3-2014

Differences in Emotional Labor Across Cultures: A Comparison of Chinese and U.S. Service Workers

Joseph A. Allen
University of Nebraska at Omaha, josephallen@unomaha.edu

James M. Diefendorff
University of Akron Main Campus

Yufeng Ma
Sinopec Management Institute

Differences in Emotional Labor Across Cultures: A Comparison of Chinese and U.S. Service Workers

Joseph A. Allen
Department of Psychology, Creighton University

James M. Diefendorff
Department of Psychology, University of Akron

Yufeng Ma
Sinopec Management Institute, Beijing, China

Abstract

Purpose In the global economy, the need for understanding cross-cultural differences and the customer service-related processes involved in emotional labor is evident. The current study attempts to examine this issue by developing and testing hypotheses pertaining to cross-cultural differences between U.S. and Chinese service workers on the levels of display rule perceptions, emotion regulation, and burnout (i.e., emotional exhaustion, personal accomplishment, and depersonalization) as well as the relationships among these variables.

Design/Methodology/Approach Data was collected from service workers in the U.S. (n=280) and China (n=231). We tested for measurement differences, mean differences, and differences in the relationships among emotional labor variables between the two samples using a variety of analyses.

Findings It was found that the relatively robust sequence of display rules to surface acting to burnout was observed in a U.S. sample but was not observed in a Chinese sample, with some relationships being significantly weaker in the Chinese sample (e.g., surface acting to burnout dimensions) and others exhibiting relationships with the opposite sign (e.g., display rules were negatively related to surface acting in the Chinese sample).

Implications The results of this study suggest that many of the relationships among emotional labor variables vary as a function of the cultural context under consideration.

Originality/Value This is the first study to directly compare emotional labor across samples from Eastern and Western cultures. Additionally, this study begins to answer questions concerning why models of emotional labor generated in a Western culture may not apply in other cultures.

Keywords

Emotional labor, Cultural differences, Emotional exhaustion, Affectivity

Differences in Emotional Labor Across Cultures: A Comparison of Chinese and U.S. Service Workers

Emotional labor, as postulated by Hochschild (1983) andmodified by others (e.g., Grandey 2000), suggests that employees engage in emotion regulation in response to organizational demands for certain types of emotional displays (e.g., display rules) and that emotion regulation can lead to negative outcomes

(e.g., emotional exhaustion) depending upon the regulation strategy used (e.g., surface acting and deep acting). In recent years, a large body of research has investigated emotional labor in its various forms and looked at how contextual factors impact the way individuals react to emotional demands placed upon them (Allen et al. 2010; Brotheridge and Grandey 2002; Brotheridge and Lee 2003; Grandey 2003; Grandey et al. 2005; Rupp and Spencer 2006). Although these studies provide substantial evidence supporting the basic tenets of the traditional model of emotional labor (Hochschild 1983), the research relies almost exclusively on samples from North American, with few exceptions (see Grandey et al. 2005).

In the global economy, the need for understanding cultural differences cannot be overstated and the need for cross-cultural research concerning emotional labor is evident.According to Grandey et al. (2005, p. 902), ‘‘cultural differences in work emotions are important to understand...as companies outsource their service functions overseas and export companies to other countries, managers need to be aware of the potential strain on employees if they require emotional displays that are incongruent with cultural norms.’’ Thus, from the perspectives of both researchers and practitioners, understanding how cultural differences impact emotional labor processes may help to better ensure that employees can provide quality service that matches the expectations of customers in different countries.

Given that cross-cultural emotional labor research is quite sparse, the current study attempts to address this gap by examining emotional labor processes in two samples of service workers from different cultural contexts. Specifically, we look at cross-cultural differences between U.S. and Chinese service workers in display rule perceptions, emotion regulation, and burnout (i.e., emotional exhaustion, personal accomplishment, and depersonalization) as well as differences in the relationships between these variables. In doing this, we extend emotional labor research by testing whether the predominant emotional labor model in the literature operates differently in the level of key variables or the relationships among key variables across two cultural contexts. Evidence for differences could provide the impetus for more formally incorporating cultural factors in multinational applications of emotional labor in service jobs. We apply existing theory on emotional labor developed in Western contexts to guide our overarching model and then consider ways in which the relationships among variables in the model might differ between the U.S. and Chinese samples. In doing so, we theorize about possible cultural mechanisms (i.e., collectivism/individualism) that might result in differences between the two samples.

A Model of Emotional Labor

The term emotional labor refers to the process of managing emotions for pay (Hochschild 1983). Engaging in emotional labor may involve faking, suppressing, enhancing, or alter- ing emotions in an effort to provide a particular emotional display as prescribed by organizational goals (Grandey 2000). Grandey (2000) proposed a model of emotional labor suggesting that service workers engage in emotional labor in response to display rules—standards indicating which emotional expressions are appropriate in a given situation (Diefendorff and Richard 2003; Rafaeli and Sutton 1987). Organizations develop display rules as a means of governing the emotional tone of the customer-service interface with these rules typically including expectations to express positive emotions and suppress negative emotions (Allen et al. 2010).

According to Grandey’s (2000) model, employees respond to display rules by engaging in two types of emotion regulation: surface acting and deep acting. Surface acting refers to modifying the outward expression of the emotion, through suppressing felt emotions, faking expressed emotions, or both. Deep acting involves consciously changing how one feels in order to express the desired emotion, through such activities as reappraisal, positive refocusing, or physiological modification (Grandey 2000; Gross 1998).

Although both strategies aid in meeting the organization's display rule requirements, surface acting involves mere compliance (i.e., acting in "bad faith;" Rafaeli and Sutton 1987) whereas deep acting provides authentic emotional displays (i.e., acting in "good faith;" Rafaeli and Sutton 1987).

Meeting display requirements through emotion regulation is thought to deplete cognitive and emotional resources which may result in emotional exhaustion and other negative outcomes (Hochschild 1983; Grandey 2003). Although theoretical models of emotional labor have suggested that burnout is an outcome of both types of emotional labor (i.e., surface acting and deep acting; Grandey 2000), the empirical findings are mixed. A recent meta-analysis by Hülsheger and Shewe (2011) found some support for the mediated process model depicted in Fig.1. Specifically, surface acting was positively correlated with display rules and the three dimensions of burnout (see also Brotheridge and Grandey 2002; Grandey 2003), with the links to burnout presumably being a result of the effort required to show the desired emotions and the emotional dissonance resulting from showing fake emotions (Cote 2005). In contrast, deep acting was unrelated to display rules and certain well-being outcomes (i.e., emotional exhaustion and depersonalization), though it did significantly positively predict the burnout dimension of personal accomplishment (Brotheridge and Grandey 2002; Brotheridge and Lee 2003). The null effects of deep acting in predicting exhaustion and depersonalization have been theorized as resulting from simultaneous positive effects (due to positive emotions individuals experience as a result of deep acting) and negative effects (due to the effort involved in deep acting and the potential to be alienated from one's true feelings; Hochschild 1983) canceling each other out (Grandey et al. 2013). Thus, across studies there appears to be empirical support for surface acting as a mediator of the display rules and burnout relationship, but no such role for deep acting. However, the majority of the studies included in the meta-analysis used samples from Western cultures. Therefore, one purpose of this study was to examine and compare support for the model in two samples from different cultures.

Emotional Labor Across Cultures

Though researchers have studied emotion regulation and display rules across cultures (Eid and Diener 2001; Masuda et al. 2008; Safdar et al. 2008), relatively little research has looked at the relationships among emotional labor constructs from a cross-cultural perspective (Grandey et al. 2005). The research on display rules has typically focused on cultural display rules (i.e., how different cultures influence the display of emotion in social settings) rather than on how culture shapes the display rules present in organizations (see Grandey et al. 2010) or impacts the way that display rules relate to emotion regulation or well-being. Grandey et al. (2010) focused primarily on specific emotions (e.g., anger and happiness) and how they are expressed relative to display rules toward organizational targets and non-organizational targets across cultures. However, unlike the present study, they did not investigate emotional labor strategies (e.g., surface and deep acting) or the outcomes associated with engaging in emotional labor and how that differs across cultures. Similarly, though research has shown that emotion regulation varies as a function of cultural context (Butler et al. 2007; Matsumoto et al. 2008), very little work has focused on how culture specifically impacts surface acting and deep acting and the way that these emotion regulation strategies relate to different aspects of employee well-being.

In the one study looking at cross-cultural differences in emotional labor, Grandey et al. (2005) showed that the way that emotional labor is enacted by service employees differed between the U.S. and France. Specifically, they found that the relationship between emotion regulation and job dissatisfaction was weaker among French customer-contact employees than among U.S. employees. They proposed that this difference stems from an impulsive orientation toward emotion among French employees and an institutional orientation toward emotions among U.S. employees. That is, they argued that the impulsive orientation of the French culture allowed more autonomy for employees to express felt emotions, whereas

the institutional orientation of the U.S. culture makes employees feel more connected to their employment situation and, as a result, show emotions in line with institutional reasons (i.e., the customer-service role).

Although the cultural mechanism described by Grandey et al. (2005) (i.e., institutional versus impulsive orientations toward emotion) provides a potential explanation for their findings, the authors acknowledged that there are other important cultural variables that could explain country-level differences in emotional labor. For instance, a common distinction made between Eastern and Western cultures is that they differ in the level of collectivism versus individualism (Hofstede 1980; Uchida et al. 2008) and related constructs (e.g., emotional moderation/expression; Eid and Diener 2001). This widely agreed upon distinction between Eastern and Western societies provides a mechanism for theorizing about cultural influences on emotional labor processes in the U.S. and China.

Emotions and Individualism-Collectivism

One of the more commonly studied factors in cross-cultural research is the degree to which a given culture is collectivistic versus individualistic (Hofstede 1980). Individualistic cultures value the individual over the group and tend to promote uniqueness, autonomy, and separateness (Markus and Kitayama 1991). Countries high in individualism (e.g., U.S.) view emotions as an individual's right and as being an important personal experience (Safdar et al. 2008). Emotional expression is typically promoted in individualistic cultures as a means of personal expression, with norms emphasizing emotional expression. In contrast, collectivistic cultures value groups over individuals (Hofstede 1980). These cultures tend to promote harmony and cooperation within the group rather than promote assertive individual behavior (Noon and Lewis 1992). In these cultures, emotions are interactive experiences that reflect the social context and downplay connections to the inner self (Safdar et al. 2008). As a result, emotional expressions tend to be much more controlled, with greater focus on how emotions might impact the relationship between the individual and others (Mesquita 2000). Additionally, collectivistic cultures tend to promote emotional moderation whereas individualistic cultures tend to promote emotional expression (Eid and Diener 2001). According to Eid and Diener (2001, p. 883), "in China there is a general attitude to consider emotions as dangerous, irrelevant, or illness causing...the moderation or suppression of emotions is generally highly valued in China." Eid and Diener (2001) found that, in comparison to western cultures (e.g., U.S.), China has a strong norm for emotional moderation.

Emotional Labor and Individualism-Collectivism

Because of the emphasis on managing relationships and emotional moderation in collectivistic cultures, it is believed that in China it may be more common to regulate emotions across contexts than in the U.S. That is, emotion regulation may be a normative process in collectivistic cultures that stems from the need to maintain harmony and cooperation among groups (Mesquita 2000; Mesquita and Delvaux 2013). As a result of this normative expectation, emotion regulation in these cultures may be less dependent upon work-specific display rules (i.e., smile at customers) and more dependent on the societal norm to maintain harmony. In contrast, the need to regulate emotions in individualistic cultures may be more dependent upon the particular display rules of the situation. Research on emotions across cultures suggests that in individualistic cultures, natural emotional expression is valued (Eid and Diener 2001). Thus, individuals in these cultures are more comfortable expressing their natural emotions and may find display rules established by their work organizations to be more intrusive on their emotions. That is, because strong emotional control is less of a social norm in the U.S. compared to in China, the presence of work display rules will make this expectation more salient and distinctive for U.S. workers compared to Chinese workers, who may be more used to the presence of norms governing emotional expression. As a result of these cultural differences, it is expected that service workers in the U.S. will perceive stronger display

rules at work compared to Chinese service workers. Additionally, because of the broader social expectation to regulate emotions, service workers in China will tend to regulate their emotions more across a broader range of contexts, including work. As such, it is expected that Chinese service workers will engage in more emotion regulation on average than U.S. service workers. Thus, the following hypotheses are proposed:

Hypothesis 1a Display rules are perceived at higher levels in the U.S. than in China.

Hypothesis 1b The emotional regulation strategies of (a) surface acting and (b) deep acting are reported as being used more often in China than in the U.S.

Also given these cultural differences, we expected that emotional display rule perceptions would exhibit stronger relationships with emotion regulation in the U.S. than in China. In China, service workers are likely to engage in more emotion regulation across contexts because of broader social norms, and as such, may not do so as a function of organizational display rules (Oetzel et al. 2001; Safdar et al. 2008). In contrast, the perception of display rules at work may more strongly influence the level of emotion regulation performed by employees from the U.S. because such rules are distinct from the prevailing cultural norm of emotional freedom and discretion in emotion management. That is, ‘‘service with a smile’’ display rules are more distinct from broader cultural norms for social interaction in the U.S. than they are from the cultural norms in China, which should result in these workplace display rules having a more powerful effect on employee behavior in the U.S. than in China. Thus, the following hypothesis is presented:

Hypothesis 2 Display rules are positively related to both (a) surface acting and (b) deep acting, with this relationship being stronger in the U.S. compared to China.

Emotional Labor and Burnout Across Cultures

Previous research and theory suggest that emotional labor is related to facets of burnout (i.e., personal accomplishment, depersonalization, and emotional exhaustion; Brotheridge and Grandey 2002). Across the many studies testing these relationships, there appears to be some consistency in the relationships between surface acting and the facets of burnout in samples from individualistic cultures, with surface acting showing a negative relationship with personal accomplishment and positive relationships with depersonalization and emotional exhaustion (Hülsheger and Shewe 2011). However, we theorize that the relationships of surface acting with burnout dimensions in a collectivistic culture may be weaker than is observed in an individualistic culture. Given the overarching norm for emotion control in collectivistic cultures, the act of suppressing felt emotions and faking unfelt emotions in such a culture may be experienced as less effortful and result in less dissonance than doing so in an individualistic culture. This may be because individuals in collectivistic cultures in general have more experience with regulating their emotions as part of the cultural norms associated with promoting harmony and avoiding conflict (Hofstede 1980). Thus, the regulation of emotions is instilled through socialization processes as a means of supporting the group/collective which may make such regulation automatic and less resource intensive. Further, surface acting may generate a positive response in Chinese employees, due to the belief that surface acting may help one maintain harmony and behave in line with social norms. Indeed, Chinese service workers may experience the act of conforming to display rules through surface acting as being beneficial due to the alignment of one’s behaviors to broader social norms. As a result, surface acting in a collectivistic culture may be less likely to harm employee well-being than surface acting in an individualistic culture because of the collectivistic norm for emotional moderation and maintaining harmony/cooperation.

Deep acting also may function differently across the two cultures. In individualistic cultures, the positive effects (i.e., positive felt emotions) and negative effects (i.e., effort and emotional alienation) of deep acting may cancel each other out, resulting in a weak or null relationship with burnout. However, in collectivistic cultures where harmony and group supporting behaviors are highly desired and where emotion regulation in response to contextual demands is valued, deep acting may primarily exhibit positive effects. That is, the beneficial effects of deep acting will be present (i.e., employees will experience more positive emotions), but the harmful effects of deep acting may be absent or greatly diminished (i.e., perceived effort due to regulation) in a collectivistic culture. Further, individuals in collectivistic cultures may not experience the process of changing one's emotions to match external demands as being as threatening to one's sense of authenticity because of the high value placed on managing emotions for social reasons. As a result, we expect that deep acting will exhibit null or weak effects with the components of burnout in the U.S. but will demonstrate significant and beneficial effects in predicting burnout components in China. Thus, the following hypotheses are proposed:

Hypothesis 3a Surface acting is positively associated with burnout (i.e., less personal accomplishment, more depersonalization, and more emotional exhaustion), with this relationship being stronger in the U.S. compared to China.

Hypothesis 3b Deep acting is negatively associated with burnout (i.e., more personal accomplishment, less depersonalization, and less emotional exhaustion), with this relationship being stronger in China compared to the U.S.

As implied by Fig.1, display rules are theorized to operate through emotion regulation to impact well-being outcomes, suggesting a mediated relationship (Grandey 2000). However, research also supports the idea that display rules may have direct effects on well-being outcomes, possibly by serving to constrain or control employee behavior (Diefendorff et al. 2011). Diefendorff et al. (2011) found that surface acting partially mediated the relationship between display rules and burnout, suggesting that display rules have both direct and indirect links to employee well-being. We build on these findings by looking at all three facets of burnout to explore whether and in what way display rules directly relate to the dimensions of burnout in both countries.

Methods

Sample and Procedure

To test these hypotheses, we administered questionnaires assessing the emotional labor constructs in Fig. 1 to two samples of customer service employees, one in China and one in the U.S. It should be noted that U.S. and China were chosen specifically for their respective emphasis on individualism versus collectivism (Hofstede 1980), though other counties have been identified as having higher levels of collectivism than China (e.g., Taiwan) and higher levels of individualism than the U.S. (e.g., The Netherlands). As such, our results may represent a more conservative test of potential differences between individualistic and collectivistic cultures (Hofstede 1980).

As a first step in testing our hypotheses we examined the cross-cultural measurement equivalence of the scales assessing constructs in Fig. 1 and then proceeded to test our hypotheses using multiple-groups structural equation modeling (SEM). In preparing the surveys, we followed previous conventions for ensuring that the measures had similar meanings across the two countries (Grandey et al. 2005; Hofstede 1980). We followed four steps, similar to Grandey et al. (2005). First, the lead author designed the survey with scales in English. Second, the survey was translated from English to Mandarin Chinese by native Mandarin speakers. Third, the items were back-translated to English by a bilingual native Mandarin

Chinese speaker (Brislin 1970). Fourth, the items were reviewed by the authors to see if the back-translated items maintained their original meanings. Following a few minor adjustments, the survey was administered to participants in both countries.

For the U.S. sample, 50 students at a large university in the South Eastern U.S. were asked to identify 10 working adults who would be willing to participate in a survey. Students received extra credit for providing email addresses for these individuals. A link to an online survey was then sent to these individuals. Of the 500 potential respondents, 342 completed the survey for an initial response rate of 68.4 %. Due to the nature of the study, those who indicated they had minimal contact with customers were removed from the sample (n=53) as well as those who were missing more than 5 % of the data (n=9). The final sample consisted of 280 individuals where 55.6 % were female, average age was 34.1 years (SD=12.8), and the average tenure with their current job was 6.2 years (SD=7.2). Of the 280 employees, 112 worked in retail (e.g., malls and grocery stores), 93 worked in restaurants, and 75 worked in a variety of service-oriented jobs (e.g., telemarketing and direct sales associates).

For the sample from China, an online survey was sent via email to 390 potential service employee participants working in one of four organizations (n=122 shopping mall employees, n=96 hotel employees, n=82 restaurant employees, and n=90 call center employees) headquartered in Beijing. The four organizations who participated in this study were invited by one of the authors through solicitations to the top management of the respective firms. In total, 321 individuals participated in the survey for an initial response rate of 82 %. As with the other sample, we removed participants who had minimal contact with customers (n=57) and those who were missing more than 5 % of the data (n=33). The final sample consisted of 231 individuals where 44.6 % were female, average age was 30.1 years (SD=7.2), and the average tenure with their current job was 7.4 years (SD=7.3).

T-tests demonstrated that the two samples differed in age ($t=-4.24$, $p<.05$) and proportion female ($t=2.44$, $p<.05$), but not on tenure on the current job ($t=1.85$, $p>.05$). Further, age was significantly correlated with several of the substantive variables in this study. As such, we performed supplemental analyses in which we controlled for these demographic variables when testing our hypotheses.[1] Including these covariates in the model did not alter the pattern of significance when testing the substantive hypotheses so we did not include them in the multiple-groups SEM reported below for ease of presentation.

Measures

Each survey contained the following measures with the survey for Chinese service workers being translated into Mandarin. The internal consistency reliability estimates for each sample are presented in Table 1.

Integrative Display Rules

Integrative display rules were measured using seven items developed by Best et al. (1997). The participants were asked, ‘‘Please indicate your level of agreement with the following items concerning your job.’’ Ratings were made on a 5-point scale, ranging from 1 being ‘‘strongly disagree’’ to 5 being ‘‘strongly agree.’’ Sample items include ‘‘Part of my job is to make the customer feel good’’ and ‘‘I am expected to suppress my bad moods or negative reactions to customers.’’

Surface Acting and Deep Acting

Surface acting and deep acting were measured using five items and four items (respectively) from Grandey (2003) which are based upon items originally developed by Brotheridge and Lee (2003). Minor modifications were made following recommendations suggested by Diefendorff et al. (2005). Participants

were asked ‘‘Please indicate your level of agreement with the following statements concerning your job:’’ Ratings were made using a 5-point Likert-type scale, ranging from 1 being ‘‘Strongly Disagree’’ to 5 being ‘‘Strongly Agree.’’ A sample surface acting item is ‘‘Put on an act in order to deal with customers in an appropriate way,’’ and a sample deep acting item is ‘‘Try to actually experience the emotions I must show to customers.’’

Burnout

Burnout was assessed using the Maslach Burnout Inventory (MBI; Maslach and Jackson 1981) with personal accomplishment assessed using eight items, depersonalization assessed using five items, and emotional exhaustion assessed using nine items. Participants were asked, ‘‘How often do you:’’ followed by a series of statements. Sample items include ‘‘I feel I’m positively influencing other people’s lives through my work’’ for personal accomplishment, ‘‘I’ve become more callous toward people since I took this job’’ for depersonalization, and ‘‘I feel emotionally drained from work’’ for emotional exhaustion. Items were rated on a seven-point scale ranging from 1 being ‘‘never’’ to 7 being ‘‘every day.’’

Results

Table 1 presents the means, standard deviations, correlations, and internal consistency reliabilities for all of the study variables in the U.S. and Chinese samples. Before testing each hypothesis, confirmatory factor analysis was used to assess the measurement model and to provide an indication of measurement equivalence across the two samples. For each CFA model, individual indicators were allowed to load on only one factor and the latent variables were allowed to freely correlate. Several indicators of model fit were examined, including (a) the goodness of fit statistic, (b) the root mean square error of approximation (RMSEA), (c) the Tucker–Lewis index (TLI), (d) the comparative fit index (CFI), and (e) the standardized root mean square residual (SRMR). The lower bound of good fit for the TLI and the CFI is considered to be .90. For the RMSEA and the SRMR, the upper bounds for good fit are considered to be .08 and .10, respectively (Vandenberg and Lance 2000). Further, because the models were nested in most cases, their fit could be directly compared using the χ^2 difference test.

Tests of Measurement Equivalence

Prior to testing the substantive hypotheses, we wanted to know whether the scales exhibited conceptual equivalence in our two samples (Cheung and Rensvold 1999). Conceptual equivalence is supported if the responses of participants from the U.S. and China have the same factor structure (i.e., same number of underlying dimensions, same items loading on the same dimensions, and item loadings of roughly the same magnitude), meaning that the same underlying latent constructs are assessed in the same way for employees in the two countries (Cheung and Rensvold 1999). Assuming most of the requirements for conceptual equivalence are met (e.g., same underlying dimensions and same items loading on the same dimensions), then cross-cultural researchers conclude that partial equivalence has occurred and cross-cultural comparisons can still be made (Cheung et al. 2004; Yoo and Donthu 2011). That is, differences in the magnitude of factor loadings suggest that the contribution of indicators to a latent construct is different, but such a difference does not mean that the substantive meaning of the construct is not comparable across groups (Byrne et al. 1989; Milfont and Fischer 2010).

As a first step in testing the conceptual equivalence of the scales, we simultaneously examined whether the six-factor solution separating all of the constructs fit the data well for our samples using multiple-groups SEM. This procedure involved testing two nested models, starting with a model imposing the same factor structure on both groups. If this model provides poor fit to the data, it is interpreted to mean that participants from the two samples disagree over the number or composition of factors contained in

the instrument. If this model fits the data reasonably well, a second model is estimated with the added constraint of equal factor loadings across the two groups. At issue here is whether the strength of the relationships between specific behaviors (items) and the underlying constructs is the same for different groups. So for example, the two groups may agree that a regulation strategy (e.g., ''I work hard to feel the emotions that I need to show to customers'') is related to an underlying construct (e.g., deep acting), but the magnitude of the loading may differ for the two groups, suggesting that for one group this regulation strategy is more important for defining the construct than the other group. If both models achieve good fit to the data, then there is complete conceptual equivalence for the scale (Cheung and Rensvold 1998; Reise et al. 1993). Although other sources of nonequivalence may be revealed, they primarily relate to psychometric properties (e.g., mean differences, reliability differences, and range differences) rather than to the underlying conceptual foundation of the instrument.

An initial analysis of the 6-factor model for both countries (Model A) demonstrated good fit on three out of four fit indices [(430) = 1027.48, p<.05, RMSEA = .076, TLI = .95; CFI = .96 SRMR US = .050, SRMR China = .110], with less than desirable fit being present for the SRMR from the Chinese sample. Inspection of the item loadings revealed that two of the surface acting items pertaining to faking emotions (''I fake a good mood when interacting with customers,'' ''I fake the emotions I show when dealing with customers'') had factor loadings close to zero in China, though the factor loadings were acceptable in the U.S. This finding means that these two items were not functioning the same in the two groups, suggesting that the concept of faking emotions either does not easily translate into Chinese or that such a concept does not co-vary with other aspects of surface acting (i.e., suppressing) in our Chinese sample. Either way, these items were not operating the same in the two samples, so we dropped them and retested the 6-factor model (Model B). We revisit this issue in the discussion section.

Model B exhibited very good fit across all four fit indices [χ^2 (348) = 720.53, p< .05, RMSEA = .064, TLI = .96; CFI = .97, $SRMR_{US}$ = .051, $SRMR_{China}$ = .065] with all of the factor loadings being above .50 for both samples (see "Appendix"). After establishing good fit for the hypothesized model, we then proceeded to test two more parsimonious models that (a) combined the surface acting and deep acting items into a single emotion regulation construct (Model C) and (b) combined the three burnout dimensions into one overall burnout construct (Model D). Both of these models exhibited acceptable fit on two fit indices (TLI and CFI), but unacceptable fit on the SRMR and RMSEA (see Table 2). Further, the change in χ^2 between models B and C ($\Delta\ \chi^2$ (10) = 1067.24, $p < .05$) and models B and D ($\Delta\ \chi^2$ (18) = 1435.80, $p < .05$) was significant, suggesting that the fit was worse for these more parsimonious models. These results suggest that the responses for the two groups had the same number of factors and that the items loaded on the same dimensions. Thus, the latent variables are defined similarly across the two groups (see "Appendix").

Next we applied the constraint of equal factor loadings on the two groups. The overall fit of this model was acceptable [χ^2 (363) = 766.04, $p < .05$, RMSEA = .066, TLI = .96; CFI = .97; $SRMR_{US}$ = .053, $SRMR_{China}$ = .076], but this model fit worse than the unconstrained model [$\Delta\ \chi^2$ (15) = 45.51, $p < .05$], suggesting that the magnitude of the factor loadings was not the same across the groups. As a next step, we explored whether this decrease in model fit was primarily attributable to certain items. To do this, we freed up the constraint of equal factor loadings one item at a time and examined whether the $\Delta\ \chi^2$ was significant, suggesting that the fit of the model improved. Such a finding would mean that the particular item in question had significantly different factor loading for the two groups. The "Appendix" displays the factor loadings for each indicator across the two groups (as well as the constrained loading) and whether the $\Delta\ \chi^2$ was significant for each loading. In total, 14 of the 21 indicators had factor loadings that were significantly different across groups, suggesting that the ways in which these indicators contributed

to the definition of the latent constructs differed among the groups. However, seven of the indicators did not have factor loadings that significantly differed across the groups. Finally, we tested a measurement model in which the seven equivalent indicators were constrained to be equal and the 14 non-equivalent indicators were unconstrained (Model F). As would be expected, this model fit the data well (see Table 2) and did not fit significantly worse than the model without any factor loading constraints (Model B), [$\Delta \chi^2$ (6) = 8.30, n.s.]. We also tested models with a smaller number of factors when applying the factor loading constraints and found that the fit was significantly worse than the 6-factor model (Model F). On the whole, we conclude that the constructs in Fig. 1 are fairly similar across the two groups, though there were some differences in how particular indicators contributed to the latent constructs in the model, suggesting partial measurement equivalence (instead of complete measurement equivalence). However, the observed differences did not suggest that the composition of the latent constructs was different across the countries (i.e., all indicators exhibited significant factor loadings and the overall model fit was excellent across the two samples), but that the relative contribution of specific indicators to the underlying latent constructs was somewhat different in the two countries. Although exact measurement equivalence is desirable, in practice, partial equivalence is more commonly observed in cross-cultural research (Milfont and Fischer 2010). Therefore, when testing our substantive hypotheses in SEM, weutilized Model F as the baseline measurement model.

Tests of Hypotheses

Prior to testing our hypotheses, we compared the results for our U.S. sample to the recent meta-analytic findings by Hülsheger and Shewe (2011). As previously discussed, the majority of the samples included in the meta-analysis were from the U.S. and other similar individualistic cultures. Hulsheger and Schewe found that surface acting and deep acting were positively correlated (r=.18), surface acting was positively correlated with both emotional exhaustion (r = .37), and depersonalization (r = .35) and negatively correlated with personal accomplishment (q = -.07). Deep acting was slightly positively correlated with emotional exhaustion (r = .08), unrelated to depersonalization (r = .04), and positively related to personal accomplishment (q = .21). As can be seen in Table 1, our U.S. sample had a pattern of correlations that is similar to Hülsheger and Shewe (2011), suggesting that our U.S. sample covariance matrix is fairly similar to what is frequently observed in the emotional labor literature.

Bold values in Table 1 indicate means or correlations that are significantly different between the U.S. and China. As can be seen in the table, display rule perceptions were higher in the U.S. than in China (t= -3.17, η^2= .02, p< .05), supporting Hypothesis 1a. Mean surface acting and deep acting were higher in China (t = 4.57, 6.26, η^2 = .04, 07, p< .05, respectively) than in the U.S. sup- porting Hypothesis 1b. Also noteworthy, personal accomplishment had a higher mean in the U.S. than in China (t = -9.75, η^2 = 15, p < .05) (see Table 1). This was not a surprising finding given each country's emphasis on individualism versus collectivism as previously discussed. No other mean differences in the burnout facets were observed.

As can be seen in the Table 1, display rules were more strongly related to surface acting in the U.S. than in China and less strongly related to deep acting in the U.S. compared to in China providing preliminary support for Hypothesis 2. Surface acting was more strongly related to the burnout dimensions in the U.S. than in China providing preliminary support for Hypothesis 3a. Deep acting was more strongly related to two of the burnout dimensions (i.e., depersonalization and emotional exhaustion) in China than in the U.S. providing preliminary support for Hypothesis 3b. Display rules appear to be more strongly related to the burnout dimensions in China than in the U.S. which is the opposite of what was proposed (in Hypothesis 4).

To provide a formal test of these hypotheses, we used SEM. The full structural model (Model G, depicted in Fig. 1) in which the theoretical model was simultaneously imposed on data from the two groups (and included equivalence constraints on some factor loadings) fit the data well (see Table 2; note that this model had identical fit to Model F as all latent variables were allowed to freely correlate with one another, but in the context of a structural model). We then applied constraints of equal paths for all paths in the model (Model H) and found that, while this model fit the data well according to three out of four criteria (see Table 2), it fit significantly worse than the unconstrained model ($\Delta \chi^2$ (11) = 74.57, $p < .05$) suggesting that the relationships among variables differed between the U.S. and China. To test specific hypotheses, we constrained one path at a time and tested whether the change in η^2 was significant between the two countries. A significant change in η^2 indicates that the relationship differs between the countries, effectively meaning that country was a significant moderator of the relationship between the variables under consideration. Bold values in Fig. 2 indicate paths that were significantly different in the U.S. compared to China.

Hypothesis 2 proposed that display rules would be more strongly related to (a) surface acting and (b) deep acting in the U.S. than in China. Hypothesis 2a was supported as display rules was a stronger predictor of surface acting in the U.S. than in China, but Hypothesis 2b was not supported as display rules was not a significantly stronger predictor of deep acting in the U.S. than in China (see Fig. 2). The moderating effect of country on the display rules and surface acting relationship is consistent with Hypothesis 2a (see Fig. 3).

Hypothesis 3a proposed that surface acting would be more strongly related to the facets of burnout in the U.S. compared to China. Hypothesis 3a was supported as surface acting was a stronger predictor of the burnout dimensions in the U.S. compared to China (see Fig. 2). Confirming the expected nature of the relationship, graphs of the country slopes for surface acting in predicting each burnout dimension are presented in Figs. 4, 5, and 6.

Hypothesis 3b, which proposed that deep acting would be more strongly related to the facets of burnout in China compared to the U.S., was only partially supported as deep acting was more strongly related to depersonalization but not the other burnout dimensions in China compared to the U.S. (see Fig. 2). Indeed, it seems that deep acting is more negatively related to depersonalization among individuals in China compared to the U.S., suggesting that putting forth effort to change one's emotions reduces the belief that people are objects more among individuals in China than in the U.S. (see Fig. 7). Deep acting is similarly beneficial for emotional exhaustion in both countries (see Fig. 2).

In terms of our exploration of display rules' direct effect on the dimensions of burnout across the two countries, display rules are negatively related to both emotional exhaustion and depersonalization in China, but not in the U.S. However, this difference is not statistically significant. The trend suggests that display rules have a beneficial impact on service workers in China as compared to service workers in the U.S. As previously discussed, it was anticipated that display rules would have less detrimental effects on individuals in collectivistic versus individualistic cultures because collectivistic cultures have a shared norm supporting emotional control and group harmony. Perhaps rather than just making display rules less harmful, it may be that having an integrative display rule in China actually produces benefits for individuals in the form of lower emotional exhaustion and depersonalization.

Discussion

The results of this study suggest that many of the relationships among emotional labor variables vary as a function of the cultural context under consideration (Grandey et al. 2005). In particular, it was found that the relatively robust sequence of display rules to surface acting to burnout was observed in a U.S. sample

but was not in a Chinese sample, with some relationships being significantly weaker (e.g., surface acting to burnout dimensions) and others exhibiting an opposite sign (e.g., display rules were negatively related to surface acting in the Chinese sample).

Though we did not observe any country differences between display rules and deep acting, we did find that deep acting has a stronger negative relationship with depersonalization in the Chinese sample compared to the U.S. sample. Also, display rule perceptions did not exhibit significantly different direct relationships with the burnout dimensions across country, but the pattern of results showed that display rules are negatively related to emotional exhaustion and depersonalization in the Chinese sample but not in the U.S. sample.

Also of interest, we observed mean differences between the U.S. and China samples on display rule perceptions, with U.S. employees reporting higher levels than the Chinese employees, suggesting that ''service with a smile'' work expectations were more salient for service workers in an individualistic culture than in a collectivistic culture (Mesquita 2000; Mesquita and Delvaux 2013). Further, the emotion regulation strategies of surface acting and deep acting were reported as being used more frequently in China than in the U.S., suggesting that employees in a collectivistic culture engage in more emotion management than employees in an individualistic culture (Matsumoto et al. 2008).

Implications for Emotional Labor Theory

The current findings provide several noteworthy theoretical implications. First, culture appears to matter in the relationships among emotional labor constructs (Grandey et al. 2005). In particular, the presumed deleterious effects of surface acting on well-being outcomes may be more likely to be observed in individualistic cultural contexts than in collectivistic cultural contexts (Mesquita and Delvaux 2013). This pattern of results is consistent with the notion that individuals in a collectivistic culture find surface acting to be less onerous than individuals in an individualistic culture, perhaps due to prevailing cultural norms that place high value on emotional control (Leu et al. 2010). Similarly, the results demonstrated the somewhat surprising finding that, for the Chinese sample, high display rule perceptions were associated with lower levels of burnout, whereas the direct effects in the U.S. sample were not significant. The beneficial effects of display rules in the Chinese sample may be due to the display rules confirming and reinforcing the broader cultural objective of maintaining social harmony and displaying emotions in a strategic way that benefits the greater good (Mesquita and Frijda 1992; Mesquita and Walker 2003). As such, there may be a greater ''display rule and culture fit'' for Chinese participants who rate integrative display rules highly.

It is also noteworthy that the relationships of deep acting with display rules were not significantly different in China compared to the U.S. while only one relationship between deep acting and the dimensions of burnout was significantly different across country (i.e., deep acting to depersonalization). This pattern of findings suggests that individuals from both countries were more likely to deep act when display rule perceptions were high compared to when they were low and that burnout was generally lower when deep acting was high compared to when it was low. In the one significant difference, the negative relationship between deep acting and depersonalization was stronger in China compared to the U.S. This suggests that the benefit of deep acting, in terms of its relationship to burnout, is more salient in the Chinese sample.

Second, although Hochschild's (1983) original theory on emotional labor continues to provide the basis for much of emotional labor research and inquiry, applications of this theory across cultures should be done with caution. The current study found both mean differences in key emotional labor constructs (i.e., display rule perceptions and emotion regulation) across countries as well as country differences in

relationships between display rules, emotional labor, and burnout dimensions. In particular, we argued that cultural elements grounded in emotional moderation versus expressivity (common to collectivistic versus individualistic cultures, respectively) may explain the pattern of results (Mesquita and Delvaus 2013). Building on the potential importance of culture to continued theoretical development in emotional labor, perhaps culture is the "third variable" that explains the lack of previous support in Western samples for the display rule to deep acting to burnout relationship (Hülsheger and Shewe 2011). Thus, perhaps the lack of findings in previous studies for what was originally theorized stems from a lack of modeling or considering cultural differences and implications on the relationships.

Third, concerning theories of measurement and measurement development across cultures, the measures of display rules, emotion regulation, and burnout seemed to generally work well across cultural contexts—suggesting that they are likely tapping the same constructs in the U.S. and China (Cheung and Rensvold 1999). The one exception was for the surface acting items describing faking emotions, which did not seem to fare very well in the Chinese sample. As previously mentioned, the items pertaining to faking emotions had factor loadings close to zero in China, suggesting that they did not effectively match-up with other items used to assess surface acting. Further, the faking items did not load on the deep acting construct either, suggesting that they simply did not assess emotional labor as represented by surface or deep acting. This could mean several things. It could be that the translation was not effective or that faking emotions simply does not translate well into Mandarin Chinese. Or, and more interesting from a theoretical perspective, faking emotions may not be viewed the same way as other surface acting processes (i.e., suppression) among individuals in collectivistic cultures. Given the emphasis on harmony and supporting the group (Hofstede 1980; Noon and Lewis 1992), the idea of faking emotions may be considered grossly inauthentic and potentially conflict evoking in a collectivistic culture. That is, faking may suggest something more disingenuous in a collectivistic culture than do strategies like suppressing or hiding feelings (Eid and Diener 2001). Further, the norm of emotional moderation would suggest that emotions should generally be reduced in level of expression and inward feeling. Faking is inconsistent with this norm in that rather than reducing the emotions expressed or the inward feelings experienced, the individual simply fakes the appropriate expression. The emotion that is being hidden is not reduced per se, but rather hidden in an inauthentic manner. Thus, faking may not be a viable emotional regulation strategy in a collectivistic culture. However, further research is needed to investigate this issue.

Implications for Practice

Given the global market place and the international nature of many corporations, the current study has several meaningful implications for managers of customer service workers. First, managers do not need to assume that emotional labor is bad for all employees in all areas. There is a general tendency toward emphasizing the negative outcomes of engaging in emotional labor (Hochschild 1983; Grandey 2003). However, the current findings suggest that in collectivistic cultures, emotional labor may not have the severe negative effects that are found among samples in individualistic cultures. Thus, managers of customer service workers in collectivistic cultures can demand positive emotional displays from their employees with a bit less concern for the possible burnout associated with such demands.

Second, managers also need to be aware of individual differences within cultures. Even though the current findings suggest that individuals in collectivistic cultures may experience less burnout in response to emotional labor, this should not be taken as an indication that ALL individuals in these cultures will not experience burnout or other negative outcomes. As such, managers should continue to assess the effects of emotional demands on their employees to ensure that the negative outcomes do not inhibit good performance as well as create an atmosphere where turnover and other undesirable outcomes go unchecked.

Third, managers may want to consider cultural variation in training needs, selection processes, and strategies for expanding operations in other countries/cultures. In other words, the ‘‘one-size-fits-all’’ approach to customer service does not appear to be appropriate or necessary and there may be beneficial cost savings when the appropriate training, selection processes, and strategies are employed. Further, diversity initiatives may be a target for emphasizing cultural differences in emotional display demands and provide a motivation for continuing to support such programs. However, related to the previous practical implication, managers should not ignore the individual differences within the culture that may make the implementation of these culturally adapted processes challenging.

Limitations and Future Directions

A potential limitation of the current study stems from the methodological approach used. The design is both cross-sectional design and potentially subject to common method bias within country. As a cross section design, causal inferences cannot be made based on the data alone (Allen et al. 2010). However, given the theoretical and meta-analytic work on the mediation process model tested here, the causal inferences invoked are solidly grounded in previous research and seems appropriate (Hülsheger and Shewe 2011). In terms of common method bias, it is possible that this concern is present for each sample (Podsakoff et al. 2003). However, given that the focus was on comparing established effects across cultures and not on testing new theoretical relationships, we do not believe this is a major concern for current study. It could be argued that if the relationships are inflated in the positive direction due to common method bias, then the differences between the two countries might be washed out by this bias. Since several theoretically meaningful differences were discovered, one could actually argue that the potential presence of this bias actually makes the current study a more conservative assessment of the potential cultural differences that explain the relationships discovered.

Another limitation related to the samples stems from concerns about the equivalence of the samples. As previously shown, the samples were not equivalent in terms age and gender. Although the results were the same even after statistically controlling for these demographic differences, it is unclear whether other differences not currently assessed and controlled in this study could, to some extent, explain the results. Future research needs to be more attentive to this issue and attempt to gather data from samples that are more equivalent. Specifically, a stratified random sample of service workers from the two countries would be an important next step for ruling out these demographic differences.

Another limitation that was previously mentioned is the concern about the measure of surface acting. Two items were removed from the measure because their factor loadings were too low within the China sample. The two items focused on the process of faking one’s emotions in order to accomplish the organizational demands for positive emotional displays. It is unclear whether this problem stems from a translation issue or a cultural difference not modeled here. Future research is needed to discover the issue surrounding this measurement problem.

Related to the previous limitation is the fact that the current study did not include the cultural variables that presumably underlie the country-level differences in the statistical models. Thus, future research should attempt to replicate the current findings while also modeling the culture variables presumed to underlie the country-level differences. Such assessments might include additional cultural constructs that were not the primary focus of the theoretical positions taken in the current investigation.

In terms of future research, researchers might be interested in looking at important individual differences relative to culture within as opposed to between cultures. The current study focused on comparing a collectivistic culture to an individualistic culture in terms of emotional labor. However, individuals within each culture vary on their level of individual versus collective identities (Brewer and Gardner 1996;

Johnson et al. 2006). Future research may want to investigate the extent to which individuals with higher levels of psychological collectivism may experience less of the negative outcomes of engaging in emotional labor. Finding that this or other individual differences related to culture (e.g., the self-concept) mitigate the negative outcomes of emotional labor would provide implications for selection of employees for service jobs as well as training/development once on the job.

Another important future direction pertains to the relative status of service jobs across cultures. Specifically, this study did not investigate the extent to which the prestige and overall view of service jobs in China and the U.S. are similar or different. There is evidence to suggest that gender and sexuality play an important role in service jobs in China whereas these issues are less salient in U.S. service jobs (Hanser 2005; Otis 2008). These differences may, in some way, impact the emotional labor process among service workers in China and the U.S. Future research should investigate these and other prestige or status differences between service jobs in these two cultures to determine if they explain the differences in emotional labor found in this study.

Appendix

See Table 3.

Notes

[1] Structural model analyses including the control variables are available upon request.

Fig. 1 Model of hypothetical relationships

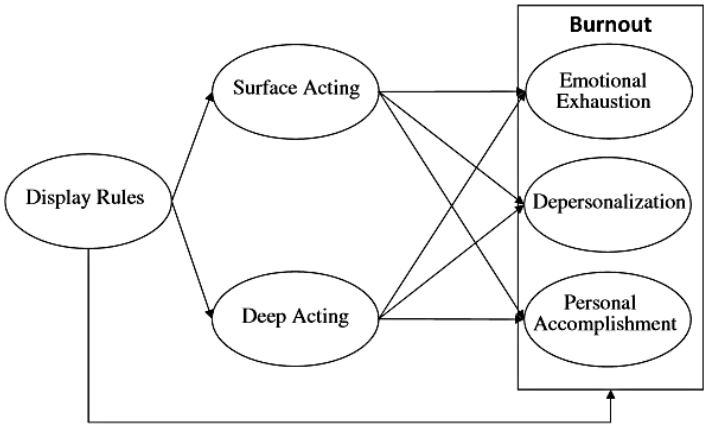
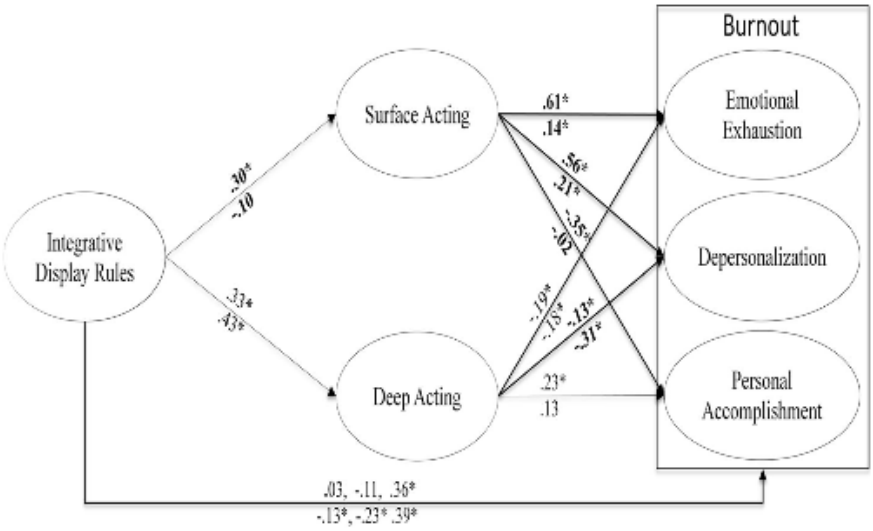


Fig. 2 Model of focal relationships. *Note.* Values above the *line* are for the US, values below the *line* are for China. Links to burnout three coefficients are in the order of: emotional exhaustion, depersonalization, and personal accomplishment. *Bold* terms indicate that the values are significantly different from each other. $*p < .05$



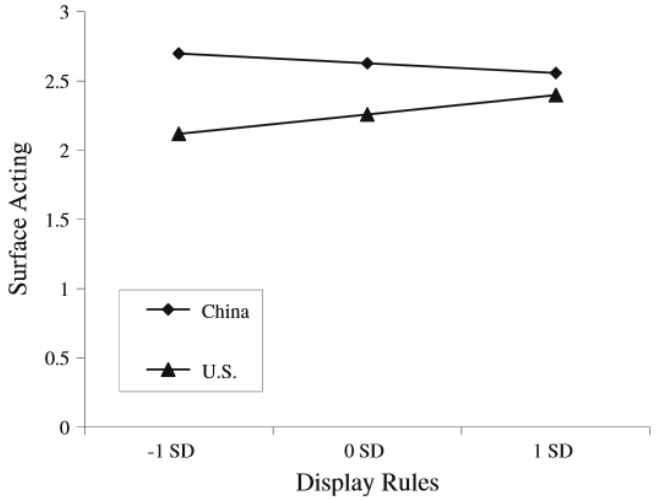


Fig. 3 Moderating effects of country on the display rule to surface acting relationship

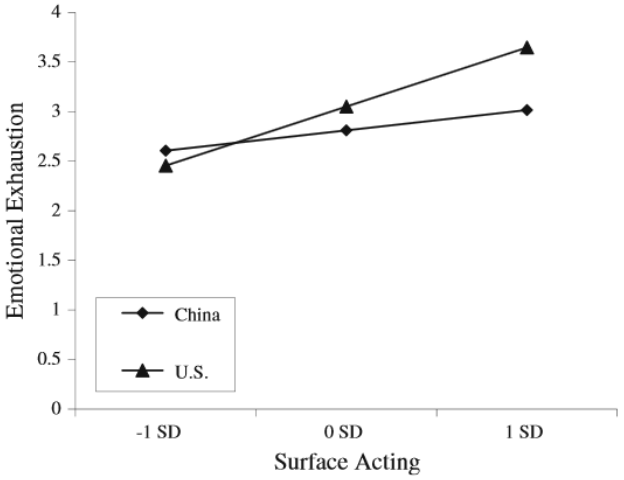


Fig. 4 Moderating effects of country on the surface acting to emotional exhaustion relationship

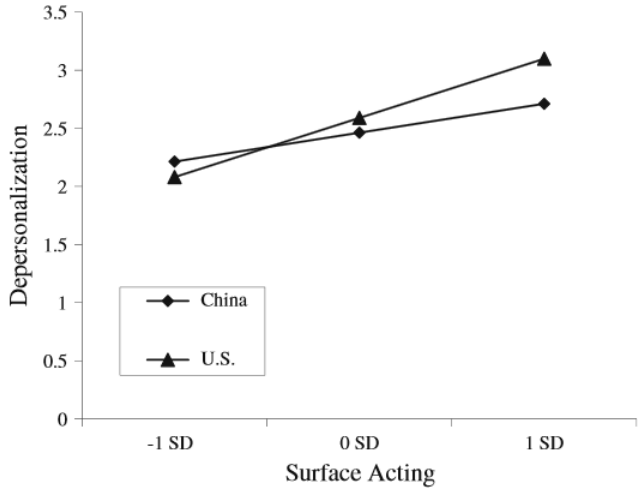


Fig. 5 Moderating effects of country on the surface acting to depersonalization relationship

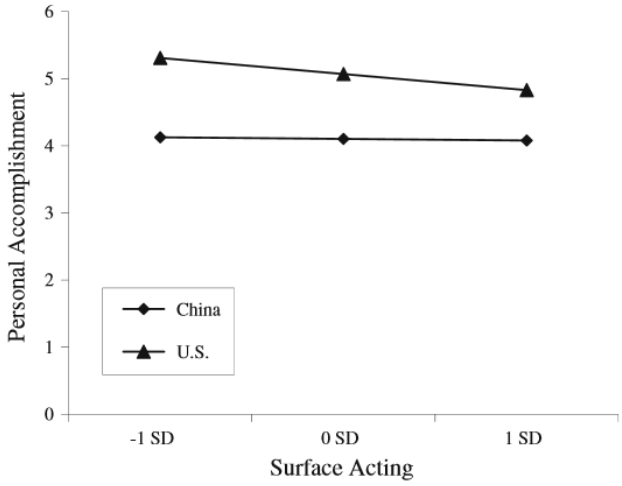


Fig. 6 Moderating effects of country on the surface acting to personal accomplishment relationship

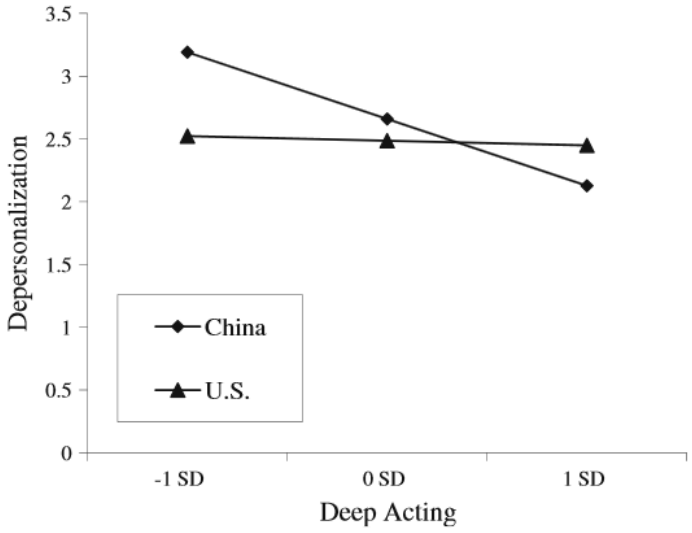


Fig. 7 Moderating effects of country on the deep acting to depersonalization relationship

Table 1 Means, standard deviations, reliabilities, and intercorrelations among variables

	1	2	3	4	5	6
1. Display rules	.81 (.87)	**−.01**	**.48***	.47*	**−.45***	**−.36***
2. Surface acting	**.19***	.93 (.78)	**−.02**	**−.01**	**.15***	**.15***
3. Deep acting	**.26***	**.16***	.92 (.79)	.30*	**−.41***	**−.34***
4. Personal accomplishment	.33*	**−.22***	.25*	.79 (.83)	**−.04**	−.09
5. Depersonalization	**−.06**	**.47***	**−.05**	**−.25***	.88 (.87)	.71*
6. Emotional exhaustion	**−.08**	**.52***	**−.07**	**−.15***	.72*	.92 (.82)
US mean	**4.03**	**2.69**	**3.35**	**5.06**	2.50	2.95
US SD	0.76	1.12	1.03	1.28	1.27	1.36
China mean	**3.82**	**3.04**	**3.81**	**4.05**	2.51	2.84
China SD	0.68	0.55	0.61	1.07	0.94	0.76

Note Coefficient alphas are presented on the diagonal; China alphas are in parentheses. US values are below the diagonal ($n = 280$); China values are above the diagonal ($n = 231$). Bold correlations and means indicate that the US and China correlations and means are significantly different from each other. All tests are two-tailed

* $p < .05$

Table 2 Summary of fit statistics

Model	π^2	df	RMSEA	$SRMR_{USA}$	$SRMR_{China}$	TLI	CFI
Set 1: 21 indicators							
A. 6 factors	1027.48*	430	.076	.050	.110	.95	.96
Set 2: 21 indicators (drop 2 surface acting items)							
B. 6 factors	720.53*	348	.064	.051	.065	.96	.97
C. 5 factors (combine DA and SA)	1787.77*	358	.130	.150	.094	.86	.88
D. 4 factors (combine PA, DP, and EE)	2156.33*	366	.170	.120	.190	.83	.85
Set 3: 21 indicators and cross-cultural constraints							
E. 6 factors (full equivalence)	766.04*	363	.066	.053	.076	.96	.97
F. 6 factors (partial equivalence)	728.83*	354	.063	.051	.068	.96	.97
Set 4: partial measurement equivalence and structural models							
G. Structural model (Fig. 1)	728.83*	354	.063	.051	.068	.96	.97
H. Structural model all constraints	803.40*	365	.068	.090	.100	.96	.96

RMSEA root mean square error of approximation, *SRMR* standardized root mean squared residual, *TLI* Tucker–Lewis index, *CFI* comparative fit index

* Significant at $p < .05$

Table 3 Primary factor loadings

	USA	China	Constrained
1. Display rule—Parcel 1*	.67	.83	.75
2. Display rule—Parcel 2	.93	.92	.93
3. Display rule—Parcel 3	.92	.90	.91
4. SA1—I put on an act in order to deal with customers in an appropriate way*	.78	.87	.80
5. SA2—I put on a "show" or "performance" when interacting with customers*	.89	.67	.83
6. SA3—I just pretend to have the emotions I need to display for my job*	.91	.58	.83
7. SA4—I put on a "mask" in order to display the emotions I need for the job*	.86	.62	.79
8. SA5—I show feelings to customers that are divergent from what I feel inside*	.80	.54	.73
9. DA1—I try to actually experience the emotions that I must show to customers	.79	.76	.77
10. DA2—I make an effort to actually feel the emotions that I need to display toward others*	.88	.79	.85
11. DA3—I work hard to feel the emotions that I need to show to customers*	.87	.74	.84
12. DA4—I work at developing the feelings inside of me that I need to show to customers*	.88	.50	.80
13. Personal accomplishment—Parcel 1*	.96	.88	.93
14. Personal accomplishment—Parcel 2	.77	.82	.79
15. Personal accomplishment—Parcel 3	.79	.79	.79
16. Depersonalization—Parcel 1*	.69	.78	.72
17. Depersonalization—Parcel 2	.89	.86	.88
18. Depersonalization—Parcel 3	.79	.80	.80
19. Emotional exhaustion—Parcel 1*	.92	.73	.88
20. Emotional exhaustion—Parcel 2*	.95	.84	.92
21. Emotional exhaustion—Parcel 3*	.90	.83	.88

Con. constrained model, which reflects the value of the loading when it is constrained to be equal across the two samples

* The factor loadings are significantly different in the U.S. and Chinese samples

References

Allen, J. A., Pugh, S. D., Grandey, A. A., & Groth, M. (2010). Display rules and emotional labor: The moderating role of customer orientation. *Human Performance, 23*(2), 101–115. doi:10.1080/08959281003621695.

Best, R. G., Downey, R. G., & Jones, R. G. (1997, April). *Incumbent perceptions of emotional work requirements.* Paper presented at the 12th annual conference of the Society for Industrial and Organizational Psychology, St. Louis, MO.

Brewer, M. B., & Gardner, W. (1996). Who is the "We"? Levels of collective identity and self representations. *Journal of Personality and Social Psychology, 71*(1), 83–93. doi:10.1037/0022-3514.71.1.83.

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1*(3), 185–216. doi:10.1177/135910457000100301.

Brotheridge, C. M., & Grandey, A. A. (2002). Emotional labor and burnout: Comparing two perspectives of "people work". *Journal of Vocational Behavior, 60*, 17–39. doi:10.1006/jvbe.2001.1815.

Brotheridge, C. M., & Lee, R. T. (2003). Development and validation of the Emotional Labour Scale. *Journal of Occupational and Organizational Psychology, 76*, 365–379. doi:10.1348/096317903769647229.

Butler, E. A., Lee, T. L., & Gross, J. J. (2007). Emotion regulation and culture: Are the social consequences of emotion suppression culture-specific? *Emotion, 7*(1), 30–48. doi:10.1037/1528-3542.7.1.30.

Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*(3), 456–466. doi:10.1037/0033-2909.105.3.456.

Cheung, G. W., & Rensvold, R. B. (1998). Cross-cultural comparisons using non-invariant measurement items. *Applied Behavioral Science Review, 6*(1), 93–110. doi:10.1016/S1068-8595(99)80006-3.

Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management, 25*(1), 1–27. doi:10.1177/014920639902500101.

Cheung, G. W., Murrmann, K. F., Murrmann, S. K., & Becker, C. (2004). Noninvariant measurement versus traditional approaches for studying cultural differences: A case of service expectations. *Journal of Hospitality and Tourism Research, 28*, 375–390. doi:10.1177/1096348004265282.

Cote, S. (2005). A social interaction model of the effects of emotion regulation on work strain. *Academy of Management Review, 30*(3), 509–530. doi:10.5465/AMR.2005.17293692.

Diefendorff, J. M., & Richard, E. M. (2003). Antecedents and consequences of emotional display rule perceptions. *Journal of Applied Psychology, 88*, 284–294. doi:10.1037/0021-9010.88.2.284.

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. doi:10.1016/j.jvb.2004.02.001.

Diefendorff, J. M., Erickson, R. E., Grandey, A. A., & Dahling, J. J. (2011). Emotional display rules as work unit norms: A multilevel analysis of emotional labor among nurses. *Journal of Occupational Health Psychology, 16*, 170–186. doi:10.1037/a0021725.

Eid, M., & Diener, E. (2001). Norms for experiencing emotions in different cultures: Inter- and intranational differences. *Journal of Personality and Social Psychology, 81*(5), 869–885. doi:10.1037/0022-3514.81.5.869.

Grandey, A. A. (2000). Emotion regulation in the workplace: A new way to conceptualize emotional labor. *Journal of Occupational Health Psychology, 5*(1), 95–110. doi:10.1037/1076-8998.5.1.95.

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. doi:10.2307/30040678.

Grandey, A. A., Fisk, G. M., Mattila, A. S., Jansen, K. J., & Sideman, L. A. (2005a). Is "service with a smile" enough? Authenticity of positive displays during service encounters. *Organizational Behavior and Human Decision Processes, 96*, 38–55. doi:10.1016/j.obhdp.2004.08.002.

Grandey, A. A., Fisk, G. M., & Steiner, D. D. (2005b). Must "service with a smile" be stressful? The moderating role of personal control for American and French employees. *Journal of Applied Psychology, 90*(5), 893–904. doi:10.1037/0021-9010.90.5.893.

Grandey, A., Rafaeli, A., Ravid, S., Wirtz, J., & Steiner, D. (2010). Emotion display rules at work in the global service economy: The special case of the customer. *Journal of Service Management, 21*(3), 388–412. doi:10.1108/09564231011050805.

Grandey, A. A., Diefendorff, J. M., & Rupp, D. E. (2013). Emotional labor: Overview of definitions, theories, and evidence. In A. A. Grandey, J. M. Diefendorff, & D. E. Rupp (Eds.), *Emotional labor in the 21st century: Diverse perspectives on emotion regulation at work*. New York: Psychology Press/Routledge.

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. doi:10.1037/0022-3514.74.1.224.

Hanser, A. (2005). The gendered rice bowl: The sexual politics of service work in urban China. *Gender & Society, 19*(5), 581–600. doi:10.1177/0891243205276794.

Hochschild, A. R. (1983). *The managed heart: Commercialization of human feeling*. Berkeley: University of California Press.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. London: Sage.

Hülsheger, U. R., & Shewe, A. F. (2011). On the costs and benefits of emotional labor: A meta-analysis of three decades research. *Journal of Occupational Health Psychology, 16*(3), 361–389. doi:10.1037/a0022876.

Johnson, R. E., Selenta, C., & Lord, R. G. (2006). When organizational justice and the self-concept meet: Consequences for the organization and its members. *Organizational Behavior and Human Decision Processes, 99*, 175–201. doi:10.1016/j.obhdp.2005.07.005.

Leu, J., Mesquita, B., Ellsworth, P. C., ZhiYoung, Z., Huijuan, Y., Buchtel, E., et al. (2010). Situational differences in dialectical emotions: Boundary conditions in a cultural comparison of North Americans and East Asians. *Cognition and Emotion, 24*(3), 419–435. doi:10.1080/02699930802650911.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*, 224–253. doi:10.1037/0033-295X.98.2.224.

Maslach, C., & Jackson, S. E. (1981). *The Maslach Burnout Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Masuda, T., Ellsworth, P. C., Mesquita, B., Leu, J., Tanida, S., & Van de Veerdonk, E. (2008). Placing the face in context: Cultural differences in the perception of facial emotion. *Journal of Personality and Social Psychology, 94*(3), 365–381. doi:10.1037/0022-3514.94.3.365.

Matsumoto, D., Yoo, S. H., & Nakagawa, S. (2008). Culture, emotion regulation, and adjustment. *Journal of Personality and Social Psychology, 94*(6), 925–937. doi:10.1037/0022-3514.94.6.92.

Mesquita, B. (2000). Emotions in collectivist and individualist contexts. *Journal of Personality and Social Psychology, 80*, 68–74. doi:10.1037/0022-3514.80.1.68.

Mesquita, B., & Delvaux, E. (2013). A cultural perspective on emotion labor. In A. A. Grandey, J. M. Diefendorff, & D. E. Rupp (Eds.), *Emotional labor in the 21st century: Diverse perspectives on emotion regulation at work*. New York: Psychology Press/Routledge.

Mesquita, B., & Frijda, N. H. (1992). Cultural variations in emotions: A review. *Psychological Bulletin, 112*(2), 179–204. doi:10.1037/0033-2909.112.2.179.

Mesquita, B., & Walker, R. (2003). Cultural differences in emotions: A context for interpreting emotional experiences. *Behaviour Research and Therapy, 41*, 777–793. doi:10.1016/S0005-7967(02)00189-4.

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research, 3*(1), 111–121.

Noon, J. M., & Lewis, J. R. (1992). Therapeutic strategies and outcomes: Perspectives from different cultures. *British Journal of Medical Psychology, 65*, 107–117. doi:10.1111/j.2044-8341.1992.tb01691.x.

Oetzel, J., Ting-Toomey, S., Masumoto, T., Yokochi, Y., Pan, X., Takai, J., et al. (2001). Face and facework in conflict: A cross-cultural comparison of China, Germany, Japan, and the United States. *Communication Monographs, 68*(3), 235–258. doi:10.1080/03637750128061.

Otis, E. M. (2008). Beyond the industrial paradigm: Market-embedded labor and the gender organization of global service work in China. *American Sociological Review, 73*, 15–36. doi:10.1177/000312240807300102.

Podsakoff, P. M., MacKenzie, S. B., Lee, J., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879–903. doi:10.1037/0021-9010.88.5.879.

Rafaeli, A., & Sutton, R. I. (1987). Expression of emotion as part of the work role. *Academy of Management Review, 12*(1), 23–37. doi:10.2307/i302988.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin, 114*, 552–566. doi:10.1037/0033-2909.114.3.552.

Rupp, D. E., & Spencer, S. (2006). When customers lash out: The effects of customer interactional injustice on emotional labor and the mediating role of discrete emotions. *Journal of Applied Psychology, 91*(4), 971–978. doi:10.1037/0021-9010.91.4.971.

Safdar, S., Friedlmeier, W., Matsumoto, D., Yoo, S. H., Kwantes, C. T., Kakai, H., et al. (2008). Variations of emotional display rules within and across cultures: A comparison between Canada, USA, and Japan. *Canadian Journal of Psychology, 41*(1), 1–10. doi:10.1037/a0014387.

Uchida, Y., Kitayama, S., Mesquita, B., Reyes, J. A. S., & Morling, B. (2008). Is perceived emotional support beneficial? Well-being and health in independent and interdependent cultures. *Personality and Social Psychology Bulletin, 34*, 741–754. doi:10.1177/0146167208315157.

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*, 4–70. doi:10.1177/109442810031002.

Yoo, B., & Donthu, N. (2011). Testing cross-cultural invariance of the brand equity creation process. *Journal of Product & Brand Management, 11*(6), 380–398. doi:10.1108/10610420210445505.