

Emotional labor, strain and performance: Testing reciprocal relationships in a longitudinal panel study

Citation for published version (APA):

Hülshager, U. R., Lang, J. W. B., & Maier, G. W. (2010). Emotional labor, strain and performance: Testing reciprocal relationships in a longitudinal panel study. *Journal of Occupational Health Psychology, 15*(4), 505-521. <https://doi.org/10.1037/a0021003>

Document status and date:

Published: 01/10/2010

DOI:

[10.1037/a0021003](https://doi.org/10.1037/a0021003)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Emotional Labor, Strain, and Performance: Testing Reciprocal Relationships in a Longitudinal Panel Study

Ute R. Hülshager and Jonas W. B. Lang
Maastricht University

Günter W. Maier
Bielefeld University

Models of emotional labor suggest that emotional labor leads to strain and affects job performance. Although the link between emotional labor, strain, and performance has been well documented in cross-sectional field studies, not much is known about the causal direction of relationships between emotional labor, strain, and performance. Goal of the present study was therefore to test the direction of effects in a two-wave longitudinal panel study using a sample of 151 trainee teachers. Longitudinal lagged effects were tested using structural equation modeling. Results revealed that the emotional labor strategy of surface acting led to increases in subsequent strain while deep acting led to increases in job performance. In contrast, there was no indication of reverse causation: Neither strain nor job performance had a significant lagged effect on subsequent surface or deep acting. Overall, results support models of emotional labor suggesting that surface and deep acting causally precede individual and organizational well-being.

Keywords: emotional labor, strain, job performance, longitudinal, cross-lagged panel study; structural equation modeling

Most jobs involving interactions with others, for example, waiters, flight attendants, salespersons, nurses, doctors, police officers, teachers, and academics, require that employees control and regulate their emotions in order to display certain emotions toward their counterpart, be it clients, patients, suspects, or students. A waiter might have to stay calm and considerate toward an angry customer, a doctor might have to display confidence although she is worried about the patient's state of health, and a teacher needs to be friendly, understanding, and patient although he might struggle with personal problems at home. These emotional job requirements are summarized under the term of emotional labor

(Grandey, 2000; Rafaeli & Sutton, 1987; Zapf, 2002). Although emotional labor might in principle involve all four possible combinations of amplification versus suppression of positive versus negative emotions, display rules mostly require suppressing negative emotions while enhancing positive ones. In order to comply with display rules—which define norms and standards for the appropriate expression of emotions (Rafaeli & Sutton, 1987)—individuals use different strategies to regulate their emotions (Diefendorff & Gosserand, 2003; Grandey, 2000). According to the integrative model proposed by Grandey (2000), the process of managing feelings and emotional expressions is the core of emotional labor, because it addresses the way employees deal with emotional job requirements. Within so-called employee-focused emotional labor (Brotheridge & Grandey, 2002) two emotional labor strategies are distinguished, namely surface acting and deep acting: While surface acting is limited to the management of observable expressions, deep acting involves the intentional management of internal feelings which in turn leads to a modification of the observable expression.

Arlie Russell Hochschild (1983)—who coined the term “emotional labor” and who was the first to direct scientific attention to this topic—argued that emotional labor would always be a burden to those showing it. In recent years it has, however, been shown that emotional labor is not always associated with

Ute R. Hülshager and Jonas W. B. Lang, Department of Work and Social Psychology, Maastricht University; and Günter W. Maier, Department of Work and Organizational Psychology, Bielefeld University.

We thank Rafaela Strube and Christiane Wolff for their assistance in data collection. Furthermore, we would like to express our gratitude to Detlev Gadow from Studienseminar für Lehrämter Bielefeld, Wolfgang Wähl, and Rainer Quadflieg from Studienseminar für Lehrämter Arnsberg, as well as Bernd Bareuther from Verband Bildung und Erziehung Hochsauerlandkreis for facilitating data collection. Finally, we are indebted to Paula Niemietz for her help in translating the emotional labor scale.

Correspondence concerning this article should be addressed to Ute R. Hülshager, Department of Work and Social Psychology, Maastricht University, P.O. Box 626, 6200 MD Maastricht, the Netherlands. E-mail: ute.hulshager@maastrichtuniversity.nl

negative psychological outcomes. Research suggests that consequences depend on the strategies individuals adopt to regulate their emotions (Côté, 2005). On one hand, strong negative associations have been documented between surface acting and psychological health outcomes. On the other hand, research revealed no or only weak negative associations between deep acting and psychological health but positive associations between deep acting and performance (e.g., Brotheridge & Lee, 2002; Brotheridge & Lee, 2003; Glomb & Tews, 2004; Grandey, 2003; Martínez-Iñigo, Totterdell, Alcover, & Holman, 2007; Totterdell & Holman, 2003).

The vast majority of evidence accumulated so far stems from cross-sectional, correlational field studies. Although these allow detecting associations between variables they do not provide any information on the direction of effects (Frese & Zapf, 1988; Spector, 2001; Zapf, Dormann, & Frese, 1996). Theoretical models of emotional labor conceptualize surface acting as a source of stress that causes strain and emotional exhaustion. Likewise, deep acting is seen as a strategy that causally precedes enhanced service performance (Grandey, 2000; Holman, Martínez-Iñigo, & Totterdell, 2008a; Rubin, Staebler Tardino, Daus, & Munz, 2005).

Although the described direction of effects is plausible, the observed associations might also be explained by alternative, specifically reverse causal pathways. Employees experiencing strain might, for instance, more likely engage in surface and less likely engage in deep acting. Researchers have therefore long been calling for longitudinal field studies that allow testing directions of effects in occupational health psychology (de Jonge et al., 2001; Frese & Zapf, 1988; Zapf et al., 1996).

The goal of the present study was therefore to conduct a longitudinal panel study that allows testing the suggested direction of effects against alternative explanations of reverse causation. In doing so, we focus on two core outcome variables in the emotional labor literature, namely strain and performance. We thereby cover the two outcome domains of individual and organizational well-being described in Grandey's (2000) model of emotional labor.

In the following we will elaborate on the two emotional labor strategies of deep and surface acting. Next, we will discuss alternative explanations for the link between surface and deep acting on the one and strain and performance on the other hand. We will start with reviewing theoretical arguments for the most plausible direction of effects, that is, emotional labor strategies lead to strain and performance. Next,

we discuss alternative views, namely the influence of strain and performance on emotional labor strategies (reverse causation).

Emotional Labor Strategies: Deep Acting and Surface Acting

In her seminal book, Hochschild (1983) was the first to differentiate two ways employees can manage their emotions in order to comply with organizational display rules: deep acting and surface acting. In the following we refer to deep and surface acting as emotional labor strategies (Diefendorff, Croyle, & Gosserand, 2005). They are specific types of emotion regulation strategies used by employees when interacting with clients as part of required job role behavior. Notably, we use "clients" here in a broad sense, referring to individuals interacting with employees, for instance customers, patients, students, suspects, defendants, debtors, or passengers (Zapf, 2002). Grandey (2000) linked deep and surface acting to emotion regulation theory (Gross, 1998) thereby providing valuable insights in the basic processes underlying these emotional labor strategies. Deep acting is an antecedent-focused form of emotion regulation in that it operates right at the onset of an emotion and affects the perception and processing of emotional cues before they elicit behavioral, experiential, or physiological responses (Gross, 1998). The goal of deep acting lies in aligning required and true feelings. To reach this goal, employees can use techniques to modify their perception of a situation: attention deployment, that is, directing attention toward pleasurable things or thoughts to stir up the required emotion, or cognitive change, that is, reappraising the situation to induce the required emotion. Surface acting is a form of response-focused emotion regulation. It comes in at a later point in time, when the emotion has already developed. Not the emotion itself, but only the observable reaction, the public display of the emotion is altered. The adjustment of the emotional response can be achieved by suppressing, amplifying, or faking emotions. In consequence, the emotional experience and the emotion expression remain discordant when engaging in surface acting (Grandey, 2000; Gross, 1998; Totterdell & Holman, 2003).

Emotional Labor Strategies, Strain, and Job Performance

Extant research provides ample evidence for the association between emotional labor strategies and

strain. These lines of research have been summarized in a quantitative review (Bono & Vey, 2005) that has recently been updated and extended in a comprehensive meta-analysis (Hülshager & Schewe, 2010). Findings suggest that surface acting is substantially related to psychological strain, emotional exhaustion, and depersonalization (corrected correlations in the .40s). Deep acting, in contrast, was uncorrelated with strain, emotional exhaustion, and depersonalization (Hülshager & Schewe, 2010). Due to the difficulties involved in collecting performance data, fewer studies have investigated the relationship between surface and deep acting and performance. Only a small number of studies have been published that rely on supervisor or peer rather than self-ratings of performance (Duke, Goodman, Treadway, & Breland, 2009; Grandey, 2003). So far, two field studies have revealed significant positive associations between deep acting and performance (Grandey, 2003; Totterdell & Holman, 2003). With regard to surface acting and performance, findings have been mixed. While Totterdell and Holman (2003) found no significant association, Grandey (2003) reported a significant negative association between surface acting and affective delivery, but not between surface acting and breaking character.

In sum, research suggests that surface acting is positively related to psychological strain and, potentially, negatively related to performance. Deep acting, in contrast, seems to be unrelated to psychological strain, but displays positive associations with performance. Yet, evidence accumulated so far stems from cross-sectional field studies that do not provide any information on whether emotional labor strategies precede strain and performance or whether strain and performance precede emotional labor strategies. Inferences on directions of effects can thus not be drawn (Finkel, 1995; Spector, 2001; Zapf et al., 1996).

The goal of the present study is therefore to test the direction of the relationship between the two emotional labor strategies of deep and surface acting with strain and performance with a full two-wave panel study. Although the ultimate test of causal relationships can only be accomplished in an experimental setting, longitudinal studies are an appropriate mean to shed light on causal relations and to rule out third-variable phenomena in a field setting (Spector, 2001; Zapf et al., 1996). They allow to test whether the potential antecedents (here, surface acting and deep acting) precede the hypothesized effect (here, on strain and performance). Apart from testing the hypothesized relationship over time we will test

for effects of reverse causation. Testing the hypothesized as well as recursive directions of effects is an important prerequisite to make inferences on causal relationships (Finkel, 1995).

Emotional Labor Strategies Precede Strain and Job Performance

Extant research and models of emotional labor imply that surface acting leads to strain and performance impairments while deep acting positively affects performance (Grandey, 2000; Rubin et al., 2005). Consequently, surface and deep acting are usually conceptualized as independent variables, whereas strain and performance are the dependent variables. Several mechanisms have been discussed in the literature suggesting the direction of effects to be such that surface and deep acting causally precede strain and performance.

Surface acting leads to strain. With regard to surface acting and its relationship with strain, five core-mechanisms have been suggested that may explain how surface acting leads to strain. First, researchers have argued that just as any other self-regulatory process, suppressing and faking emotions is effortful and depletes mental resources (Baumeister, Bratslavsky, Muraven, & Tice, 1998; Hochschild, 1983; Martinez-Iñigo et al., 2007). Research building on Baumeister and colleagues' ego-depletion model (Baumeister et al., 1998; Muraven, Tice, & Baumeister, 1998) has revealed that the need to alter the expression of one's emotions continuously entails psychological effort and depletes mental resources (Richards & Gross, 1999, 2000). Resource loss is stressful and seen as one of the key reasons why surface acting makes people feel drained and exhausted (Côté, 2005; Martinez-Iñigo et al., 2007).

Second, as surface acting alters the facial and bodily expression while leaving the true emotion intact, the individual still feels the original inner emotion that is only suppressed and disguised by a faked outer expression. As emotional labor usually involves the suppression of negative emotions, the individual will continue to experience this very negative emotion which remains unresolved, surviving beneath the masked face and thus continuing to negatively affect an individual's feelings and psychological state of health (cf. Gross & John, 2003).

Third, suppressing one's own true negative feelings and pretending to have positive feelings result in a mismatch between outer expressions and inner feelings. This, in turn, leads to a state of tension. In the

field of emotion regulation research, it has been shown that suppressing negative emotions leads to enhanced sympathetic activation of the cardiovascular system (Gross, 1998; Richards & Gross, 1999) which can be interpreted as a physiological indicator of this inner tension. This physiological inner tension may be experienced as strain and may lead to emotional exhaustion and burnout in the long run.

Fourth, it has been argued that surface acting makes people feel incongruent and corrupts their sense of authenticity (Brotheridge & Lee, 2002; Erickson & Ritter, 2001). The emotional display generated by surface acting is a conscious façade rather than an authentic form of self-expression. Surface acting thus lowers the actor's sense of personal authenticity which, in turn, has been shown to be associated with depression, perceived stress, and self-esteem (Sheldon, Ryan, Rawsthorne, & Ilardi, 1997).

Fifth, in addition to these intrapersonal mechanisms presented so far, an interpersonal mechanism has been proposed. In his social interaction model of emotion regulation at work, Côté (2005) drew attention to interpersonal processes that help understand how emotional labor impacts psychological strain. Building on an emotional transactions framework (Rafaeli & Sutton, 1987), the model suggests that the sender's emotional display is appraised by the target person (e.g., the customer) who in turn responds accordingly and thereby reffects the sender and his or her emotional, psychological, and physical state. As authentic differ from inauthentic emotion displays (Ekman, 2003; Ekman, Friesen, & O'Sullivan, 1988; Grandey, Fisk, Mattila, Jansen, & Sideman, 2005) and as interaction partners are able to differentiate between the two (Grandey et al., 2005), they react more unfavorably to inauthentic compared to authentic displays of positive emotions. Surface acting—which results in inauthentic emotion displays—is therefore obstructive to positive interactions and evokes overall negative reactions from interaction partners. These negative reactions (e.g., anger expressed by a customer) are stressors that reffect the sender and lead to psychological strain. Supporting this notion, Brotheridge and Lee (2002) provided evidence that the quality of relationships with clients mediated the positive relationship between surface acting and emotional exhaustion and depersonalization.

In light of these arguments we expect surface acting to precede strain. We will test this in a longitudinal panel study where we expect surface acting to display a positive cross-lagged panel correlation with strain. Put differently, we expect Time 1 surface

acting to display positive correlations with Time 2 strain, controlling for Time 1 levels of strain.

Hypothesis 1: Surface acting has a positive lagged effect on strain.

Surface acting leads to impaired job performance. Ego depletion might also explain why surface acting leads to impaired job performance. The ego depletion model (Baumeister et al., 1998; Muraven et al., 1998) posits that any self-regulatory behavior draws on a limited supply of energy. When this supply diminishes it impairs subsequent activities that involve self control or executive functioning. Building on the ego depletion model, Richards and Gross (2000, 1999) argued that response-focused emotion regulation is cognitively demanding and depletes mental resources. In a series of experiments they illustrated that suppressing the expression of negative feelings impaired subsequent memory performance and performance on subsequent tasks. Similar to emotion suppression, individuals engaging in surface acting have to constantly monitor their emotion expressions and alter them so that they are in line with display rules. This process involves self control, it depletes mental resources and detracts attention from other cognitive tasks employees have to perform as part of their job (Holman, Martinez-Iñigo, & Totterdell, 2008b). Accordingly, Sideman-Goldberg and Grandey (2007) showed that participants that had to fulfill positive display rules in a call center simulation task made more errors during service interactions. Similarly, Shull, Wallace, Edwards, and Kantowitz (2006, as cited in Holman et al., 2008b) revealed that in comparison to reappraisal (similar to deep acting) suppression (similar to surface acting) requires more resources and diminishes task focus, thereby leading to comparatively lower performance.

Furthermore, surface acting is thought to influence performance through the perceived inauthenticity of the emotional display (Grandey, 2003; Grandey et al., 2005; Groth, Hennig-Thurau, & Walsh, 2009). Satisfying organizational display rules is an important element of task performance. Flight attendants who are cheerful and friendly toward their customers and nurses who show sympathy toward their patients fulfill their work roles successfully and display higher overall job performance. Supporting this notion, research has documented that employee affective delivery (i.e., displayed positive emotions) is associated with positive customer reactions, an important indicator of service performance (Tsai, 2001; Tsai & Huan, 2002). It has, however, been argued

that the display of positive emotions is not sufficient but that emotion expressions have to be authentic to elicit positive customer reactions (Grandey et al., 2005; Hennig-Thurau, Groth, Paul, & Gremler, 2006). Authentic emotional displays differ from fake ones (Ekman et al., 1988) and experimental research on emotions has shown that individuals react more favorably to authentic (Duchenne) than to inauthentic smiles (Frank, Ekman, & Friesen, 1993). Although employees try to express the appropriate emotions while surface acting, they consciously fake the emotional display and the result is an inauthentic emotion expression. Experimental research demonstrates that the authenticity of emotional displays affects important customer service outcomes, such as customer-employee rapport and customer positive affect (Hennig-Thurau et al., 2006). Researchers have even argued that inauthentic emotion displays do not only yield less positive reactions than authentic displays, but they may even produce negative reactions from customers (Grandey, 2003; Groth et al., 2009). For instance, an inauthentic display of friendliness and sympathy toward an angry customer may be interpreted as dishonesty and deceit and signal to the customer that the employee is not truly motivated to help him or her solve the problem. This may elicit negative customer reactions and negatively affect behavioral outcomes such as customer loyalty. We therefore expect that due to the inauthenticity of the emotional display, surface acting negatively affects job performance.

Hypothesis 2: Surface acting has a negative lagged effect on job performance.

Deep acting leads to enhanced job performance.

Similarly to surface acting and job performance, the authenticity of the emotional display is the key mechanism that may explain how deep acting positively affects performance.

According to the Emotion as Social Information Model (van Kleef, 2009, 2010), emotions facilitate social interaction by conveying important information about the expresser's feelings, attitudes, and intentions. Thereby, emotional expressions help to influence interaction partners; they trigger affective reactions and provide the observer with information about the social situation. In contrast to spontaneous emotional expressions, emotional display rules may elicit emotional expressions that are more or less authentic. However, only genuine expressions entail the relevant cues that serve important social functions and have the desired effects on other individuals

(Keltner & Haidt, 1999). For instance, the genuine display of understanding and sympathy signals to an angry customer the intention to be friendly and to help him or her with a problem and this may in turn elicit favorable customer reactions. Accordingly, experimental and field research has documented that positive emotional displays elicit more favorable customer reactions in terms of customer satisfaction and customer-employee rapport when they are perceived as authentic (Grandey et al., 2005; Hennig-Thurau et al., 2006). When employees engage in deep acting they adjust their inner emotional states with their outer expressions. The positive emotions displayed, are consequently authentic reflections of their true feelings and attitudes and they should therefore elicit positive customer reactions. Employees regulating their emotions with deep acting thereby fulfill their work roles successfully and display high service performance.

Similarly, the authentic expression of appropriate emotions is essential for the fulfillment of primary job tasks in professions where emotional display rules do not only require the expression of positive but also of negative emotions, for example, in the teaching professions. Teachers' emotions that are perceived as genuine by students convey important information and thereby help teachers fulfill their teaching tasks (Sutton & Wheatley, 2003). For instance, the expression of anger can help teachers enforce proper classroom conduct and the expression of disappointment in students' grades can influence their attributional style and thereby affect their motivation (Sutton & Wheatley, 2003). In sum, deep acting leads to the authentic expression of emotions and thereby facilitates the fulfillment of primary job tasks which should result in higher overall job performance.

These lines of reasoning suggest that deep acting precedes performance. We therefore expect that Time 1 deep acting displays a positive relationship with Time 2 performance controlling for initial performance levels.

Hypothesis 3: Deep acting has a positive lagged effect on job performance.

Deep acting and strain. Numerous cross-sectional studies have investigated the link between deep acting and psychological strain or related outcomes such as emotional exhaustion (Brotheridge & Grandey, 2002; Brotheridge & Lee, 2002; Grandey, 2003; Martinez-Iñigo et al., 2007; Prati, 2004; Totterdell & Holman, 2003). Researchers have sug-

gested that although deep acting involves less effort than surface acting, managing emotions by deep acting still is an effortful regulatory process that depletes mental resources and may thereby lead to exhaustion (Grandey, 2003; Holman et al., 2008a; Totterdell & Holman, 2003). However, deep acting does not only result in a resource loss but also in a resource gain: Deep acting promotes self-authenticity and helps creating rewarding social interactions with customers (Brotheridge & Lee, 2002; Holman et al., 2008a; Martinez-Iñigo et al., 2007). It has been argued that these different processes might compensate for one another so that there is no net gain or loss for deep acting (Martinez-Iñigo et al., 2007). Indeed, many studies have failed to find a significant positive relationship between deep acting and emotional exhaustion or psychological strain (Brotheridge & Grandey, 2002; Brotheridge & Lee, 2002; Grandey, 2003; Martinez-Iñigo et al., 2007; Prati, 2004; Totterdell & Holman, 2003). This has been confirmed in a recent meta-analysis revealing mean corrected correlations of about 0 between deep acting and emotional exhaustion and psychological strain (Hülshager & Schewe, 2010). We therefore analyze the relationship without presenting a specific hypothesis regarding the relation between deep acting and strain.

Alternative Directions of Effects—Reverse Causation

Researchers have convincingly argued that stressor-strain relationships might not only be explained by the effects of stressors, but also by reverse or mutual causation (Garst, Frese, & Molenaar, 2000; Zapf et al., 1996). In consequence, although arguments suggesting that deep and surface acting are the causal precedents of strain and performance are compelling, alternative theoretical explanations might also account for the relationship between deep and surface acting, strain, and performance.

Research on the antecedents of emotional labor suggests that individuals' state of affect influences the way emotions are regulated. Consequently, strain might be a causal precedent of surface acting. Individuals with different affective tones differ in their perception of emotional display rules which, in turn, influences their tendency to regulate their emotions by deep or surface acting (Diefendorff & Richard, 2003; Gosserand & Diefendorff, 2005). Individuals high in negative affectivity perceive more demands to suppress negative emotions than individuals low in negative affectivity (Schaubroeck & Jones, 2000).

The contention is that, as individuals high in negative affectivity feel negative emotions more frequently and more intensively, organizational display rules are frequently at odds with their natural feelings. Consequently, they experience more demands to suppress negative emotions at work. As a result they need to regulate their emotions more often than individuals low in negative affectivity and they often do so by surface acting. The idea has been corroborated by a number of cross-sectional field studies revealing a positive association between trait negative affectivity (or neuroticism) and surface acting and a negative association between trait positive affectivity (or extraversion) and surface acting (Brotheridge & Grandey, 2002; Brotheridge & Lee, 2003; Diefendorff et al., 2005; Gosserand & Diefendorff, 2005).

Psychological strain is an adverse individual experience that is characterized by anxiety, unhappiness, anger, and irritability, among others (Beehr & Newman, 1978). Thus, individuals suffering from strain are in a negative state of affect and they might therefore be more likely to regulate their emotions by surface acting.

Similarly, performance might impact surface acting and deep acting through its effect on positive affect. Expectancy theories of motivation suggest that performance is associated with intrinsic rewards satisfying needs of self-actualization and inducing positive feelings of success and personal accomplishment (Judge, Thoresen, Bono, & Patton, 2001; Lawler & Porter, 1967). As noted earlier, individuals experiencing more positive emotions tend to engage less often in surface acting and more often in deep acting (Gosserand & Diefendorff, 2005). Thus, by fostering positive feelings, performance might reduce employees' tendency to engage in surface acting and increase their inclination to use deep acting.

The above lines of arguments stand counter to models of emotional labor (Grandey, 2000; Rubin et al., 2005). They suggest causal pathways to be such that strain and performance influence subsequent surface and deep acting. To rule out these alternative explanations we will specify a full cross-lagged panel model that does not only account for the hypothesized directions of effects but also for these reverse pathways.

Method

Participants and Procedure

The sample consisted of trainee teachers working in elementary and secondary schools in North Rhine-

Westphalia, Germany. After completing their studies of Teacher Education with the First State Examination at a German university, German trainee teachers get appointed as teachers-in-training for a period of 2 years by the Ministry of Education. During that period of preparatory service, further practical pedagogical training takes place on and off the job and they undergo regular evaluation by senior teachers. Throughout each week, trainee teachers work at the schools to which they have been appointed. On a weekly basis, trainee teachers follow advanced professional training at specialized institutions which are under the control of the Ministry of Education.

Data were collected at these centers of advanced training at two time points, lying 2 months apart. Two months are time frames long enough to allow emotional labor strategies, strain, and job performance to change but still short enough to detect short-term longitudinal effects. All variables were measured at both time points. A total of 214 trainee teachers agreed to participate and provided valid questionnaires at Time 1. Overall 151 trainee teachers (70.6% of Time 1) with a mean age of 27.4 years ($SD = 1.2$ years) continued their participation at Time 2. One-hundred and 11 participants (73.5%) were women and 39 were men (25.8%). One participant did not provide any information on gender. To estimate whether drop-outs and the follow-up sample differed in study characteristics, we compared the two groups with regard to age and gender, and all measures assessed at Time 1. With two exceptions, there were no significant differences between groups. However, drop-outs were older ($M = 29.95$, $SD = 5.62$, $p < .01$) and reported more strain ($M = 2.58$, $SD = .97$, $p < .01$) than participants who continued their participation (age: $M = 27.45$, $SD = 3.33$; strain: $M = 2.17$, $SD = .86$).

Measures

Emotional labor strategies. To assess deep acting and surface acting we used Brotheridge and Lee's (2003) 6-item Emotional Labor Scale (3 items for each subscale). It was translated to German by the first author and a bilingual German American philologist in a translation-backtranslation process. All items were answered on a scale ranging from 1 (never) to 5 (always). A sample item for surface acting read "How often do you pretend to have emotions that you don't really have?" and for deep acting "How often do you try to actually experience the emotion that you must show?"

Psychological strain. Strain was assessed with the emotional strain component of the irritation scale developed by Mohr and colleagues (Mohr, Müller, Rigotti, Aycan, & Tschan, 2006). The subscale consists of five items assessing emotional irritation and mental strain, which are seen as precursors of more serious psychological impairments such as psychosomatic complaints or depression. The scale has previously been employed in a variety of cross-sectional and longitudinal research studying the consequences of work stressors (e.g., Dormann & Zapf, 2002; Frese, 1999; Schyns, 2006). Sample items are "From time to time I feel like a bundle of nerves," "I get irritated easily, although I don't want this to happen." Items were rated on a scale ranging from 1 (not at all true) to 5 (completely true).

Job performance. As trainee teachers undergo regular formal evaluation by senior teachers (called mentors) and supervisors as part of their 2-year professional development program, we assessed job performance by asking them about these evaluations with the following two items: "What grade did you receive in your last evaluation?" and "Overall, how does your mentor evaluate your performance?" In accordance with the evaluation scale used in the teacher development program, they indicated their answer on a 5-point scale ranging from "very good" to "not sufficient." High values on that scale indicate high performance. Our measure of job performance can thus best be described as a self-reported supervisor rating and comes close to an independent measure of job performance. Trainee teachers did not indicate their own evaluation of job performance, but their supervisor's evaluation. For a subsample of 89 participants from Time 1 we were able to compute the concordance between the present measure of job performance and supervisors' ratings of job performance at Time 1. The correlation of .64 ($p < .001$) indicates that trainee teachers provided accurate reports of their supervisors' job performance ratings.

Analytical Procedure

Data were analyzed with structural equation modeling (SEM) techniques. We relied on SEM techniques because SEM techniques are generally considered to be the most adequate method to analyze cross-lagged panel designs (Finkel, 1995; Little, Preacher, Selig, & Card, 2007; Williams & Podsakoff, 1989). Specifically, SEM has advantages over other analytical approaches because SEM allows researchers (a) to correct for measurement error, (b) to test latent variables for measurement invariance

across time to ensure that latent variables share a common meaning at each time point, and (c) to avoid bias from spurious correlations between measurement errors by allowing item-specific errors to correlate over time (Finkel, 1995).

As recommended in the literature (cf. Finkel, 1995; Meade & Kroustalis, 2006; Vandenberg & Lance, 2000), we started SEM analyses by examining measurement invariance across time for each of the latent variables. We examined two types of measurement invariance: Configural (or form invariance) and metric invariance (or loading invariance). These types of measurement invariance are considered to be preconditions for adequately testing cross-lagged effects (Finkel, 1995; Hu & Cheung, 2008; Little et al., 2007; Steenkamp & Baumgartner, 1998). Configural invariance indicates that items can be assigned to factors as theoretically suggested. In order to test this type of invariance in longitudinal analyses, researchers examine the overall level of fit for measurement models containing measurements of one construct at more than one time point. Metric invariance indicates that the relation between the latent variable and the

items is constant over time. This assumption is tested by applying equality constraints on the factor loadings of the indicators and comparing the fit of these constraint models with the fit of the basic unconstrained models. When the difference between models on recommended fit indices is below the cutoff, this suggests that metric invariance exists.

After examining configural and metric invariance, we proceeded by examining cross-lagged panel models to test our hypotheses. We specified four bivariate models: In Model 1 we tested cross-lagged effects between surface acting and strain (Hypothesis 1), in Model 3 we tested cross-lagged effects between surface acting and job performance (Hypothesis 2), and in Model 4 we tested cross-lagged effects between deep acting and job performance (Hypothesis 3). Although we did not hypothesize a lagged effect from deep acting on strain, we specified a model testing cross-lagged effects between deep acting and strain (Model 2) for exploratory purposes.

Cross-lagged panel analyses using SEM provide a number of relevant indicators which are depicted in Figure 1: (1) stabilities, that is, correlations of con-

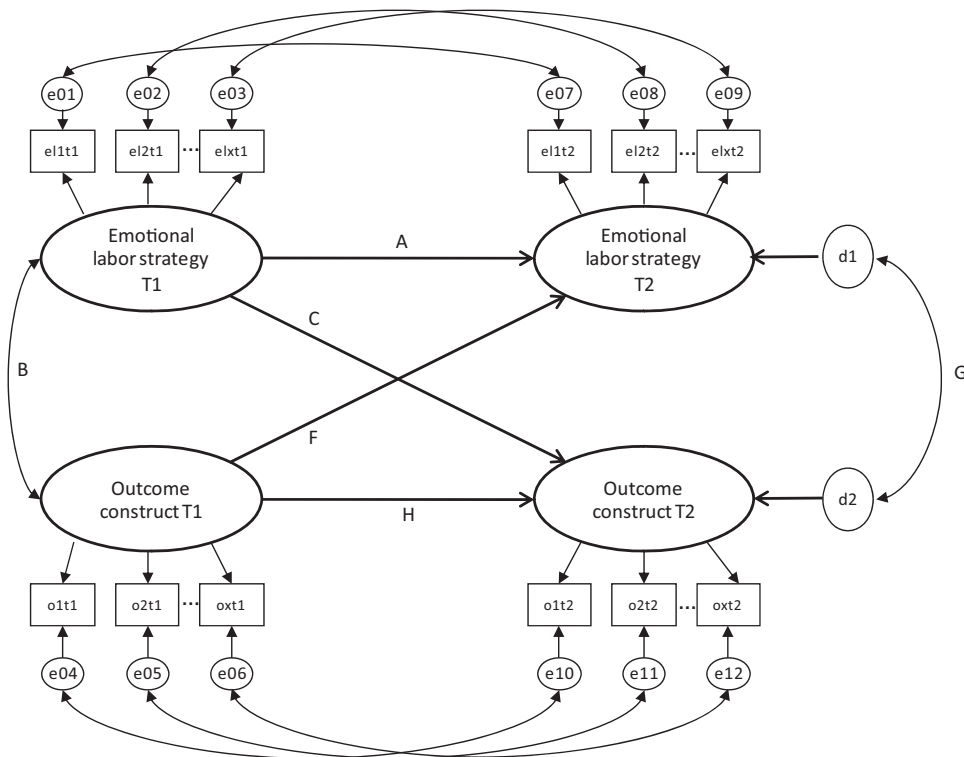


Figure 1. Cross-lagged structural equation model.

structs between Time 1 and Time 2 (depicted as A and H in Figure 1); (2) the cross-lagged effects of emotional labor strategies, that is, correlations between surface or deep acting at Time 1 and strain or performance at Time 2 (C); (3) cross-lagged effects of strain or performance, that is, correlations between strain or performance at Time 1 and surface or deep acting at Time 2 (F); (4) predictor correlations at Time 1 (B), that is, correlations between surface or deep acting and strain or performance measured at Time 1; and (5) disturbance correlations at Time 2, that is, correlations between the error terms of the two latent variables at Time 2 (G). The key information in cross-lagged panel SEM analyses comes from the cross-lagged effects, which provide estimates of the influence of Time 1 variables on Time 2 variables.

All SEM analyses were conducted using LISREL 8.72 (Jöreskog & Sörbom, 2005) and maximum likelihood estimation. For overall model evaluation, we relied on the comparative fit index (CFI) and the Tucker-Lewis index (TLI). CFI and TLI commonly have values ranging from 0 to 1. Although the use of strict cut-off values is generally not recommended in the literature (Marsh, Hau, & Wen, 2004; Vernon & Eysenck, 2007), we note that TLI and CFI values of .90 or higher are commonly considered to be indicative of acceptable fit, and values of .95 or higher are considered to be indicative of good fit (Kline, 1998). For the sake of completeness we also report values for the Root Mean Square of Error Approximation (RMSEA).

To evaluate the relative fit of models in metric invariance analyses, we relied on recommendations of Cheung and Rensvold (2001). Cheung and Rensvold recommended that researchers should rely on both the χ^2 -difference test and differences in CFI in

evaluating metric invariance and suggested that a nonsignificant χ^2 -difference and a CFI-difference of less than .01 provide evidence of measurement invariance.

Results

Descriptive Statistics and Dimensionality of Emotional Labor

Means, standard deviations, reliability coefficients, and bivariate correlations for all study variables are displayed in Table 1. Reliabilities were acceptable to good, overall. Lowest reliabilities were received for the two-item job performance measure (.62, Time 1; .65, Time 2), and are thus slightly higher than reliability estimates for supervisory ratings commonly reported in the literature (Viswesvaran, Ones, & Schmidt, 1996). Stabilities were lying between .63 ($p < .01$; surface acting) and .69 ($p < .01$; strain). Further, surface and deep acting displayed modest but significant positive intercorrelations, .36 ($p < .01$) at Time 1 and .27 ($p < .01$) at Time 2. We therefore investigated the dimensionality of the emotional labor measures before we proceeded and examined measurement invariance and structural models. Specifically, we compared a model treating deep acting and surface acting items as originating from a unified emotional labor construct and a model treating deep acting and surface acting as separate dimensions. Both models included measures at both Time 1 and Time 2 and allowed measurement errors for the same items to correlate over time. Allowing measurement errors to correlate over time is in line with recommendations regarding tests for longitudinal measurement models (Finkel, 1995; Little et al.,

Table 1
Descriptive Statistics, Reliabilities (Cronbach's Alpha) and Zero-Order Correlations for Study Variables

| Variable | <i>M</i> | <i>SD</i> | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|-------------------------|----------|-----------|-------|--------|-------|--------|-------|--------|-------|-------|
| 1. Deep acting (T1) | 2.62 | .80 | (.80) | | | | | | | |
| 2. Surface acting (T1) | 2.61 | .59 | .36** | (.71) | | | | | | |
| 3. Strain (T1) | 2.50 | .82 | .04 | .20* | (.87) | | | | | |
| 4. Job performance (T1) | 3.93 | .69 | -.16* | -.28** | .02 | (.62) | | | | |
| 5. Deep acting (T2) | 2.69 | .75 | .68** | .32** | .04 | -.09 | (.84) | | | |
| 6. Surface Acting (T2) | 2.63 | .60 | .22** | .63** | .15 | -.27** | .27** | (.72) | | |
| 7. Strain (T2) | 2.41 | .79 | .12 | .29** | .69** | -.05 | .12 | .23** | (.89) | |
| 8. Job performance (T2) | 4.04 | .75 | -.04 | -.20* | .00 | .68** | -.03 | -.33** | -.14 | (.65) |

Note. *N* = 151. T1 = Time 1; T2 = Time 2.

* $p < .05$. ** $p < .01$.

2007; Williams & Podsakoff, 1989) and is a reasonable assumption given that measurement errors from an identical measure over time should share common meaning (Finkel, 1995; Little et al., 2007; Zapf et al., 1996). The first set of these analyses included only the emotional labor items. Results are depicted in the first three rows of Table 2. To further test for potential conceptual overlap between emotional labor and outcome variables we repeated these analyses including the outcome variables (strain and performance). Results are depicted in rows 4, 5, and 6 of Table 2. Results of both sets of analyses confirm that surface acting and deep acting were distinct from each other and were also distinct from the outcome measures.

As can be seen from Table 2, treating surface and deep acting as separate constructs (2-factor model) provided a better fit than treating all emotional labor items as one integrated emotional labor construct (1-factor model).

Measurement Invariance

Table 2 provides tests of configural and metric invariance. As shown in Table 2, measurement models of surface acting and deep acting provided a good fit to the data, supporting configural invariance. Furthermore, constraining factor loadings to be equal over time resulted in changes in CFI that did not

Table 2
Dimensionality of Emotional Labor Measures, Measurement Invariance Analyses, and Measurement Model Fit

| Variable | χ^2 | <i>df</i> | <i>p</i> | CFI | TLI | RMSEA | RMSEA 90% CI |
|---|----------|-----------|----------|-------|-------|-------|-----------------|
| Dimensionality of emotional labor (EL) | | | | | | | |
| 2-factor model | 40.972 | 42 | .516 | 1.000 | 1.001 | .000 | (.000; .053) |
| 1-factor model | 173.298 | 47 | .000 | .900 | .856 | .134 | (.113; .155) |
| Model difference | 132.326 | 5 | .000 | .100 | .145 | -.134 | |
| Dimensionality of all study measures | | | | | | | |
| 2 EL-factors + strain + job performance | 329.065 | 258 | .000 | .981 | .976 | .043 | (.027; .056) |
| 1 EL-factor + strain + job performance | 538.123 | 271 | .000 | .927 | .913 | .081 | (.071; .091) |
| Model difference | 209.058 | 13 | .000 | .054 | .063 | -.038 | |
| Measurement invariance | | | | | | | |
| Surface acting (SA) | | | | | | | |
| Free loadings | 4.277 | 5 | .510 | 1.000 | 1.006 | .000 | (.000; .105) |
| Loadings invariant | 5.674 | 8 | .684 | 1.000 | 1.011 | .000 | (.000; .075) |
| Model difference | 1.397 | 3 | .706 | .000 | -.005 | .000 | |
| Deep acting (DA) | | | | | | | |
| Free loadings | 3.248 | 5 | .662 | 1.000 | 1.008 | .000 | (.000; .090) |
| Loadings invariant | 4.472 | 8 | .812 | 1.000 | 1.010 | .000 | (.000; .060) |
| Model difference | 1.224 | 3 | .747 | .000 | -.002 | .000 | |
| Strain | | | | | | | |
| Free loadings | 76.429 | 29 | .000 | .973 | .958 | .104 | (.076; .133) |
| Loadings invariant | 81.325 | 34 | .000 | .973 | .964 | .096 | (.070; .123) |
| Model difference | 4.896 | 5 | .429 | .000 | -.006 | .008 | |
| Job performance + invariant SA & DA | | | | | | | |
| Free loadings | 90.111 | 87 | .388 | .998 | .997 | .015 | (.000; .048) |
| Loadings invariant | 91.578 | 89 | .405 | .998 | .997 | .014 | (.000; .047) |
| Model difference | 1.467 | 2 | .480 | .000 | .000 | .001 | |
| Measurement models | | | | | | | |
| SA & Strain | 171.290 | 98 | .000 | .969 | .962 | .071 | (.053; .088) |
| DA & Strain | 155.658 | 98 | .000 | .977 | .971 | .063 | (.043; .081) |
| SA & job performance | 20.712 | 29 | .869 | 1.000 | 1.017 | .000 | (.000; .033) |
| DA & job performance | 36.656 | 29 | .155 | .991 | .987 | .042 | (.000; .079) |

Note. *N* = 151. EL = emotional labor; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root-mean square error of approximation; CI = confidence interval. Because job performance was operationalized using two items and a test of measurement invariance using only two indicators with correlated errors cannot be conducted, we included the restricted versions of the SA and DA models to identify the model.

exceed the cutoff value of $\Delta\text{CFI} \leq .01$ recommended by Cheung and Rensvold (2001) and changes in model χ^2 statistics that were statistically not significant, supporting metric invariance (see Table 2).

Structural Models and Tests of the Hypotheses

Table 3 provides results for the structural models. As indicated in Table 3, all bivariate and the multivariate models provided good fit to the data. We accordingly proceeded by examining the cross-lagged effects.

As shown in Table 3, surface acting had a significant positive lagged effect on strain (Model 1; stan-

dardized coefficient $\gamma = .19$; $p < .01$), supporting Hypothesis 1. In contrast, the opposite (and not hypothesized) reversed effect of strain on surface acting was not significant ($\gamma = .03$; ns). These results thus provide evidence that the relation between the two variables primarily stems from the influence of surface acting on strain and not vice versa. For exploratory reasons, we also examined potential (nonhypothesized) effects between deep acting and strain in Model 2. Both, the lagged effect of deep acting on strain ($\gamma = .10$; ns) and the opposite effect of strain on deep acting ($\gamma = .05$; ns) were not significant.

In Models 3 and 4 we tested cross-lagged effects between emotional labor strategies and job performance. Model 3 revealed that neither the cross-

Table 3
Structural Model Estimates

| Variable | Model 1 | Model 2 | Model 3 | Model 4 |
|--|--------------|--------------|--------------|--------------|
| Stabilities | | | | |
| Surface acting (SA) | .81 (.09)** | | .79 (.10)** | |
| Deep acting (DA) | | .81 (.07)** | | .81 (.07)** |
| Strain | .69 (.06)** | .75 (.06)** | | |
| Job performance (JP) | | | .95 (.13)** | .97 (.13)** |
| Cross-lagged effects of emotional labor | | | | |
| SA → Strain | .19 (.07)** | | | |
| DA → Strain | | .10 (.06) | | |
| SA → JP | | | .09 (.10) | |
| DA → JP | | | | .20 (.08)* |
| Cross-lagged effects of outcomes (reversed causation) | | | | |
| Strain → SA | .03 (.08) | | | |
| Strain → DA | | .05 (.04) | | |
| JP → SA | | | .05 (.09) | |
| JP → DA | | | | .02 (.07) |
| Predictor correlations T1 | | | | |
| SA ↔ Strain | .29 (.09)** | | | |
| DA ↔ Strain | | -.08 (.09) | | |
| SA ↔ JP | | | -.48 (.09)** | |
| DA ↔ JP | | | | -.21 (.10)* |
| Disturbance correlations T2 | | | | |
| SA ↔ Strain | .01 (.05) | | | |
| DA ↔ Strain | | .05 (.04) | | |
| SA ↔ JP | | | -.19 (.05)** | |
| DA ↔ JP | | | | -.03 (.05) |
| Model fit | | | | |
| χ^2 | 174.926 | 160.128 | 32.560 | 45.052 |
| p | .000 | .000 | .212 | .016 |
| df | 96 | 96 | 27 | 27 |
| CFI | .967 | .974 | .993 | .980 |
| TLI | .958 | .968 | .988 | .966 |
| RMSEA | .074 | .067 | .037 | .067 |
| RMSEA 90% CI | (.056; .091) | (.048; .085) | (.000; .077) | (.029; .100) |

Note. N = 151. T1 = Time 1; T2 = Time 2; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root-mean square error of approximation; CI = confidence interval. All values are standardized coefficients with standard errors in parentheses.

* $p < .05$. ** $p < .01$.

lagged effect of surface acting on job performance ($\gamma = .09$; *ns*) nor the reversed effect of job performance on strain ($\gamma = .05$; *ns*) were significant. Results did thus not lend support for Hypothesis 2. Model 4 displayed a significant positive cross-lagged effect of deep acting on job performance ($\gamma = .20$; $p < .05$), supporting Hypothesis 3. Again, the nonhypothesized, opposite lagged effect of job performance on deep acting was not significant ($\gamma = .02$; *ns*).

Discussion

Research into the consequences of emotional labor has blossomed in recent years and has provided ample evidence that surface acting is positively associated with strain and negatively associated with performance while deep acting seems to be positively related to performance (Bono & Vey, 2005; Hülsheger & Schewe, 2010). Yet, since extant studies were based on data collected at a single point in time they do not lend any insights into the direction of effects. Although there are a multitude of theoretical arguments suggesting that emotional labor causally precedes strain and performance, this is a theoretical assumption that has, to date, not been tested empirically. Goal of the present study was therefore to investigate how surface and deep acting causally relate to strain and performance. To this end we investigated reciprocal relationships between the two emotional labor constructs and strain and job performance in a full cross-lagged panel study.

Results supported the proposition that surface acting temporally precedes strain (Hypothesis 1), while deep acting precedes job performance (Hypothesis 3). Thus, individuals engaging in surface acting experience increases of strain over time, while individuals engaging in deep acting experience increases in job performance over time. With standardized coefficients of .19 (effect of surface acting on strain) and .20 (effect of deep acting on job performance), the effects are to be evaluated as small to medium in size (applying Cohen's, 1992 bench marks for correlation coefficients). This is in line with Frese and colleagues' (Frese & Zapf, 1988; Zapf et al., 1996) statement that due to theoretical and methodological reasons researchers should expect small effect sizes for stressor-strain relationships. In contrast, Hypothesis 2, stating that surface acting is negatively related to subsequent job performance, was not supported.

Our finding that surface acting precedes strain is in line with traditional stressor-strain approaches to occupational stress presuming a unidirectional causal

link between work-related characteristics or events and employees' psychological responses to it (Hart & Cooper, 2001). Overall, our findings thus lend empirical support to the conceptual model of emotional labor proposed by Grandey (2000) and refined by Rubin et al. (2005) contending that deep and surface acting have long-term consequences on individual and organizational well-being.

Furthermore, findings provide evidence for the differential effects of deep and surface acting. Consequences of emotional labor depend on the regulation strategies individuals adopt when facing emotional job requirements: While surface acting negatively affects individual well-being, deep acting positively affects performance and thereby organizational well-being. Further, our results add longitudinal evidence to findings from previous cross-sectional and diary studies showing that surface acting is related to impaired well-being (e.g., Brotheridge & Lee, 2002; Brotheridge & Lee, 2003; Glomb & Tews, 2004) while deep acting is not (e.g., Brotheridge & Lee, 2002; Martínez-Iñigo et al., 2007; Totterdell & Holman, 2003). This is in line with the assumption that opponent processes are underlying deep acting and that their differential effects on strain level each other out (Martinez-Iñigo et al., 2007): On one hand, deep acting is an effortful regulation strategy that depletes mental resources (Baumeister et al., 1998). On the other hand deep acting serves a social interactive role and leads to rewarding relationships with clients (Brotheridge & Lee, 2002; Côté, 2005) resulting in a resource gain (Martinez-Iñigo et al., 2007). These findings stand counter to Hochschild's (1983) early assertion that emotional labor in general harms the employee while it helps the organization. It seems like deep acting is the strategy that kills two birds with one stone in that it has the potential to benefit the organization without harming the employee. That of course provides grounds for optimism, not only for all of those who have to perform emotional labor but also for organizational decision makers. Apart from being interested in maintaining a healthy workforce, organizations are interested in achieving high levels of performance.

Our study results confirmed that trainee teachers engaging in deep acting received higher job performance ratings two months later. This longitudinal result is in line with previous cross-sectional research documenting a positive association between deep acting and different performance aspects: Deep acting was positively associated with emotional performance in a student lab study (Bono & Vey, 2007), with affective delivery in administrative assistants

(Grandey, 2003), and with self-rated performance in call-center employees (Totterdell & Holman, 2003). These studies focused on emotional performance aspects: They either explicitly assessed emotional performance and affective delivery (Bono & Vey, 2007; Grandey, 2003) or they assessed performance in samples from the service sector (call-center agents; Totterdell & Holman, 2003) in which dealing with the customer and hence emotional delivery is the primary aspect of performance. The present study adds to these findings by focusing on overall job performance ratings in a sample drawn from the educational rather than the service sector. It is noteworthy that deep acting is associated with increases in overall job performance ratings of trainee teachers, because these ratings do not focus primarily on emotional performance aspects. They rather reflect the overall educational proficiency and success a teacher displays in the classroom. This exemplifies that the impact of emotional labor is far wider than we might previously have expected. Following the present findings, emotional labor does not only influence service performance but it also facilitates general job performance and in consequence organizational outcomes. It is therefore an issue not only for organizations active in the service sector, but basically for all organizations whose employees need to interact with others, be it customers, patients, or students.

As stated in Hypothesis 2, we expected surface acting to have a negative lagged effect on job performance. Yet, we found no significant cross-lagged correlation between surface acting and subsequent job performance. We expected surface acting to have a negative effect on job performance following Grandey's lines of reasoning (2003; 2005). Accordingly, surface acting yields emotional displays that are inauthentic. The inauthenticity is thought to be detected by interaction partners (Ekman, 2003; Ekman et al., 1988), who are expected to react negatively to the falseness of surface acting which leads to negative ratings of role performance. This hypothesis could only partly be confirmed in previous research (Grandey, 2003). Surface acting was negatively related to affective delivery, "the extent that service delivery is perceived as friendly and warm" (p. 89; Grandey, 2003). Yet, it was not significantly related to breaking character, "the extent to which employees reveal negative moods or reactions to customers" (p. 89; Grandey, 2003). In light of these inconsistent findings we conclude that the relationship between surface acting and job performance might be more complex than previously expected. Future research

might therefore focus on potential moderator variables that influence the surface acting—job performance relationship.

In sum, the overall picture emerging from the present findings suggests the following: In dealing with emotional job requirements, deep acting is preferable over surface acting. It does not impair well-being while it enhances job performance. Evidently, the question emerges whether and how it is possible to encourage and support employees to engage in deep rather than in surface acting. A first indication that this might indeed be possible comes from Totterdell and Parkinson (1999). They analyzed the use of engagement versus diversion mood-regulation strategies among trainee teachers. Reappraisal, a form of deep acting, was one of the investigated engagement strategies. Results showed that a mere instruction to use more engagement strategies significantly increased the use of cognitive reappraisal and had a positive effect on cheerfulness. Richard (2006) investigated the effectiveness of training customer service employees in reappraisal strategies. Results were mixed. While the training led to increases in positive emotions and decreases in depersonalization after the training, no effects were found on a host of other outcome variables, such as negative emotions, job satisfaction, emotional exhaustion, and personal accomplishment. More research into the possibilities of training employees in deep acting is certainly needed. Researchers should also consider the possibility that trained deep acting has other effects on well-being and performance than spontaneous deep acting (Totterdell & Holmann, 2003). Increasing reappraisal and perspective taking in response to training might require additional effort and deplete rather than conserve mental resources.

Directions for Future Research

Various directions for future research can be deduced from the current findings. Here, we want to outline four key issues that could be fruitful targets of future research endeavors. First, given the specific sample used in the present study, future studies could explore whether the present longitudinal findings can be replicated in different occupational samples and settings. It would be especially interesting to investigate whether the positive relationship between deep acting and changes in job performance can be found in more occupations that don't pertain to the service sector and where job performance is not primarily an

indicator of affective delivery or emotional performance. Second, in light of our findings on the relationship between deep acting and job performance the question emerges as to why this relationship exists. Supervisors usually are the primary source of job performance ratings. Do they differentiate between authentic and inauthentic emotional displays and in consequence directly reward more authentic displays by providing higher overall job performance ratings? Or is the relationship mediated by other mechanisms? Taking the example of our trainee teachers, it could well be the case that deep acting had an effect on the learning climate in the classroom, on the students' behavior, motivation, and ability to learn and thereby affected teaching performance which in turn led to higher job performance ratings. Future research could direct attention to those potential mediating mechanisms. Somewhat related, emotional intelligence might be a variable worth considering in this context. Possibly, emotional intelligence moderates the emotional labor-performance relationship, such that individuals who are able to perceive which emotional display is appropriate in a given situation (high emotional intelligence) and subsequently engage in deep acting display the highest performance. Third, a thorough understanding of the mechanisms that underlie the emotional labor-strain relationship is indispensable. Throughout the introduction we have compiled a number of mechanisms that have been suggested and—in some cases—already been investigated in the literature. We believe that now the field would benefit from combining these various proposed mechanisms in one study. This would allow accounting for overlaps, interaction effects, and opponent processes in those mechanisms. Fourth and finally, we would like to point to the necessity to draw more attention to the antecedents of surface acting and deep acting. In light of the significant consequences of these emotional labor strategies, it is important to learn more about what makes individuals engage in surface or deep acting in work situations. Are they rather stable personality traits or coping styles or is it dependent on the situation and the specific emotional requirements whether individuals deep act or surface act or is it an interaction between person and situation? Answering these questions is necessary to decide to what extent interventions aiming at alleviating the burden of emotional labor should aim at the person and/or the situation and would help us develop such intervention programs.

Limitations

The present study is certainly not without its limitations. Our findings are based on a sample of trainee teachers and their generalizability might therefore be constrained to this specific kind of occupation. Another limitation concerns our assessment of job performance. We were not able to receive actual supervisor ratings for all trainee teachers at all measurement points and therefore had to rely on trainee teachers' reports of their performance evaluation results. However, by asking them about the grades they received in those formal evaluation sessions, our measure comes close to an actual supervisor rating. This is evidenced by the substantial correlation ($r = .64$) between trainee teachers' reports and the actual supervisors' job performance ratings at Time 1 for a subsample of 89 participants. Furthermore, we relied on self-report data on strain. Future studies might consider objective measures in addition to self-report measures of strain, for example, psychosomatic complaints or registered sickness absence.

Since we used a complete two-wave panel design and rigorous data analytical techniques, our data provide some insights into the causal order of relationships (Finkel, 1995; Zapf et al., 1996). Yet, to definitely establish cause and effect, it is necessary to verify this initial evidence in an experimental set-up. This would involve the experimental manipulation of surface and deep acting and the control of other confounding variables in the laboratory.

Conclusion

Longitudinal studies are thought to overcome some of the problems associated with cross-sectional studies (Zapf et al., 1996). It was therefore timely to investigate the direction of effects between emotional labor and individual and organizational well-being. By adhering to a rigorous longitudinal panel design and thorough longitudinal analyses, results revealed that surface acting does indeed lead to increased strain, while deep acting leads to enhanced job performance. Results bring us one step further to draw inferences about causal relationships underlying the emotional labor process.

References

- Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited

- resource? *Journal of Personality and Social Psychology*, 74, 1252–1265.
- Beehr, T. A., & Newman, J. E. (1978). Job stress, employee health, and organizational effectiveness: A facet analysis, model, and literature review. *Personnel Psychology*, 31, 665–698.
- Bono, J. E., & Vey, M. A. (2005). Toward understanding emotional management at work: A quantitative review of emotional labor research. In C. E. Hartel, W. J. Zerbe, & N. M. Ashkanasy (Eds.), *Emotions in organizational behavior* (pp. 213–233). Mahwah, NJ: Erlbaum Publishers.
- Bono, J. E., & Vey, M. A. (2007). Personality and emotional performance: Extraversion, neuroticism, and self-monitoring. *Journal of Occupational Health Psychology*, 12, 177–192.
- Brotheridge, C. M., & Grandey, A. A. (2002). Emotional labor and burnout: Comparing two perspectives of ‘people work’. *Journal of Vocational Behavior*, 60, 17–39.
- Brotheridge, C. M., & Lee, R. T. (2002). Testing a conservation of resources model of the dynamics of emotional labor. *Journal of Occupational Health Psychology*, 7, 57–67.
- Brotheridge, C. M., & Lee, R. T. (2003). Development and validation of the Emotional Labour Scale. *Journal of Occupational and Organizational Psychology*, 76, 365–379.
- Cheung, G. W., & Rensvold, R. B. (2001). The effects of model parsimony and sampling error on the fit of structural equation models. *Organizational Research Methods*, 4, 236–264.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159.
- Côté, S. (2005). A social interaction model of the effects of emotion regulation on work strain. *Academy of Management Review*, 30, 509–530.
- de Jonge, J., Dormann, C., Janssen, P. P. M., Dollard, M. F., Landeweerd, J. A., & Nijhuis, F. J. N. (2001). Testing reciprocal relationships between job characteristics and psychological well-being: A cross-lagged structural equation model. *Journal of Occupational and Organizational Psychology*, 74, 29–46.
- Diefendorff, J. M., Croyle, M. H., & Gosserand, R. H. (2005). The dimensionality and antecedents of emotional labor strategies. *Journal of Vocational Behavior*, 66, 339–357.
- Diefendorff, J. M., & Gosserand, R. H. (2003). Understanding the emotional labor process: A control theory perspective. *Journal of Organizational Behavior*, 24, 945–959.
- Diefendorff, J. M., & Richard, E. M. (2003). Antecedents and consequences of emotional display rule perceptions. *Journal of Applied Psychology*, 88, 284–294.
- Dormann, C., & Zapf, D. (2002). Social stressors at work, irritation, and depressive symptoms. Accounting for unmeasured third variables in a multi-wave study. *Journal of Occupational and Organizational Psychology*, 75, 33–58.
- Duke, A. B., Goodman, J. M., Treadway, D. C., & Breland, J. W. (2009). Perceived organizational support as a moderator of emotional labor/outcomes relationships. *Journal of Applied Social Psychology*, 39, 1013–1034.
- Ekman, P. (2003). *Emotions revealed*. New York: Henry Holt.
- Ekman, P., Friesen, W. V., & O’Sullivan, M. (1988). Smiles when lying. *Journal of Personality and Social Psychology*, 54, 414–420.
- Erickson, R. J., & Ritter, C. (2001). Emotional labor, burnout, and inauthenticity: Does gender matter? *Social Psychology Quarterly*, 64, 146–163.
- Finkel, S. E. (1995). *Causal analysis with panel data*. Thousand Oaks, CA: Sage.
- Frank, M. G., Ekman, P., & Friesen, W. V. (1993). Behavioral markers and recognizability of the smile of enjoyment. *Journal of Personality and Social Psychology*, 64, 83–93.
- Frese, M. (1999). Social support as a moderator of the relationship between work stressors and psychological dysfunctioning: A longitudinal study with objective measures. *Journal of Occupational Health Psychology*, 4, 179–192.
- Frese, M., & Zapf, D. (1988). Methodological issues in the study of work stress: Objective vs subjective measurement of work stress and the question of longitudinal studies. In C. L. Cooper & R. Payne (Eds.), *Causes, coping and consequences of stress at work* (pp. 375–411). Oxford: Wiley.
- Garst, H., Frese, M., & Molenaar, P. C. M. (2000). The temporal factor of change in stressor-strain relationships: A growth curve model on a longitudinal study in East Germany. *Journal of Applied Psychology*, 85, 417–438.
- Glomb, T. M., & Tews, M. J. (2004). Emotional labor: A conceptualization and scale development. *Journal of Vocational Behavior*, 64, 1–23.
- Gosserand, R. H., & Diefendorff, J. M. (2005). Emotional display rules and emotional labor: The moderating role of commitment. *Journal of Applied Psychology*, 90, 1256–1264.
- Grandey, A. A. (2000). Emotional regulation in the workplace: A new way to conceptualize emotional labor. *Journal of Occupational Health Psychology*, 5, 95–110.
- Grandey, A. A. (2003). When “the show must go on”: Surface acting and deep acting as determinants of emotional exhaustion and peer-rated service delivery. *Academy of Management Journal*, 46, 86–96.
- Grandey, A. A., Fisk, G. M., Mattila, A. S., Jansen, K. J., & Sideman, L. A. (2005). Is “service with a smile” enough? Authenticity of positive displays during service encounters. *Organizational Behavior and Human Decision Processes*, 96, 38–55.
- Gross, J. J. (1998). Antecedent- and response-focused emotion regulation: Divergent consequences for experience, expression, and physiology. *Journal of Personality and Social Psychology*, 74, 224–237.
- Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, 85, 348–362.
- Groth, M., Hennig-Thurau, T., & Walsh, G. (2009). Customer reactions to emotional labor: The roles of employee acting strategies and customer detection accuracy. *Academy of Management Journal*, 52, 958–974.
- Hart, P. M., & Cooper, C. L. (2001). Occupational stress: Toward a more integrated framework. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial and organizational psychology: Organizational psychology* (pp. 93–114). London, U.K.: Sage.

- Hennig-Thurau, T., Groth, M., Paul, M., & Gremler, D. D. (2006). Are all smiles created equal? How emotional contagion and emotional labor affect service relationships. *Journal of Marketing*, *70*, 58–73.
- Hochschild, A. R. (1983). *The managed heart*. Berkeley, CA: University of California Press.
- Holman, D., Martínez-Iñigo, D., & Totterdell, P. (2008a). Emotional labour and employee well-being: An integrative review. In *Research companion to emotion in organizations* (pp. 301–315). Cheltenham, U.K.: Edward Elgar.
- Holman, D., Martínez-Iñigo, D., & Totterdell, P. (2008b). Emotional labour, well-being and performance. In C. L. Cooper and S. Cartwright (Eds.), *The Oxford handbook of organizational well-being* (pp. 331–355). Oxford, U.K.: Oxford University Press.
- Hu, C., & Cheung, G. W. (2008, August). *Eight general questions on measurement equivalence/invariance that bother me most*. Paper presented at the annual meeting of the Academy of Management, Anaheim, CA.
- Hülsheger, U. R., & Schewe, A. F. (2010). *On the costs and benefits of emotional labor: A meta-analysis of three decades of research*. Manuscript submitted for publication.
- Jöreskog, K. G., & Sörbom D. (2005). LISREL (Version 8.72) [Computer software]. Chicago, IL: Scientific Software International.
- Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The job satisfaction-job performance relationship: A qualitative and quantitative review. *Psychological Bulletin*, *127*, 376–407.
- Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition and Emotion*, *13*, 505–521.
- Kline, R. B. (1998). *Principles and practices of structural equation modeling*. New York: Guilford Press.
- Lawler, E. E., & Porter, L. W. (1967). The effect of performance on job satisfaction. *Industrial Relations*, *7*, 20–28.
- Little, T. D., Preacher, K. J., Selig, J. P., & Card, N. A. (2007). New developments in latent panel analyses of longitudinal data. *International Journal of Behavioral Development*, *31*, 357–365.
- 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 overgeneralising Hu & Bentler's (1999) findings. *Structural Equation Modelling*, *11*, 320–341.
- Martínez-Iñigo, D., Totterdell, P., Alcover, C., & Holman, D. (2007). Emotional labour and emotional exhaustion: Interpersonal and intrapersonal mechanisms. *Work and Stress*, *21*, 30–47.
- Meade, A. W., & Kroustalis, C. M. (2006). Problems with item parceling for confirmatory factor analytic tests of measurement invariance. *Organizational Research Methods*, *9*, 369–403.
- Mohr, G., Müller, A., Rigotti, T., Aycan, Z., & Tschan, F. (2006). The assessment of psychological strain in work contexts: Concerning the structural equivalency of nine language adaptations of the irritation scale. *European Journal of Psychological Assessment*, *22*, 198–206.
- Muraven, M., Tice, D. M., & Baumeister, R. F. (1998). Self-control as a limited resource: Regulatory depletion patterns. *Journal of Personality and Social Psychology*, *74*, 774–789.
- Prati, L. M. (2004). *Emotional intelligence as a facilitator of the emotional labor process* (Unpublished doctoral dissertation). The Florida State University, Tallahassee.
- Rafaeli, A., & Sutton, R. I. (1987). Expression of emotion as part of the work role. *Academy of Management Review*, *12*, 23–37.
- Richard, E. M. (2003). *Applying appraisal theories of emotion to the concept of emotional labor* (Unpublished doctoral dissertation). Louisiana State University, Baton Rouge.
- Richards, J. M., & Gross, J. J. (1999). Composure at any cost? The cognitive consequences of emotion suppression. *Personality and Social Psychology Bulletin*, *25*, 1033–1044.
- Richards, J. M., & Gross, J. J. (2000). Emotion regulation and memory: The cognitive costs of keeping one's cool. *Journal of Personality and Social Psychology*, *79*, 410–424.
- Rubin, R. S., Staebler Tardino, V. M., Daus, C. S., & Munz, D. C. (2005). A reconceptualization of the emotional labor construct: On the development of an integrated theory of perceived emotional dissonance and emotional labor. In C. E. J. Härtel, W. J. Zerbe, & N. M. Ashkanasy (Eds.), *Emotions in organizational behavior* (pp. 189–211). Mahwah, NJ: Erlbaum.
- Schaubroeck, J., & Jones, J. R. (2000). Antecedents of workplace emotional labor dimensions and moderators of their effects on physical symptoms. *Journal of Organizational Behavior*, *21*, 163–183.
- Schyns, B. (2006). Are group consensus in leader-member exchange (LMX) and shared work values related to organizational outcomes? *Small Group Research*, *37*, 20–35.
- Sheldon, K. M., Ryan, R. M., Rawsthorne, L. J., & Iardi, B. (1997). Trait self and true self: Cross-role variation in the Big-Five personality traits and its relations with psychological authenticity and subjective well-being. *Journal of Personality and Social Psychology*, *73*, 1380–1393.
- Sideman-Goldberg, L. S., & Grandey, A. A. (2007). Display rules versus display autonomy: Emotion regulation, emotional exhaustion, and task performance in a call center simulation. *Journal of Occupational Health Psychology*, *12*, 301–318.
- Spector, P. E. (2001). Research methods in Industrial and Organizational Psychology: Data collection and data analysis with special consideration to international issues. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work & organizational psychology: Personnel psychology* (Vol. 1; pp. 10–26). London, UK: Sage.
- Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*, 78–90.
- Sutton, R. E., & Wheatley, K. F. (2003). Teachers' emotions and teaching: A review of the literature and directions for future research. *Educational Psychology Review*, *15*, 327–358.
- Totterdell, P., & Holman, D. (2003). Emotion regulation in customer service roles: Testing a model of emotional labor. *Journal of Occupational Health Psychology*, *8*, 55–73.
- Totterdell, P., & Parkinson, B. (1999). Use and effective-

- ness of self-regulation strategies for improving mood in a group of trainee teachers. *Journal of Occupational Health Psychology, 4*, 219–232.
- Tsai, W. C. (2001). Determinants and consequences of employee displayed positive emotions. *Journal of Management, 27*, 497–512.
- Tsai, W. C., & Huan, Y. M. (2002). Mechanisms linking employee affective delivery and customer behavioral intentions. *Journal of Applied Psychology, 87*, 1001–1008.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4–69.
- Van Kleef, G. A. (2009). How emotions regulate social life: The emotions as social information (EASI) model. *Current Directions in Psychological Science, 18*, 184–188.
- Van Kleef, G. A. (2010). The emerging view of emotion as social information. *Social and Personality Psychology Compass, 4*, 331–343.
- Vernon, T., & Eysenck, S. (Eds.). (2007). Special issue on structural equation modeling. *Personality and Individual Differences, 42*.
- Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology, 81*, 557–574.
- Williams, L. J., & Podsakoff, P. M. (1989). Longitudinal field methods for studying reciprocal relationships in organizational behavior research: Toward improved causal analysis. In B. M. Staw, & L. L. Cummings (Eds.), *Research in organizational behavior* (vol. 11, pp. 247–292). London: JAI.
- Zapf, D. (2002). Emotion work and psychological well-being: A review of the literature and some conceptual considerations. *Human Resource Management Review, 12*, 237–268.
- Zapf, D., Dormann, C., & Frese, M. (1996). Longitudinal studies in organizational stress research: A review of the literature with reference to methodological issues. *Journal of Occupational Health Psychology, 1*, 145–169.

Received December 12, 2008

Revision received July 24, 2010

Accepted July 24, 2010