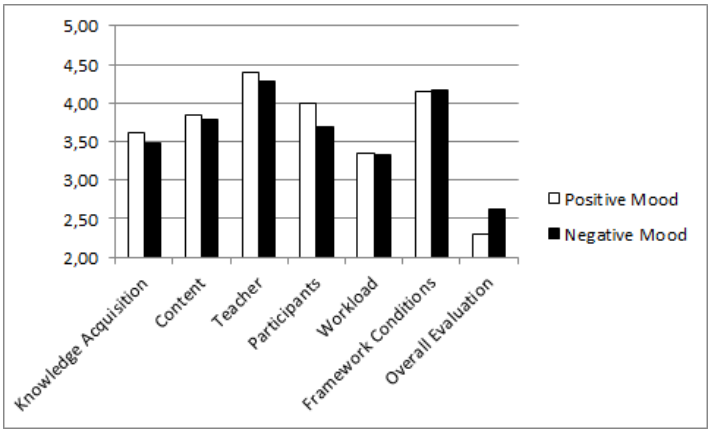

Figure 1: Descriptive results from Experiment 1 (higher values indicate a more positive evaluation except "Overall Evaluation").

Descriptive results reveal that across all measurements, the positive-mood condition leads to better course evaluation results than the negative-mood condition (see Table 1 and Figure 1).

Discussion of Experiment 1

The results from this first study reveal that participants' mood might be a source of systematic bias

Table 1: Mean values (standard deviations in brackets) and inferential statistics for the seven sub-scales of the evaluation questionnaire from Experiment 1 (N=109).

Dependent Variable	Positive Mood	Negative Mood	Type III Square Sum	F	p	η^2
Knowledge Acquisition	3.61 (0.50)	3.48 (0.56)	0.45	1.57	0.21	0.02
Teacher	4.39 (0.40)	4.28 (0.35)	0.37	2.65	0.11	0.03
Framework Conditions	4.16 (0.62)	4.17 (0.67)	0.07	0.02	0.90	0.00
Workload	3.34 (0.50)	3.33 (0.57)	0.03	0.01	0.92	0.00
Participants	3.99 (0.56)	3.70 (0.57)	2.26	7.07	0.01	0.06
Content	3.85 (0.34)	3.78 (0.49)	0.10	0.54	0.47	0.01
Overall Evaluation	2.30 (0.63)	2.63 (0.65)	2.88	7.00	0.01	0.06

in academic course evaluation. It seems that, due to the Facial Feedback Hypothesis, we were able to induce a positive (pen or pencil between the teeth) or a negative (pen or pencil between the lips) mood. The induced mood appears to have influenced information processing and, thus, judgment in this course evaluation. Although the positive-mood condition produced more positive results on all sub-scales of the course evaluation instrument used here, significant group differences could be found with regard to two sub-scales ("Participants" and "Overall Evaluation"). An explanation might be that both sub-scales are in the final part or at the end of the questionnaire, and that the intervention may have taken some time to be effective. The limitations of this study are as follows: Firstly, the pooling of the two samples is problematic, because it might lead to the Yule's or Simpson's Paradox (cf. Rosenthal, 1991). According to this paradox, the true mean values of two pooled samples cannot be computed and pooling might even lead to an inversion of values of dependent variables within sub-samples. Secondly, the internal consistency of the values obtained for some sub-scales in this first study is rather poor, which might also be a consequence of sample pooling. Thirdly, we did not control mood itself. Thus, we can only assume that the intervention led to different mood conditions, which is supported by the hypothesized findings. However, at this stage, we cannot prove that it was indeed mediated by mood. Therefore, these limitations made a second replication and extension study necessary.

Experiment 2

Method

Participants

Sixty four participants (41 female, 23 male, mean age=22.16; *SD*=3.51) were recruited for this study. All were undergraduate university students of Psychology at the University of Salzburg (Austria) attending an introductory lecture in Educational Psychology. Participants were randomly assigned to one of two conditions (positive- vs. negative-mood condition). No reward was given for participation as the course evaluation was part of the academic program.

Material

Mood induction. The mood induction procedure was the same as in Experiment 1. Participants were randomly assigned to one of two conditions (we asked participants, in accordance with the study arm they had been assigned to, to keep a pencil or a pen either between their lips or their teeth).

Course evaluation questionnaire. Again, participants had to fill in a course evaluation questionnaire. The questionnaire used in this second experiment was a further development of the same questionnaire used in the first study (see Appendix 1; Stehle, 2011). Due to problems with the internal consistency of some scales in the previous version, this version now consisted of 41 items assessing 9 dimensions. The dimensions were (1) Overall Impression (e.g., "The lecture is well structured."; *Cronbach's Alpha* in this sample = 0.76), (2) Knowledge Acquisition (*Cronbach's Alpha* = 0.82), (3) Content (e.g., "Participation in this lecture provided me with basic expertise in this area."; *Cronbach's Alpha* = 0.69 with one item dropped), (4) Teacher (*Cronbach's Alpha* = 0.84), (5) Student-Teacher Interaction (*Cronbach's Alpha* = 0.83), (6) Framework Conditions (*Cronbach's Alpha* = 0.68), (7) Participants (*Cronbach's Alpha* = 0.75), (8) Workload (*Cronbach's Alpha* = 0.68), (9) Overall Grading (participants were required to give the lecture an overall grade corresponding to the German/Austrian school grading system from 1 – very good to 5 – failed). In addition, and in order to validate the results, we assessed participants' mood by means of a manipulation test on a 5-point Likert-scale from failed (1) to very good (5). This was, in addition to the improvement of the course evaluation scales and the recruitment of participants, the most important difference to Experiment 1.

Design and Procedure. This study also uses a one-factorial design with the independent variable "Mood" (positive vs. negative) and the mean values of the sub-scales of the course evaluation questionnaire as well the overall grade as dependent variables. The procedure was the same as in Experiment 1.

Results

A MANOVA revealed a significant overall effect of "Mood" (F (11, 47)=2.24, p=0.028; η^2=0.34). Analysis of between-subject effects showed significant differences regarding the "Teacher" (F (1, 57)=5.23, p=0.026; η^2=0.08), "Content" (F (1, 57)=13.99, p<0.001; η^2=0.20), "Knowledge Acquisition" (F (1, 57)=4.94, p=0.03; η^2=0.08), and the "Overall Grading" (F (1, 57)=6.89, p=0.01; η^2=0.11) sub-scales. Also regarding the control item "Mood", there was a significant group difference (F (1, 57)=5.56, p=0.021; η^2=0.09). There were no significant group differences with regard to "Workload" (F (1, 57)=0.37, p=0.54; η^2<0.01),"Participants" (F (1, 57)=2.18, p=0.15;

η^2=0.04), "Framework Conditions" (F (1, 57)=0.02, p=0.61; η^2<0.00), "Teacher-Student Interaction" (F (1, 57)=2.20, p=0.14; η^2=0.04), and "Overall Evaluation" (F (1, 57)=1.81, p=0.18; η^2=0.03). A correlation analysis shows a highly significant correlation between "Mood" and "Overall Grading" with r=-0.50 (p<0.01). Further significant correlations of "Mood" and other dependent measures were found between "Teacher" (r=0.33; p<0.01), "Framework Conditions" (r=0.35; p<0.01), "Teacher-Student Interaction" (r=0.27; p<0.05), "Content" (r=0.43; p<0.01), "Knowledge Acquisition" (r=0.39; p<0.01), and "Overall Impression" (r=0.39; p<0.01). Descriptive results reveal that in all measures except "Framework Conditions", the positive-mood condition leads to better course evaluation results than the negative-mood condition (see Table 2).

Table 2: Mean values (standard deviations in brackets) and inferential statistics for the nine sub-scales of the evaluation questionnaire and for mood from Experiment 2 (N=64).

Dependent Variable	Positive Mood	Negative Mood	Type III Square Sum	F	p	η^2
Knowledge Acquisition	3.38 (0.44)	3.08 (0.59)	1.32	4.94	0.03	0.08
Teacher	4.26 (0.55)	3.87 (0.76)	2.24	5.23	0.03	0.08
Framework Conditions	4.31 (0.54)	4.39 (0.68)	0.10	0.26	0.61	0.01
Workload	3.27 (0.44)	3.11 (0.42)	0.36	1.97	0.17	0.03
Participants	2.99 (0.64)	2.74 (0.70)	0.97	2.18	0.15	0.04
Content	4.30 (0.46)	3.70 (0.75)	5.27	13.99	0.00	0.20
Overall Evaluation	2.58 (0.62)	2.48 (0.62)	0.76	1.81	0.18	0.03
Teacher-Student Interaction	4.06 (0.56)	3.78 (0.87)	1.16	2.21	0.14	0.04
Overall Grading	1.92 (0.47)	2.25 (0.5)	1.61	6.89	0.01	0.11
Mood	3.58 (0.50)	3.21 (0.69)	1.98	5.56	0.02	0.09

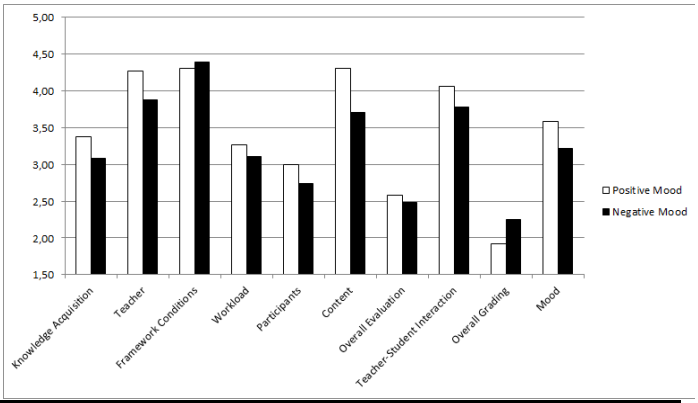

Figure 2: Descriptive results from Experiment 2 (higher values indicate a more positive evaluation except "Overall Grading").

In addition, the control item "Mood" also shows that participants with a pen or pencil between their teeth report a better mood than those with the pen or pencil between their lips (see also Figure 2).

Discussion of Experiment 2

The results from the second study support the assumption that mood contributes to a systematic bias in academic course evaluation. The data obtained here also supports the assumption that the Facial Feedback Hypothesis was applicable and that we have managed to induce a positive or a negative mood. The induced mood seems to have had an influence on information processing, and thus on students' judgments made in this course evaluation. With this second study, we also managed to enhance the psychometric quality of the scales used, which in turn led to acceptable values regarding internal consistency. Interestingly, similarly to the first study, not all sub-scales showed significant group differences. An explanation for this might be that, e.g., "Overall Evaluation", which, in Experiment 1, was in the latter part of the instrument, was moved to the beginning of the questionnaire in the version used for this second study. The duration of the intervention might have been too short in order to induce the intended moods to such a degree that they may have had an influence on participants' judgmental processes. The fact that significant mood differences could be identified in this study supports the effectiveness of the intervention. A problem from an experimental perspective might be that we did not test for mood before the intervention. The major reason for this was that we did not want to prime participants and that we preferred them to be naïve to the aim of the study and the influence of facial feedback. By means of

randomization, we expect to have avoided a sampling bias. In addition, replication of the findings supports the assumption that mood differences result from systematic mood induction rather than from a sampling bias. The high correlation between "Overall Grading" and "Mood" also supports the assumption that there is a strong relationship between these two variables. Although this is a correlation value, data supports a causal relationship suggesting that a positive mood leads to better grading and, vice versa, a negative mood to worse grading. However, not all dimensions of course evaluation are affected. It seems that only those variables that really provide enough variance in the judgmental process seem to be affected. A final problematic issue that should be addressed is the assessment of mood as conducted here. We used a single-item approach that does not allow analysis of internal consistency. However, using a direct single-item assessment appears to be a valid approach here, which is also used in other areas of applied research. One example is the Mental Effort Rating Scale that is frequently used in Cognitive Load Research (cf. Paas, Tuovinen, Tabbers & Van Gerven, 2003). This is also a valid single-item approach that is inspired by Sports Psychology research showing that such an economical approach to collecting data does not automatically lead to inferior psychometric quality than using multi-item scales. However, a cross-validation of the scale as applied here with standardized mood scales should be applied in further controlled studies.

General Discussion

In general, we have found a weak effect. Yet, the theoretical importance of this effect might be greater than its practical value: an intentional misuse of this effect would not change the course evaluation by 180 degrees. But together with other biases, it might add to an unfair evaluation; therefore, awareness of these biases is advisable.

The findings of this research support the assumption, that mood can lead to different judgmental processes with regard to a single target. Although different moods were experimentally induced by means of the Facial Feedback Hypothesis in the present studies, systematic biases in course evaluation could also be likely in everyday academic practice. We would usually expect mood influences to be balanced out by a normal distribution among a specific population. However, systematic environmental factors can contribute to biases due to their impact on human mood (e.g., weather condition, state of lecture theatres, etc.). In both of our experiments, we found that a positive mood leads to better evaluation scores than a negative mood. Across almost all sub-scales of the instruments applied here, a positive mood induced by means of a pen or a pencil between the teeth led to better scores than a negative mood induced by a pen or pencil between the lips.

Explanations can be derived from the Elaboration Likelihood Model (ELM; Petty & Cacioppo, 1986), according to which a positive mood leads to a peripheral route of information processing while a negative mood leads to a central route of information processing. Following the assumptions of Fiedler (2001), positive affect might have led to assimilative processing in the present case, while a negative mood might have triggered accommodative processing. Thus, a negative mood might have caused students to focus on criticizing details of a course while a positive mood might have contributed to a global and mostly positive overall impression. Although we assume that this difference is caused by different types of information processing, we cannot prove that different judgments result from either assimilative or accommodative processing due to a lack of additional measures of whether processing is assimilative and accommodative. However, when looking at the correlations between mood and other dependent measures, it seems likely that mood as induced in these studies leads to different information processing strategies which, in turn, have an impact on course evaluation. Another limitation to be discussed is the fact that participants did not have the basic intervention of these studies during lectures (when comparing the methodology with the experiments of Strack et al., 1988), but only during course evaluation. This suggests that basic information processing while completing the course assessment was not influenced by the intervention. Further criticism relate to the Facial Feedback Hypothesis itself. Firstly, it remains to be clarified how the mechanism as such actually works. Furthermore, recent research by Ito, Chiao, Devine et al. (2006) shows that facial feedback affects implicit rather than explicit judgment processes. It has also been assumed that the effect is due to the simple fact that participants find it easier to keep a pencil between their teeth than between their lips rather than to different muscle positions (cf. McIntosh, 1996). According to McIntosh (1996, p. 132), this criticism can be disproved: "This criticism is refuted by the studies that find reports of effort or difficulty to be unrelated to the manipulation or emotion."

We argue in this respect that the task of completing a course evaluation questionnaire can be regarded as a separate and different task to what is usually performed in lectures. While individuals may randomly reflect on the quality of their course during the lecture period, this reflection is not as focused and systematic as in standardized evaluation settings. What is more, if students monitor the quality of teaching during the lecture period, their observations might also contribute to the outcome of their course evaluation. Assuming that students reflect upon the reasons for deficits in the quality of teaching, their dissatisfaction with a course (e.g., waste of time) might have a negative impact on their mood. According to the ELM, this also might contribute to a more central information processing. Reflecting upon a lecture experienced as 'good' might contribute to the development or maintenance of a positive mood because students get the impression that they really learn and benefit. However, this does not necessarily lead to deep analysis because there is no need for analyzing the reasons. Thus, a more peripheral route of information processing might apply to the assessment of course quality in this case. Following the mood-congruity principle as also stated in the ELM, the intervention applied in the present studies might have led to different judgments because participants with a better mood might have remembered more positive and less negative aspects than students with a more negative mood, who might have remembered several negative issues they had already reflected upon during the term.

Practical Implications

Possible practical implications might be that "true" values of academic course evaluation might be obtained by assessing mood or possible other sources of systematic bias (e.g., how popular a specific subject is perceived by students), and to calculate these variables from the raw data. Without controlling systematic error sources, one might receive biased evaluation values. This is especially problematic when financial resources of a department are dependent on course evaluations – something which has already become common practice at some academic institutions. Future research will have to explore the biases mentioned and described in the present studies as well as other potentially interfering variables.

The nature of the academic courses that were involved in the present studies leads to further limitations. In both samples, the objects of course evaluation were lectures with a minimum of 50 participants. This might constitute a different setting than, e.g., in seminars or other small-group courses. Especially student involvement is usually rather low in lectures compared to more interactive courses. According to Reinhard and Sporer (2010), this might contribute to low task involvement and lead to the use of non-content information instead of intensive processing of content information for course evaluation. This factor should therefore be controlled in subsequent research.

Conclusion

As a matter of course, it is necessary to use instruments for course evaluations that are reliable and valid (see Stehle, 2011; Zhao & Gallant, 2012). Self-made questionnaires should not be used for official purposes. Also, public evaluation databases (like RateMyProfessor) are not recommended (see Legg & Wilson, 2012). But even if high-quality materials are used, the results may still be influenced by other biases. One way to deal with these biases in general is the use of statistical corrections (see Greenwald & Gillmore, 1997, concerning grading leniency). However, this may also be fraught with difficulty.

For discussions with academic policy makers who are willing to approve additional budgets in return for good course evaluations, the reported results demonstrate that the validity of these evaluations depend at least partly on situational effects of the evaluation situation itself. This problem might be resolved by means of higher degrees of standardization, including situational aspects like a "neutral" atmosphere for responding to the questions. Ultimately, it is important to help teachers and educators to improve the quality of their courses - and there are numerous ways of doing that (see, e.g., Keeley, 2012) besides the use of questionnaires. Last but not least: The effect is not restricted to course evaluations. When evaluating any questionnaire (and making any other kinds of decisions), the effects of moods and emotions must be taken into account. In the meantime, we suggest to our colleagues to conduct academic course evaluation on fancy programs, on a sunny day, in small groups, in a well-lit classroom, immediately before the Christmas holidays – and with students having a pen or pencil between their teeth.

References

Abrami, P. C., d'Apollonia, S. & Rosenfield, S. (2007). The dimensionality of student ratings of instruction: What we know and what we do not. In R. P. Perry & J. C.

Smart (Eds.), *The scholarship of teaching and learning in Higher Education: An evidence-based perspective* (pp. 385-456). Dordrecht: Springer.

Bock, F. A. (1979). Student expectations of course content affect: Faculty evaluations in an abnormal psychology course. *Teaching of Psychology*, *6*, 167-169.

Bodenhausen, G. V., Kramer, G. P., & Süsser, K. (1994). Happiness and stereotypic thinking in social judgment. *Journal of Personality and Social Psychology, 66*, 621-632.

Braun, E., & Leidner, B. (2009). Academic course evaluation. *European Psychologist, 14*, 297–306.

Buehner, M. (2010). *Einführung in die Test- und Fragebogenkonstruktion [Introduction into test and questionnaire construction]*. München: Pearson Education.

Daniel, H. (1994). Hörerbefragung an der Universität Mannheim: Konzeption, Erhebung, Auswertung [Students' evaluation at the University of Mannheim: Concept, method, and analysis]. *Empirische Pädagogik [Empirical Pedagogy], 16*, 149-164.

Döring, N. (2005). Für Evaluation und gegen Evaluitis [For evaluation, against evaluitis]. In B. Berendt, H. Voss & J. Wildt (Eds.), *Neues Handbuch Hochschullehre [New handbook of university teaching]* (p.1-22). Bonn: Raabe-Verlag.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes.* New York: Harcourt Brace & Jovanovich.

Feldman, K. A. (1984). Class size and college students' evaluations of teachers and courses: A closer look. *Research in Higher Education*, *21*(1), 45-116. doi:10.1007/BF00975035

Feldman, K. A. (1998). Reflections on the Study of Effective College Teaching and Student Ratings: One Continuing Quest and Two Unresolved Issues. In J. C. Smart (Ed.), *Higher Education: Handbook of Theory and Research, Vol. 13* (pp. 35-74). New York: Agathon Press.

Fiedler, K. (2001). Affective states trigger processes of assimilation and accommodation. In L. L. Martin & G. L. Clore (Eds.), *Theories of mood and cognition: A User's Guidebook* (pp. 85-98). Hillsdale, NJ: Erlbaum

Forgas, J. P. (1995). Mood and judgment: The Affect Infusion Model (AIM). *Psychological Bulletin, 117*, 39-66.

Forgas, J. P., Goldenberg, L., & Unkelbach, C. (2009). Can bad weather improve your memory? An unobtrusive field study of natural mood effects on real-life memory. *Journal of Experimental Social Psychology, 54*, 254-257.

Gawronski, B., & Bodenhausen, G. V. (2005). Accessibility effects on implicit social cognition: The role of knowledge activation and retrieval experiences. *Journal of Personality and Social Psychology, 89*, 672-685.

Gillmore, G. M., & Greenwald, A. G. (1999). Using statistical adjustment to reduce biases in student ratings. *American Psychologist, 54*, 518-519.

Greenwald, A. G. (1997). Validity concerns and usefulness of student ratings of instruction. *American Psychologist, 52*, 1182–1186.

Greenwald, A. G., & Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist*, *52*, 1209-1217.

Greenwald, A. G., & Gillmore, G. (1998). How useful are student ratings? *American Psychologist, 53,* 1228–1229.

Griskevicius, V., Shiota, M. N., & Neufeld, S. L. (2010). Influence of different positive emotions on persuasion processing: A functional evolutionary approach. *Emotion, 10*, 190–206.

Havas, D. A., Glenberg, A. M., Gutowski, K. A., Lucarelli, M. J., & Davidson, R. J. (2010). Cosmetic use of Botulinum Toxin-A affects processing of emotional language. *Psychological Science, 21*, 895-900.

Huntsinger, J. R., Clore, G. L., & Bar-Anan, Y. (2010). Mood and global–local focus: Priming a local focus reverses the link between mood and global–local processing. *Emotion, 10*, 722-726.

Inbar, Y., Cone, J., & Ginovich, T. (2010). People's intuitions about intuitive insight and intuitive choice. *Journal of Personality and Social Psychology, 99*, 232–247.

Ito, T. A., Chiao, K. W., Devine, P. G., Lorig, T. S., & Cacioppo, T. T. (2006). The influence of facial feedback on race bias. *Psychological Science, 17,* 256-261.

Keeley, J. W. (2012). Course and instructor evaluation. In W. Buskist & V. A. Benassi (Eds.), *Effective college and university teaching: Strategies and tactics for the new professoriate* (pp. 173-180). Thousand Oaks, CA: Sage Publications.

Kleinke, C. L., Peterson, T. R., & Rutledge, T. R. (1998). Effects of self-generated facial expressions on mood. *Journal of Personality and Social Psychology, 74*, 272-279.

Kromrey, H. (1995). Evaluation. Empirische Konzepte zur Bewertung von Handlungsprogrammen und die Schwierigkeit ihrer Realisierung [Evaluation. Empirical concepts for analyzing action programs and the difficulty of their implementation]. *Zeitschrift für Sozialisationsforschung und*

Erziehungssoziologie [Journal of Research in Socialization and Sociology in Education], 15, 313-336.

Kromrey, H. (1996). Qualitätsverbesserung in Lehre und Studium statt sogenannter Lehrevaluation. Ein Plädoyer für gute Lehre und gegen schlechte Sozialforschung [Improving quality in teaching and studies instead of course evaluation. A plea for good teaching and against bad social science research]. *Zeitschrift für Pädagogische Psychologie [German Journal of Educational Psychology], 10*, 153-166.

Legg, A. M., & Wilson, J. H. (2012). RateMyProfessors.com offers biased evaluations. *Assessment & Evaluation in Higher Education, 37*(1), 89-97.

Marsh, H. W. & Roche, L. A. (1999). Rely upon SET research. *American Psychologist, 54*, 517-518.

Mateo, M. A., & Fernández, J. (1996). Incidence of class size on the evaluation of university teaching quality. *Educational and Psychological Measurement, 56*(5), 771-778.

McIntosh, D. N. (1996). Facial feedback hypotheses: Evidence, implications, and directions. *Motivation and Emotion, 20*, 121-147.

McPherson, M. B., Kearney, P., & Plax, T. G. (2003). The dark side of instruction: Teacher anger as classroom norm violations. *Journal of Applied Communication Research, 31*(1), 76-90.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. (2003). Cognitive load measurement as a means to advance Cognitive Load Theory. *Educational Psychologist, 38*, 63-71.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology, 19,* 123–205.

Reinhard, M.-A., & Sporer, S. L. (2010). Content versus source cue information as a basis for credibility judgments. *Social Psychology, 41*, 93–104.

Reiß, S., Mildner, D., Nagler, H., & Schweizer, K. (2010). Teilnahmemotive an universitären Vorlesungen in Abhängigkeit vom Lehrveranstaltungsinhalt und Erfassungszeitpunkt [Motives for participation in academic lectures depending on lecture content and time of assessment]. Paper presented at the 8th Meeting for Didactics of Psychology and Evaluation, May 14-15, 2010, Frankfurt a.M., Germany.

Rindermann, H. & Schofield, N. (2001). Generalizability of multidimensional student ratings of university instruction across courses and teachers. *Research in Higher Education, 42*, 377-399.

Rosenthal, R. (1991). *Meta-analytic procedures for social research.* Newbury Park: Sage.

Schwarz, N., Bless, H., & Bohner, G. (1991). Mood and persuasion: Affective states influence the processing of persuasive communications. *Advances in Experimental Social Psychology, 24*, 161-199.

Schwarz, N., & Clore, G. L. (2007). Feelings and phenomenal experiences. In E. T. Higgins & A. Kruglanski (Eds.), *Social psychology. A handbook of basic principles* (2nd ed., pp. 385-407). New York: Guilford Press.

Spinath, B., & Stehle, S. (2011). Evaluation von Hochschullehre [Evaluation in academic teaching]. In L. Hornke & M. Amelang (Eds.), *Enzyklopädie der Psychologie, Grundfragen und Anwendungsfelder psychologischer Diagnostik: Psychologische Diagnostik 1* [Encyclopedia of Psychology: Foundations and application of psychological diagnostics] (Vol. 1, pp. 617-667). Göttingen: Hogrefe.

Spooren, P., Mortelmans, D., & Thijssen, P. (2012). 'Content' versus 'style': Acquiescence in student evaluation of teaching? *British Educational Research Journal, 38*, 3-21.

Stehle, S. (2011). *Quality development of university teaching: The role of student evaluations* (Doctoral Dissertation). Heidelberg: Heidelberg University, Department of Psychology.

Strack, F., Martin, L., & Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: A non-obtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology, 54,* 768-777.

Strack, F., & Neumann, R. (2000). Furrowing the brow may undermine perceived fame: The role of facial feedback in judgments of celebrity. *Personality and Social Psychology Bulletin, 26*, 762-768.

Svanum, S., & Aigner, C. (2011). The influences of course effort, mastery and performance goals, grade expectancies, and earned course grades on student ratings of course satisfaction. *British Journal of Educational Psychology, 81*, 667-679.

Zhao, J., & Gallant, D. J. (2012). Student evaluation of instruction in higher education: Exploring issues of

validity and reliability. *Assessment & Evaluation in Higher Education, 37*, 227-235.

Zumbach, J., Spinath, B., & Schahn, J. (2007). Eine Kurzskala zur Lehrevaluation [A short scale for academic course evaluation]. In M. Krämer, S. Preiser & K. Brusdeylins (Eds.), *Psychologiedidaktik und Evaluation VI [Didactics of Psychology and Evaluation VI]* (pp. 317-126). Göttingen: V&R unipress.

Appendix

Items of the questionnaire as used in Experiment 2 (Stehle, 2011).

Knowledge Acquisition
By taking this course…
… I acquired broad working knowledge of the subject matter
… I acquired specific factual knowledge of the subject matter
… I acquired functional/practical knowledge of the subject matter
… I learned how to identify complex interrelations within the domain
… I learned how to apply the acquired knowledge (to new contexts)
… I learned how to apply the discussed methods autonomously
… I learned how to work independently
… I learned how to deal with scientific texts
Content
The course contents build on my previous knowledge
The course contents are up to date
The course contents are highly relevant
The course gives a good overview of the subject matter
Teacher
Instructor clearly communicated the course objectives
Instructor gives the course an appropriate structure
Instructor is adequately prepared for each class
Instructor adequately presents and explains course contents
Instructor is able get highly complex and difficult contents across to students
Teacher-Student Interaction
Instructor encourages students to express own ideas and/or question the instructor
Instructor encourages students to ask questions and gives meaningful answers
Instructor is available for advice in or outside of class
Instructor provides students with valuable resources that improve understanding
Instructor accepts constructive criticism
Instructor creates a motivating and inspiring atmosphere in class
Workload
In my perception, the course level is...
In my perception, the amount of material covered by this course is …
In my perception, the course pace is ...
In my perception, the workload required for this course is …
To follow the course, students need to invest considerable time studying at home
Participants
The majority of students attend the class regularly
The majority of students are well prepared for each class
The majority of students follow the course with interest and attention

Framework Conditions
The room where the course takes place is adequate
The technical facilities (i.e. media. etc.) are adequate
The time frame is adequate (date, duration, interference with other courses)
Overall Evaluation
My interest in the subject has increased as a consequence of this course
I consider my personal learning gains in this seminar to be as follows
Overall I rate the course with the following school grade
Overall Grading
Taken together, I would grade this course as 1 (very good) – 2 (good) – 3 (satisfactory) – 4 (sufficient) – 5 (failed)

Note. Scale (except Overall Grading): 1 = totally disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = totally agree.

Citation:

Zumbach, Joerg & Joachim Funke (2014). Influences of mood on academic course evaluations. *Practical Assessment, Research & Evaluation*, 19(4). Available online: http://pareonline.net/getvn.asp?v=19&n=4.

Authors:

Joerg Zumbach
University of Salzburg, Austria
Email: joerg.zumbach [at] sbg.ac.at

Joachim Funke
University of Heidelberg, Germany
Email: joachim.funke [at] psychologie.uni-heidelberg.de