1-1-2020

General emotional labor scale: Development and establishing psychometric properties

Falak Zehra Mohsin
University of Karachi, Pakistan

Nadia Ayub
Institute of Business Administration, Karachi, Pakistan

Recommended Citation

General Emotional Labor Scale: Development and establishing psychometric properties

Falak Zehra Mohsin · Nadia Ayub

Abstract This study outlines the process followed in the development of the General Emotional Labor Scale-English (GELS-E). The scale construction process was divided into different phases. The first phase entailed the item generation and the second phase was the pilot study. Phase three and four dealt with developing the norms and establishing the psychometric properties (reliability and validity) of the final 30 items. It can be concluded that the GELS-E was developed as a measure that can be used by researchers to assess emotional labor in professionals. Such a measure of emotional labor can prove to be beneficial for researchers and organizations alike. It can be used to understand the performance and effects of emotional labor and arrange trainings needed to help employees with the performance of emotional labor.

Keywords Emotional labor · Scale construction · Reliability · Validity

1 Introduction

Emotional labor is a concept introduced by Hochschild (1983) who described it as the work of modifying ones' emotions to display appropriate work-related emotions (Hochschild 2003). Emotional labor is the conscious control of emotions in order to exhibit suitable verbal and body language. Ashforth and Humphrey (1993) described it further by focusing on the outward expressions of emotions i.e. 'impression management'. According to them performance of emotional labor involves surface and deep acting, as well as genuine acting. This was considered because even though an employee might not be genuinely feeling an emotion the expression of it is part of their job requirement and therefore is labor.

Falak Zehra Mohsin
University of Karachi-Pakistan
E-mail: f.zehra.m@outlook.com

Nadia Ayub
Institute of Business Management Karachi-Pakistan
©Mohsin, Z. F. and Ayub, N. 2020

Morris and Feldman (1996) related emotional labor to organizational goals. They defined it as the 'effort, planning, and control' necessary to exhibit the emotions desired by the organization. The interactionist viewpoint was followed in this conceptualization i.e. the interaction between the employee as an individual and the organizational factors. Grandey (2000), in an attempt to consolidate these viewpoints, presented a comprehensive definition integrating the feelings, the process and the expression of emotions at work.

Theories of emotional labor have presented three different dynamics - internal states, internal processes and external displays of emotions. The definition and concept of emotional labor followed in the current study encompasses these three dynamics. Internal states includes the concept of emotional dissonance (the state when individuals experience an incongruity between their felt and displayed emotions) (Hochschild 2003). Internal processes involve the regulation of emotions, which include surface, deep and genuine acting, and act as a bridge between the felt state and the displayed emotions (TM 2004).

When surface acting is the mode of performance, expression of emotion occurs only at the face level. An example could be a flight attendant having a bad day, but they would 'put on a mask' and display a positive attitude towards the passengers (Miller et al 2007). The second form is deep acting. This involves working to change ones' felt emotions. Hochschild (2003) described it as 'getting into character', this can be done through simulations of the required feelings until these feelings are actually felt (Ashkanasy and Daus 2002). Ashforth and Humphrey (1993) presented genuine acting - they believe that it is not necessary that employees do not feel the emotions they are required to display. For instance, a nurse who is supposed to display sympathy towards a patient may actually feel it as well.

The Pakistani society differs greatly from the developed western world. Many of the western countries are individualistic societies, however, the Pakistani society is predominantly collectivistic in nature. Pakistanis are typically close to their families and friends and many live in joint families or at least elderly parents still reside with their children even after the children are married (Taqui et al 2007). Additionally, Pakistan is majorly an Islamic nation and the masses try to abide by the Islamic laws and guidelines. Religion provides Muslims with guidelines about outward behavior with regard to situations. For instance, females are to display modesty when interacting with the opposite sex. Hence, the performance and experience of emotional labor by the Pakistani workforce is likely to differ from that of the labor force of other countries.

Emotional labor is a predominant requirement in the service sector. The economy of Pakistan consists of a relatively large service sector; research has shown that it contributes more than fifty percent to the country's GDP (Ahmed and Ahsan 2011). The objective of the current study was to tap this niche by developing a scale to measure the phenomena. The idea behind the development of the scale is to construct a measure that is culturally relevant to Pakistan and can provide greater understanding of emotional labor and its experience by the Pakistani workforce.

A review of existing scales showed gaps in measurement of emotional labor, for instance the Teacher Emotional Labor Scale (Cukur 2009) and the

Hospitality Emotional Labor Scale were developed and validated as measures particularly for teachers and the hospitality industry; the Discrete Emotions Emotional Labor Scale by TM (2004) measures 14 particular emotions - limiting the range of emotions measured; and the Emotional Labor Scale (Brotheridge and Lee 2003) does not incorporate genuine acting as means of emotional labor. Additionally, the aim was to use easy English so that the Pakistani masses can understand and to tap the cultural effect on the performance of emotional labor.

2 Method

Following Grandey's comprehensive approach, emotional labor was operationalized as the process of managing internal emotions and the outward expression of emotions in order to adhere to organizational rules. Deep, surface and genuine acting were determined as the subscales. Surface acting was operationalized as the display of emotions, even if they are actually not being felt. Deep acting was defined as the effort put in to modify felt emotions in order to try and feel them. Genuine acting was operationalized as the expression of spontaneous and genuinely felt emotions. Moreover, three facets were determined for each subscale: variety, refers to the diversity of emotions; intensity denotes the strength of the emotions and the 'general' items measure aspects of emotional display through different means such as the tone of voice, body language and so on.

2.1 Phase I - Part 1

Procedure: A thorough literature review was conducted to understand the phenomenon and to generate items for the scale. Initially a total of 100 items were generated that measured deep, surface and genuine acting, duration, frequency, variety and intensity of emotional labor and the experience of emotions at work. All the items were theoretically based. 49 items were then drawn through thorough subjective analysis and with the input of two organizational psychologists.

The 49 item scale was sent to an English Language expert and to four psychologists for language assessment and face and subjective validity of the scale. The corrections and modifications suggested by the above experts were analyzed and incorporated.

2.2 Phase I - Part 2: Pilot Study

Sample: The sample consisted of 510 working individuals living in Karachi. Snowball sampling was used. Of the participants that answered 48.2% were females and 51.8% were males. A majority of them (48%) were between the ages of 20 and 30 years, 49.4% of the respondents were married and 50.6% were postgraduates.

Procedure: The 49 items developed for the GELS were distributed to the target population. Consent was obtained from the respondents. In addition the participants were made aware of the fact that their participation in the study

was voluntary and they have the right to withdraw. Also, confidentiality was assured. The sample was asked to fill out a demographic information form and the entire GELS via Google Docs.

The data collected was used for item analysis, establish internal consistency and to define the cut-off points. For this purpose statistical analysis such as descriptive statistics, item total correlations, factor analysis and factor loadings of each subscale were conducted using the Statistical Package for the Social Sciences version 17.

2.3 Phase II - Part 1: Establishing Reliability

Sample: The sample consisted of 110 individuals of which 46.4% were females and 53.6% were males. 58.2% of the participants were between the ages of 20-30, 23.6%, 15.5% and 2.7% were between the ages 31-40, 41-50, and 51-60 respectively. Majority of the respondents (39.1%) held a postgraduate degree, and majority of them (47.3%) were married.

Procedure: The respondents answered a set of questionnaires which included a consent form, demographic information form and the GELS via Google Docs. For test-retest reliability, after a two week (15 days) interval, the participants were asked to answer the GELS again.

Cronbach's alpha of the GELS-E, item-subscale correlations, Cronbach's alpha values of each subscale, and correlation between the two administrations of the scale were analyzed. Statistical analysis was conducted using the Statistical Package for Social Sciences version 17.

2.4 Phase II - Part 2: Establishing Validity

Sample: The scale was administered on a sample of 303 individuals. In this sample 55.4% of the participants were females and 44.6% were males; 46.2% were between the ages of 20-30, 26.1% were between 31-40 years, 11.9% were between 41-50 years, 13.2% were between 51-60 years; and 2.6% were above 60 years of age. Majority of the respondents (59.1%) held a postgraduate degree. The participants were from a variety of different professions and occupations. Most of them (57.8%) were married. Non-random, snowball sampling was used for the purpose of this study.

Research Instruments: The demographic information form comprised of items related to the respondents' age, gender, education, and marital-status, number of dependents, socio-economic status, profession/occupation, and nature of organization (public or private).

The Emotional Labor Scale was developed by Brotheridge and Lee (2003). This scale measures Emotional Labor according to international norms and therefore could help establish concurrent and convergent validity. The items were to be answered on a 5-point Likert scale ranging from never to always. The Emotional Labor scale includes five subscales: frequency, intensity, variety, surface acting and deep acting. In the current sample, the internal reliability of the scale was 0.788.

Emotion Regulation Questionnaire (ERQ) was developed by Gross and John (2003). The ERQ was used as a measure of emotional regulation, which is related and influences how individuals perform emotional labor. ERQ consists of two subscales: reappraisal and suppression. It is a 10-item instrument with a 7-point response scale (strongly disagree to strongly agree). In the current sample the internal consistency of the scale was 0.814.

Discrete Emotions Emotional Labor Scale (DEELS) was developed by TM (2004). It consists of three subscales; Genuine Acting, Faked Expression and Suppression and measures them on 14 distinct emotions. The DEELS subscale of genuine expression established convergent validity of the genuine acting subscale of GELS-E, and the faked expression and suppression subscales performed validity analysis of the surface acting subscale of GELS-E. The response scale ranges from 1 (never) to 5 (many times a day) for genuine expression and faked expression of emotions. However, for suppression the response scale includes a 0 (I never feel this), 1 (never) to 5 (many times a day). The alpha value for the scale in the current sample was 0.788.

Schutte Self-report Emotional Intelligence Test (SSEIT) (1998) was developed by Schutte, Malouff, Hall, Haggerty, Cooper, Golden, and Dornheim. SSEIT assesses emotional intelligence and comprises of 33 self-report items. The items are to be responded on a 5-point scale (1 is strongly disagree and 5 is strongly agree) and three items (5, 28 and 33) are reverse scored. In the current sample the scale yielded an alpha value of 0.852.

Positive Affect and Negative Affect Schedule (PANAS) was developed by Watson, Clark and Tellegen in 1998. The scale measures positive and negative affectivity which have been described as moderators and consequences of emotional labor. Therefore it was imperative to establish discriminant validity. PANAS is a 10 item inventory, to be responded on a 5-point scale where 1= very slightly or not at all and 5 = extremely. The internal reliability of the scale in the current sample was 0.779.

Satisfaction with Life Scale (SWLS) was developed by Diener et al (1985). It is a 5-item scale scored on a 7-point scale ranging from Strongly Disagree (1) to Strongly Agree (7). The scale shows high internal reliability of 0.886 in the current sample.

Copenhagen Burnout Inventory (CBI): Burnout is said to be a common outcome of emotional labor, therefore distinguishing them was essential. The CBI is a public domain instrument developed as part of a PUMA study conducted in 2005. The authors of the scale are Kristensen, Borritz, Villadsen and Christensen. It contains 19 items answered on a 5-point scale (from always or to a very high degree to never/almost never or to a very low degree). In the current sample, the internal reliability of the scale was 0.755.

Ryff's Psychological Well-Being Scale (1989, 1995) the original scale consisted of 120 items and was developed by Ryff. For the current research, the shorter 18-item version of the scale was used. The responses range from strongly agree to strongly disagree answered on a 7-point scale. Based on the current sample, the internal reliability of the scale yielded an alpha value of 0.863.

Procedure: Face and subjective validity of the scale was established in two stages. In the first stage the initial item pool developed was sent to a language

expert and four psychologists. In the second stage, 14 psychology students (7 undergraduate and 7 graduate students) were approached and were asked to review the items.

To establish convergent and discriminant validity permission to use the above mentioned scales was obtained from the relevant authors. The questionnaires, along with an informed consent form and the demographic information form were sent to the participants via Google Docs. Statistical analysis of the data was conducted using Statistical Package for the Social Sciences version 17.

3 Results

The initial pool of 49 items was reduced to 30 items through factor analysis (table 2), descriptive statistics and item-total correlations (table 1). Corresponding items that had high loadings and item-total correlations greater than 0.50 were retained as part of the scale. Research posits that in exploratory factor analysis, factor loadings are deemed to be significant only when they are higher than 0.30 or 0.40 (Floyd and Widaman 1995). The Kaiser-Meyer-Olkin Measure (KMO) value was found to be 0.889 and the Bartlett test result in the current sample was found to be highly significant ($p < 0.001$), and therefore factor analysis is deemed appropriate.

Exploratory factor analysis determined the construct validity of the scale and examined the underlying structure of the scale. Table 2 displays the factor loadings for the 49 items GELS-E. The table shows that the three hypothesized subscales in fact came forth as the three factors; Surface, Deep and Genuine Acting and each comprised of 10 items. Item-Subscale Correlations were computed to assess each item of a factor significantly related to the particular subscale (tables 3, 4, & 5).

The Cronbach's alpha score obtained for the GELS-E reveals a highly significant correlation (table 6). Table 7 presents the internal consistency of all the three subscales. Additionally, test-retest reliability was calculated and the correlation coefficient shows that the test retest reliability of the GELS-E is highly significant ($r = 0.912$; $p < 0.001$; $n = 110$).

Convergent validity of GELS-E was assessed by correlating it with Emotional Labor Scale (Brotheridge and Lee 2003); the two facets of Emotion Regulation Questionnaire and with Discrete Emotions Emotional Labor Scale (TM 2004). Statistical analysis revealed a positive correlation between GELS-E with Emotional Labor Scale ($r = 0.671$; $p < 0.001$; $n = 303$) and a positive relationship was observed between GELS-E with ERQ's Reappraisal ($r = 0.686$; $p < 0.001$; $n = 303$) and GELS-E with ERQ's Suppression ($r = 0.544$; $p < 0.001$; $n = 303$). A positive correlation was also observed between GELS-E with DEELS ($r = 0.713$; $p < 0.001$; $n = 303$).

Moreover, discriminant validity was evaluated by correlating the scale with Schutte Self-report Emotional Intelligence Test (SSEIT); Positive Affect and Negative Affect Schedule (PANAS); Satisfaction with Life Scale (SWLS); Copenhagen Burnout Inventory (CBI); and Ryff's Psychological Well-Being Scale. Statistical analysis revealed no correlations between GELS-E and SSEIT ($r =$

Table 1: Descriptive statistics and item total correlations of 49 items of the GELS-E

No.	M	SD	Pearson r	Sig	No.	M	SD	Pearson r	Sig
1	2.880	1.192	0.532	0.000	26	2.630	1.255	0.521**	0.000
2	2.580	1.176	0.686**	0.000	27	2.550	1.327	0.645**	0.000
3	3.140	1.296	0.596**	0.000	28	2.840	1.284	0.401**	0.000
4	3.840	0.956	0.106*	0.017	29	3.310	1.249	0.590**	0.000
5	3.600	1.192	0.738**	0.000	30	2.890	1.211	0.013	0.767
6	2.850	1.199	0.617**	0.000	31	3.400	1.188	0.254**	0.000
7	2.600	1.144	0.544**	0.000	32	3.250	1.163	0.760**	0.000
8	3.060	1.213	0.323**	0.000	33	2.540	1.231	0.133**	0.003
9	3.060	1.295	0.126**	0.004	34	3.400	1.222	0.291**	0.000
10	3.190	1.191	0.015	0.728	35	2.450	1.283	0.670	0.000
11	3.130	1.375	0.248**	0.000	6	2.920	1.265	0.560**	0.000
12	3.190	1.365	0.175**	0.000	37	2.600	1.102	0.894**	0.000
13	3.280	1.294	0.196**	0.000	38	2.520	1.131	0.260**	0.000
14	3.530	1.250	0.540**	0.000	39	2.530	1.076	-0.062	0.164
15	2.590	1.231	0.520**	0.000	40	2.650	1.247	0.565**	0.000
16	2.200	1.005	0.670*	0.000	41	3.270	1.321	0.738**	0.000
17	3.560	1.233	0.589**	0.000	42	3.080	1.325	0.059	0.186
18	2.650	1.210	0.600**	0.000	43	2.360	1.164	0.763**	0.000
19	2.910	1.245	0.790	0.000	44	2.470	1.273	0.651	0.000
20	3.220	1.165	0.511**	0.000	45	3.510	1.094	0.372**	0.000
21	3.270	1.229	0.501**	0.000	46	3.050	1.298	0.633**	0.000
22	2.420	1.199	0.232**	0.000	47	3.350	1.310	0.824**	0.000
23	2.710	1.283	0.566**	0.000	48	3.220	1.289	0.218**	0.000
24	3.210	1.322	0.602**	0.000	49	2.780	1.238	0.668**	0.000
25	3.350	1.190	-0.020	0.645					

**. Correlation is significant at the 0.01 level (2-tailed),*. Correlation is significant at the 0.05 level (2-tailed).

0.003; p < 0.001; n = 303); SWLS (r = 0.204; p < 0.001; n = 303) and Negative Affect of PANAS (r = 0.011; p < 0.001; n = 303); and negative correlations between GELS-E and Positive Affect of PANAS (r = -0.179; p < 0.001; n = 303); CBI (r = 0.451; p < 0.001; n = 303) and Ryff's Psychological Well-being Scale (r = 0.393; p < 0.001; n = 303).

4 Discussion

The aim of the research study was to develop a consistent, reliable and valid measure of emotional labor. The scale developed consists of 30 items that measure Emotional Labor through the three subscales of Surface, Deep, and Genuine Acting. The scale was developed as a self-report questionnaire that is to be answered on a Likert type 5-point response scale.

Whilst developing the items certain basic rules and principles were kept in mind (Clark and Watson 2016). The language of the items was kept simple, direct, colloquialisms and slang language was avoided and language that could be understood by a range of different populations was utilized; difficult and complex items were avoided as were double-barreled items. Negatively coded items were not included in the scale, studies show that scales that include positively and negatively worded items contain measurement errors (DiStefano and Motl 2006; Quilty et al 2006).

F. Mohsin, N. Ayub

Table 2: Factor structure and eigen values of GELS-E with varimax rotation

No.	Item No.	Factor I	Factor II	Factor III
1.	1	0.649	0.011	-0.456
2.	6	0.676	-0.286	-0.325
3.	7	0.689	-0.080	0.040
4.	12	0.394	-0.003	-0.431
5.	13	0.335	0.135	-0.377
6.	15	0.591	0.060	-0.465
7.	16	0.688	-0.044	-0.145
8.	19	0.668	-0.006	-0.337
9.	23	0.633	-0.065	-0.371
10.	27	0.553	0.150	-0.408
11.	30	0.344	-0.044	-0.464
12.	33	0.212	-0.156	-0.289
13.	35	0.695	0.069	-0.467
14.	39	0.233	-0.209	-0.281
15.	44	0.588	-0.049	-0.335
16.	2	0.022	0.612	-0.063
17.	4	-0.099	0.478	0.168
18.	8	-0.030	0.215	0.008
19.	11	-0.102	0.468	0.024
20.	18	0.270	0.536	0.199
21.	20	-0.221	0.729	0.206
22.	22	-0.083	0.222	-0.008
23.	26	-0.217	0.736	-0.092
24.	28	-0.104	0.477	-0.457
25.	32	0.071	0.651	-0.112
26.	36	0.163	0.563	0.151
27.	38	0.670	0.189	-0.108
28.	40	0.004	0.663	0.031
29.	42	-0.436	0.247	0.136
30.	43	-0.297	0.528	0.243
31.	45	-0.153	0.323	0.030
32.	47	-0.03	0.633	-0.034
33.	49	-0.146	0.597	0.168
34.	3	-0.116	-0.036	0.668
35.	5	-0.130	0.129	0.640
36.	9	-0.318	0.171	0.294
37.	10	-0.420	0.123	0.338
38.	14	-0.252	-0.124	0.621
39.	17	-0.063	-0.114	0.690
40.	21	-0.468	0.034	0.609
41.	24	-0.027	-0.094	0.802
42.	25	-0.065	-0.352	0.477
43.	29	0.045	0.047	0.690
44.	31	-0.140	-0.117	0.458
45.	34	0.020	-0.092	0.394
46.	37	0.710	-0.177	0.700
47.	41	0.002	0.040	0.647
48.	46	-0.320	0.009	0.676
49.	48	0.235	-0.128	0.491
Eigen Values		11.866	8.362	7.079
% Variance		24.217	12.984	10365
Cumulative %		24.217	37.201	47.566

items with .50 or above loading are boldfaced in the corresponding factor.

Table 3: GELS-E: Item subscale correlations for surface acting

Item No.	Pearson r	Sig.
1	0.496**	0.000
6	0.751**	0.000
7	0.470**	0.000
15	0.803**	0.000
16	0.566**	0.000
19	0.640**	0.000
23	0.655**	0.000
27	0.773**	0.000
35	0.705**	0.000
44	0.802**	0.000

**. Correlation is significant at the 0.01 level (2-tailed),*. Correlation is significant at the 0.05 level (2-tailed), n=110.

Table 4: GELS-E: Item subscale correlations for deep acting

Item No.	Pearson r	Sig.
2	0.461**	0.000
18	0.540**	0.000
20	0.567**	0.000
26	0.594**	0.000
32	0.447**	0.000
36	0.511**	0.000
40	0.606**	0.000
43	0.628**	0.000
47	0.615**	0.000
49	0.641**	0.000

**. Correlation is significant at the 0.01 level (2-tailed),*. Correlation is significant at the 0.05 level (2-tailed), n=110.

Table 5: GELS-E: Item subscale correlations for genuine acting

Item No.	Pearson r	Sig.
3	0.667**	0.000
5	0.695**	0.000
14	0.739**	0.000
17	0.791**	0.000
21	0.813**	0.000
24	0.737**	0.000
29	0.756**	0.000
37	0.685**	0.000
41	0.757**	0.000
46	0.694**	0.000

**. Correlation is significant at the 0.01 level (2-tailed),*. Correlation is significant at the 0.05 level (2-tailed), n=110.

Table 6: Internal Consistency of the GELS E

	Cronbach's Alpha	No. of Items
GELS E	0.835	30

Table 7: Cronbach's alpha scores for the subscales of the GELS-E

Subscale	Cronbach's Alpha	No. of Items
Surface Acting	0.847	10
Deep Acting	0.719	10
Genuine Acting	0.882	10

Table 8: Test-retest reliability of the GELS-E

	Pearson r	Sig.
General Emotional Labor Scale English (GELS-E)	0.912	0.000

Table 9: Convergent validity of GELS-E

Scale		Pearson r	Sig.
ELS		0.671**	0.000
ERQ	Reappraisal	0.686**	0.000
	Suppression	0.544**	0.000
DEELS		0.713**	0.000

**. Correlation is significant at the 0.01 level (2-tailed).

The next consideration was to determine the response format. In the case of the GELS, a 5-point Likert response format was deemed appropriate. This is because the five options are allowed for the measurement of not only the occurrence but also for the frequency with which emotional labor occurred. Comrey and Lee (1992) favor multiple choice response format saying that "multiple-choice item formats are more reliable, give more stable results, and produce better scales". Likert type response scale that has five to eight options is appropriate and desirable for questionnaires (Lietz 2010).

Moreover, reverse worded or reverse scored items were not developed for the scale. There has been a certain amount of controversy surrounding whether such items should be included in a questionnaire or not. Negatively worded

Table 10: Discriminant validity of GELS-E

Scale		Pearson r	Sig.
SSEIT		0.003	0.000
PANAS	Positive Affect Score	-0.179**	0.000
	Negative Affect Score	0.011**	0.000
SWLS		0.204**	0.000
CBI		-0.451**	0.000
Ryff's Psychological Well-Being Scale		-0.393**	0.000

**. Correlation is significant at the 0.01 level (2-tailed).

items tend to confuse respondents and hence they decrease the reliability of an instrument. Research shows that when some of these reverse worded items are randomly included in a measure, it has adverse effects on the psychometric properties of the instrument (Harrison and McLaughlin 1991). Empirical studies have been conducted; the findings of which indicate that scales that include positively and negatively worded items contain measurement errors which can lead to problems in the analysis and interpretation of the results (DiStefano and Motl 2006; Quilty et al 2006).

Determining the reliability of a measure is an integral part of scale construction. Reliability refers to how consistent a measure or measurement procedure is, the reliability coefficient or indices are indicative of the reproducibility of the scores of the measure (John and Benet-Martínez 2000). Two different forms of reliability, internal consistency and test-retest reliability were established. Reliability of a scale is an essential aspect as without reliability, validity cannot exist.

Determining the validity of a scale is also a significant aspect of developing a new scale. In fact, developing a valid measure of the chosen construct is probably the first and foremost goal of scale development. Validity refers to the extent to which an instrument actually measures the construct it was developed to measure (Kimberlin and Winterstein 2008). Content validity of the scale was established via subjective validity and face validity by approaching psychologists, language experts and subject matter experts to review and critique the items developed for the General Emotional Labor Scale. Additionally, 14 students further evaluated the scale. According to Cook and Beckman (2006), the content measures the entirety of a construct and nothing else and this can be obtained by wording the items appropriately and by seeking critique by subject matter experts.

Establishing construct validity is essential for developing a sound measure. Construct validity is said to encompass other forms of validity, such as concurrent, predictive, convergent and discriminant validities (Haynes et al 1995). Concurrent validity has been defined as the extent to which the scores obtained from a test agree with the scores obtained from another test that measures the same or similar construct (Zheng and De Jong 2011). Similar to concurrent validity, convergent validity also observes the relationship between similar measures evaluating the same construct. Convergent validity has been defined as the extent to which different measures that are designed to tap the same construct correlate with each other (Cunningham et al 2001). Discriminant validity on the other hand, refers to how distinct and dissimilar different measures and scales are (Guo et al 2008).

Content validity of both versions of the scale was established by asking 14 students, 7 undergraduate level psychology students and 7 graduate level psychology students to read each of the 30 items and then match it with the subscale they understood it to be measuring. The students matched each item to its predetermined subscale, thus showing that the items effectively measured the subscales.

The convergent validity of the General Emotional Scale-English (GELS-E) was assessed against the Emotional Labor Scale by Brotheridge and Lee (2003)

the Emotion Regulation Questionnaire (ERQ) by Gross and John (2003) and the Discrete Emotions Emotional Labor Scale (DEELS) by TM (2004). The correlation between the three pre-established scales and GELS-E was observed in order to determine that GELS-E actually measures emotional labor. Strong positive correlations were observed between GELS-E and the three pre-existing scales, thus establishing convergent validity and concurrent validity of the GELS-E.

Discriminant validity of the General Emotional Labor Scale-English (GELS-E) was measured by correlating the scores of GELS-E against five different pre-existing scales namely; the Schutte Self-report Emotional Intelligence Test (SSEIT) (Schutte et al 1998), the Positive Affect and Negative Affect Schedule (PANAS) by Watson et al (1988), the Satisfaction with Life Scale (SWLS) by Diener et al (1985), the Copenhagen Burnout Inventory (CBI) developed as part of a PUMA study (Kristensen et al 2005), and Ryff's Psychological Well-Being Scale (Ryff and Keyes 1995). The correlation coefficients obtained depicting the relationship between the GELS-E and the above mentioned five scales showed negligible positive correlation for SSEIT and SWLS and negative correlations or inverse relationships for PANAS, CBI and Ryff's Psychological Well-being Scale. These observations depict that GELS-E is distinct and measures a construct separate from the constructs measured by the above stated five instruments.

5 Implications

Developing a scale measuring emotional labor can prove to be quite beneficial for those interested in research as well as for professionals. Researchers studying emotions and the role of emotions in work life can make use of such a scale. The unique quality of the GELS is that it was developed in a third-world Asian country - Pakistan. Considering the dynamics of Pakistan and its demographics, the practice of emotional labor was likely to differ to some extent from western, developed countries. Therefore, the norms established for the scale are pertinent to the said country. However, the scale can be utilized in other countries that have similar demographics, such as India and Bangladesh. Moreover, the scale was developed in a manner to avoid cultural boundaries, so that it can be used world over.

The psychometric properties of GELS were determined in the urban city of Karachi. Karachi as a city consists of a diverse range of demographics. Numerous individuals come to the city in search of work. Therefore, the residents of the current city are multi-ethnic. They belong to different religions, different areas of Pakistan, work in a variety of occupations and the age range is wide. Thus, testing the scale in such a city allowed it's norms to be developed over a range of population.

A recent study by Mastracci et al (2010) brought forth the significance of training in what they call 'emotive skills'. According to them, these emotive skills are imperative for employees working in highly emotional jobs such as law enforcement and disaster services. They suggested that training of such

skills should be part of the teaching curriculum. The GELS-E can prove to be beneficial for conducting need analysis for such trainings and assessment of current performance of emotional labor. This can eventually contribute towards the reduction of the negative impact of emotional labor on employees.

6 Limitations and directions for future research

The current study was not free of limitations. Firstly, although the norms of the GELS-E were established in Pakistan, the respondents were all residents of the metropolitan, urban city of Karachi. Future research ought to focus on assessing the psychometric properties of the scale in different cities and rural areas of Pakistan (and other countries).

7 Conclusion

The current study set out to create a reliable, valid and sound measure of emotional labor. The idea was to develop an instrument that has its norms established in the developing, Asian country of Pakistan, where religion, traditions and customs play a major role in an individuals' work life and their expression of emotions. Psychometric properties of the scale were rigorously checked so that the scale assessed emotional labor in a thorough manner.

Three subscales were determined as part of the GELS-E which are Surface, Deep and Genuine Acting. These three subscales relate to the method and expression of emotions at work to meet organizational demands. Furthermore, the three facets of each subscale enable in-depth analysis of the mode and expression of surface, deep and genuine acting and the intensity and variety of emotions expressed.

References

Ahmed A, Ahsan H (2011) Contribution of services sector in the economy of pakistan. Working Papers & Research Reports 2011

Ashforth BE, Humphrey RH (1993) Emotional labor in service roles: The influence of identity. Academy of Management Review 18(1):88–115

Ashkanasy NM, Daus CS (2002) Emotion in the workplace: The new challenge for managers. Academy of Management Perspectives 16(1):76–86

Brotheridge CM, Lee RT (2003) Development and validation of the emotional labour scale. Journal of Occupational and Organizational Psychology 76(3):365–379

Clark LA, Watson D (2016) Constructing validity: Basic issues in objective scale development.

Comrey A, Lee H (1992) A first course in factor analysis (2e éd.). hillsdale, é. U: Erlbaum

Cook DA, Beckman TJ (2006) Current concepts in validity and reliability for psychometric instruments: theory and application. The American journal of medicine 119(2):166–e7

Cukur CS (2009) The development of the teacher emotional labor scale (tels): Validity and reliability. Educational Sciences: Theory and Practice 9(2):559–574

Cunningham WA, Preacher KJ, Banaji MR (2001) Implicit attitude measures: Consistency, stability, and convergent validity. Psychological science 12(2):163–170

Diener E, Emmons R, Larsen R, Griffin S (1985) The life satisfaction scale. Journal of personality Assessment 49(1):71–75

DiStefano C, Motl RW (2006) Further investigating method effects associated with negatively
worded items on self-report surveys. Structural Equation Modeling 13(3):440–464

Floyd FJ, Widaman KF (1995) Factor analysis in the development and refinement of clinical
assessment instruments. Psychological assessment 7(3):286

Grandey AA (2000) Emotional regulation in the workplace: A new way to conceptualize
emotional labor. Journal of occupational health psychology 5(1):95

Gross JJ, John OP (2003) Individual differences in two emotion regulation processes: implica-
tions for affect, relationships, and well-being. Journal of personality and social psychology
85(2):348

Guo B, Aveyard P, Fielding A, Sutton S (2008) Testing the convergent and discriminant
validity of the decisional balance scale of the transtheoretical model using the multi-trait
multi-method approach. Psychology of Addictive Behaviors 22(2):288–294

Harrison DA, McLaughlin ME (1991) Exploring the cognitive processes underlying responses
to self-report instruments: Effects of item context on work attitude measures. In: Academy
of Management Proceedings, Academy of Management Briarcliff Manor, NY 10510, vol
1991, pp 310–314

Haynes SN, Richard D, Kubany ES (1995) Content validity in psychological assessment: A
functional approach to concepts and methods. Psychological assessment 7(3):238–247

Hochschild AR (1983) The managed heart. berkeley

Hochschild AR (2003) The managed heart: Commercialization of human feeling. Univ of
California Press

John OP, Benet-Martínez V (2000) Measurement: Reliability, construct validation, and scale
construction. n H T Reis, & C Judd, Handbook of Research Methods in Social and
Personality Psychology (pp 339-369) New York: Cambridge University Press

Kimberlin CL, Winterstein AG (2008) Validity and reliability of measurement instruments
used in research. American journal of health-system pharmacy 65(23):2276–2284

Kristensen TS, Borritz M, Villadsen E, Christensen KB (2005) The copenhagen burnout
inventory: A new tool for the assessment of burnout. Work & Stress 19(3):192–207

Lietz P (2010) Research into questionnaire design: A summary of the literature. International
journal of market research 52(2):249–272

Mastracci SH, Newman MA, Guy ME (2010) Emotional labor: Why and how to teach it.
Journal of Public Affairs Education 16(2):123–141

Miller KI, Considine J, Garner J (2007) let me tell you about my job: Exploring the terrain
of emotion in the workplace. Management Communication Quarterly 20(3):231–260

Morris JA, Feldman DC (1996) The dimensions, antecedents, and consequences of emotional
labor. Academy of management review 21(4):986–1010

Quilty LC, Oakman JM, Risko E (2006) Correlates of the rosenberg self-esteem scale method
effects. Structural Equation Modeling 13(1):99–117

Ryff CD, Keyes CLM (1995) The structure of psychological well-being revisited. Journal of
personality and social psychology 69(4):719–727

Schutte NS, Malouff JM, Hall LE, Haggerty DJ, Cooper JT, Golden CJ, Dornheim L (1998)
Development and validation of a measure of emotional intelligence. Personality and indi-
vidual differences 25(2):167–177

Taqui AM, Itrat A, Qidwai W, Qadri Z (2007) Depression in the elderly: Does family system
play a role? a cross-sectional study. BMC psychiatry 7(1):57

TM G (2004) Tews mj. Emotional labor: A conceptualization and scale development J Voc
Behs 64(1):1–23

Watson D, Clark LA, Tellegen A (1988) Development and validation of brief measures of
positive and negative affect: the panas scales. Journal of personality and social psychology
54(6):1063

Zheng Y, De Jong J (2011) Research note: Establishing construct and concurrent validity of
pearson test of english academic. Pearson Education Ltd Retrieved April 15:2019