

# Emotional Awareness: Computer and Hand Scoring of an Open-Ended Test

Kimberly A. Barchard<sup>a</sup>, Richard D. Lane<sup>b</sup>, and Bryan D. Watson<sup>a</sup>

a University of Nevada, Las Vegas

b University of Arizona

**Reference:** Barchard, K.A., Lane, R.D. & Watson, B.D. (2010, Aug). *Emotional Awareness: Computer and Hand Scoring of an Open-Ended Test*. Poster presented at the 2010 American Psychological Association convention, San Diego, CA.

**Contact Information:** Kimberly A. Barchard, Department of Psychology, University of Nevada, Las Vegas, 4505 S. Maryland Parkway, P.O. Box 455030, Las Vegas, NV, 89154-5030, USA, barchard@unlv.nevada.edu

## ABSTRACT

Open-ended questions may be the most appropriate method of measuring Emotional Awareness, but they can be difficult and time-consuming to score. The purpose of this paper was to examine the effectiveness of hand scoring and computer scoring of the Levels of Emotional Awareness Scale (Lane, 1991). The first study included 268 undergraduates and the second study included 200 community members. In both studies, the various scoring methods demonstrated internal consistency reliability; convergent validity with tests of emotional intelligence, social intelligence, and perception of affect; and discriminant validity with respect to vocabulary and response length. The most effective scoring method was the Highest40-AllinOne technique, which calculates the sum of the 40 words and phrases with the highest values, across all 20 items, after duplicate words and phrases have been eliminated. This technique seems to capture a new level of emotional awareness, beyond the five levels given in the original Lane, Quinlan, Schwartz, Walker, and Zeitlin (1990) model: the ability to make fine discriminations between similar emotions. This scoring technique may not be as valuable in groups with lower levels of emotional awareness, such as children or clinical populations. Future research should determine when each scoring method is most appropriate to capture each level of emotional awareness.

## INTRODUCTION

Emotional Awareness is central to healthy psychological functioning. For example, Emotional Awareness is lower in people with somatoform disorders (Subic-Wrana et al., 2002), depression (Berthoz et al., 2000), and eating disorders (Bydlowski et al., 2005), and people with higher Emotional Awareness have a greater sense of well-being (Ciarrochi et al., 2003). However, measuring Emotional Awareness is difficult. Self-report (e.g., Bagby et al., 1994) may be inappropriate because people with low awareness might not recognize their deficiencies (Lane, Sechrest, Reidel, Weldon, Kasniak, & Schwartz, 1996), and one-on-one interviews (e.g., Bagby et al., 1994; Sifneos, 1973) are time-consuming and require trained interviewers and scorers. Closed-ended written questions can measure related constructs like Emotional Intelligence (e.g., Mayer et al., 2002), but are likely inappropriate for measuring Emotional Awareness: If someone reports being depressed, how can we score them as correct or incorrect? The ideal measurement may be written open-ended questions that are scored based upon structure, not the specific content of the response. The Levels of Emotional Awareness scale (LEAS; Lane et al., 1990) uses precisely this strategy. Participants describe how they would feel in 20 emotionally evocative situations, and their responses are scored based upon the complexity and variety of emotion words they use (Lane, 1991). However, scoring the LEAS by hand is time-consuming and difficult, and some scorers do not obtain adequate inter-rater reliability (Barchard, 2009). The purpose of this research was to compare the reliability and validity of nine computerized scoring methods with the reliability and validity of hand scoring.

Program for Open-Ended Scoring (POES; Leaf & Barchard, 2009) includes several scoring methods, all loosely based upon hand scoring. In hand scoring, words are scored based upon emotional complexity. A vague description of an emotion (I feel bad) receives a lower score than a detailed description (I feel angry). The total score is based on the complexity and variety of the emotion words used. If someone uses synonyms (I feel mad and angry), they receive a lower score than if they describe distinct emotions (I feel angry and guilty). Although all POES methods are based loosely upon hand scoring, they vary in their complexity. Some methods mimic hand scoring as closely as possible. Others combine word-level scores in simpler ways. For example, one method simply adds together all word-level scores. POES is distributed under the GNU General Public License (Free Software Foundation, 1992) and therefore can be modified by others.

## STUDY 1

### Method

#### Participants

A total of 268 undergraduates (154 female, 114 male) participated in return for course credit. They ranged in age from 18 to 50 (mean 20.02, SD 3.59). They identified themselves as follows: 58.4% Caucasian, 12.7% Hispanic, 11.6% Asian, 8.6% African American, 5.2% Pacific Islander, and 3.4% other.

Participants completed the study on the Internet. They were asked to complete the study in university computer labs, but may have completed it from other locations.

#### Measures

The Levels of Emotional Awareness Scale (LEAS; Lane, Quinlan, Schwartz, Walker, & Zeitlin, 1990) contains 20 open-ended items. Each item presents a brief scenario involving the self and other person, followed by two questions: "How would you feel?" and "How would the other person feel?" The LEAS was scored by hand and using Program for Open-Ended Scoring (POES; Leaf & Barchard, 2010) version 1.4.1 with LEAS Wordlist 2.4 (Barchard, 2010), which specifies the scores given for specific words and phrases.

To assess discriminant validity, we counted how many words participants used across their 20 LEAS responses, and had participants complete a 60-item vocabulary test (Barchard, 2004a). To assess convergent validity, participants completed the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT; Mayer, Salovey, & Caruso 2002), which includes four branches (Perceiving, Facilitating, Understanding, Managing); the four O'Sullivan and Guilford (1976) Social Intelligence tests (Expression Grouping, Missing Cartoons, Cartoon Predictions, Social Translations); and the Metaphors Test (Barchard, 2004b). All measures were computer administered.

## Results

All LEAS scoring methods had acceptable internal consistencies (see Table 2). The correlations with the four branches of the MSCEIT and with the four O'Sullivan and Guilford Tests of Social Intelligence were usually small, positive, and significant (see Table 3). The scoring method with the highest correlations was the Highest40-AllinOne technique, which is the sum of the scores from the 40 words with the highest scores across the entire set of 20 items, after duplicates have been removed. The correlations with Vocabulary were small, positive, and significant (see Table 3),

indicating that the LEAS is not simply a Vocabulary test. For some POES methods, correlations with Response Length were high (see Table 3), however, almost all of the LEAS scores retained their significant correlations with the MSCEIT and the O'Sullivan and Guilford tests, once Response Length had been partialled out (see Table 4).

Table 1  
*LEAS Scoring Methods*

Method	Calculation	Example Scores <sup>a</sup>
<i>Gives Credit when Same Emotion Word is Used in Multiple Items</i>		
Hand Scoring	Item Score is based upon Self and Other scores, as described above	Item 1: 4 Item 2: 3
AllSum	Item Score is the sum of all Values in the Valuables List	Item 1: 9 Item 2: 7 Total: 16
Highest-4	Item Score is the sum of the four highest Values in the Valuables List	Item 1: 9 Item 2: 7 Total: 16
334	Item Score is the maximum Value in the Valuables List, unless two non-identical Valuables with a Value of 3 are present, in which case the Item Score is 4.	Item 1: 4 Item 2: 3 Total: 7
3345	Self and other Valuables Lists are scored separately using the 334 method. Item Score is 5 if the self and other scores are both 4; otherwise, Item Score is the maximum of the self and other scores. This method requires separate responses for self and other.	Item 1 Self: 4 Item 1 Other: 3 Item 1: 4 Item 2 Self: 3 Item 2 Other: 2 Item 2: 3 Total: 7
<i>Only Gives Credit the First Time an Emotion Word is Used</i>		
AllSum-AllinOne	A combined Valuables List is created using the responses to all items. The Total Score is the sum of all Values for all non-identical Valuables in the Valuables List.	Total: 10
HighestN-AllinOne	A combined Valuables List is created using the responses to all items. The Total Score is the sum of the highest <i>N</i> Values for non-identical Valuables in the Valuables List. <i>N</i> can be set to any number. In this study, <i>N</i> = 10, 20, 30, 40, 50, 60 were used.	<i>N</i> = 2 Total: 6 <i>N</i> = 5 Total: 10

a. These are the scores for the responses: Item 1: "I would be happy he likes me. He would also be happy." Item 2: "I do not think I would be happy. He would feel bad. He might cry." The Valuables Lists are as follows: Item 1: happy 3, likes 3, happy 3. Item 2: happy 3, bad 2, cry 2. Combined: happy 3, likes 3, happy 3, happy 3, bad 2, cry 2.

Table 2  
*Internal Consistencies of Scoring Methods that Calculate Item-Level Scores, Study 1*

Scoring Method	Coefficient Alpha
Hand Scoring	.84
All-Sum	.95
Highest-4	.91
334	.77
3345	.85

Table 3  
*Correlations of the LEAS with Criterion Variables, Study 1*

Scoring Method	MSCEIT				O'Sullivan and Guilford Test				Meta.	Vocab.	Length
	Perc.	Using	Und.	Man.	EG	CP	MC	ST			
<i>Gives credit each time an emotion word is used</i>											
Hand Scoring	.07	.13*	.25**	.14*	.15*	.10	.23**	.13*	.18**	.15*	.48**
AllSum	.17**	.16**	.30**+	.23**+	.15*	.21**	.24**	.16**	.27**	.24**+	.86**
Highest-4	.25**+	.22**+	.39**+	.31**+	.23**+	.27**+	.31**+	.20**	.32**	.26**+	.78**
334	.20**	.18**	.30**	.25**	.22**	.27**	.32**+	.19**	.28**	.23**	.59**
3345	.20**	.18**	.35**+	.26**+	.21**	.26**	.31**+	.19**	.29**	.26**	.71**
<i>Gives credit only the first time an emotion word is used</i>											
AllSum-AllinOne	.20**	.20**+	.39**+	.30**+	.23**+	.24**	.34**+	.26**+	.29**	.32**+	.80**
Highest80-AllinOne	.19**	.20**+	.39**+	.30**+	.23**+	.24**	.34**+	.26**+	.29**	.32**+	.80**
Highest60-AllinOne	.19**	.20**+	.40**+	.30**+	.24**+	.23**	.33**+	.26**+	.30**	.31**+	.78**
Highest40-AllinOne	.21**	.23**+	.42**+	.32**+	.27**+	.27**+	.34**+	.24**+	.30**	.31**+	.68**
Highest20-AllinOne	.12*	.20**	.33**	.27**	.23**	.24**	.28**	.19**	.27**	.27**	.42**
Response Length	.14*	.11	.22**	.17**	.13*	.17**	.20**	.13*	.25**	.18**	

Note. Perc. = Perceiving Emotions. Using = Using Emotions. Und. = Understanding Emotions. Man. = Managing Emotions. EG = Expression Grouping. CP = Cartoon Predictions. MC = Missing Cartoons. ST = Social Translations. Meta. = Metaphors Test. Vocab. = Vocabulary Test. Length = Response Length.

\*  $p < .05$ . \*\*  $p < .001$ . + indicates that this correlation is significantly larger ( $p < .05$ ) than the correlation between this criterion variable and response length, which is given in the last line of this table.

Table 4  
*Partial Correlations of the LEAS with Criterion Variables, Controlling for Response Length, Study 1*

Scoring Method	MSCEIT				O'Sullivan and Guilford Test				Meta.	Vocab.
	Perc.	Using	Und.	Man.	EG	CP	MC	ST		
<i>Gives credit each time an emotion word is used</i>										
Hand Scoring	.00	.08	.16*	.06	.10	.01	.16*	.07	.06	.08
AllSum	.10	.12	.21**	.17*	.09	.13*	.13*	.09	.11	.18*
Highest-4	.22**	.21**	.35**	.30**	.22**	.22**	.25**	.15*	.20**	.20*
334	.15*	.14*	.21**	.19*	.18*	.21**	.26**	.14*	.16*	.16*
3345	.14*	.14*	.27**	.20*	.18*	.20*	.24**	.13*	.16*	.19*
<i>Gives credit only the first time an emotion word is used</i>										
AllSum-AllinOne	.15*	.18*	.36**	.27**	.22**	.17*	.30**	.26**	.16*	.31**
Highest80-AllinOne	.14*	.18*	.36**	.27**	.22**	.17*	.30**	.26**	.16*	.31**
Highest60-AllinOne	.14*	.18*	.37**	.28**	.23**	.16*	.29**	.25**	.17*	.28**
Highest40-AllinOne	.17*	.20*	.37**	.27**	.25**	.21**	.29**	.21**	.18*	.26**
Highest20-AllinOne	.07	.17*	.27**	.22**	.20**	.18*	.22**	.15*	.18*	.21**

Note. Perc. = Perceiving Emotions. Using = Using Emotions. Und. = Understanding Emotions. Man. = Managing Emotions. EG = Expression Grouping. CP = Cartoon Predictions. MC = Missing Cartoons. ST = Social Translations. Meta. = Metaphors Test. Vocab. = Vocabulary Test. Length = Response Length.

\*  $p < .05$ . \*\*  $p < .001$ .

## STUDY 2 Method

### Participants

A total of 200 adult community members (103 females, 95 males, 2 did not answer) were each paid \$10 to participate in this study. They were recruited through posted announcements, invitations sent to church and community groups, and word of mouth. Participants ranged in age from 18 to 85 (mean 51.9, SD 18.7). Most participants identified themselves as Caucasian (87.5%), Hispanic (6%), or Black (2.5%). Participants were screened to ensure that they were native English speakers, and that they had no history of serious psychiatric disorders, substance abuse, or cognitive impairment. Participants completed this study in a quiet setting with an examiner present.

### Measures

To assess discriminant validity, we counted how many words participants used across their 20 LEAS responses. To assess convergent validity, participants completed the Perception of Affect Task (Rau, 1993). All measures were completed on paper.

### Results

All LEAS scoring methods had adequate internal consistency (see Table 5). All methods had significant moderate to high correlations with the Perception of Affect Task (see Table 6). The scoring methods with the highest correlations were the Highest20-AllinOne and the Highest40-AllinOne techniques. Once again, some LEAS scores had high correlations with Response Length, but in this study, all retained their significant correlations with the Perception of Affect Task when we controlled for Response Length (see Table 6).

Table 5  
*Internal Consistencies of Scoring Methods that Calculate Item-Level Scores, Study 2*

Scoring Method	Coefficient Alpha
Hand Scoring	.88
All-Sum	.95
Highest-4	.93
334	.85

Table 6  
*Correlations of LEAS with Perception of Affect Task and Response Length, Study 2*

Scoring Method	Correlation with PAT	Correlation with Response Length	Partial Correlation with PAT, Controlling for Response Length
<i>Gives credit each time an emotion word is used</i>			
Hand Scoring	.41**	.49**	.32**
All-Sum	.37**	.79**	.22**
Highest-4	.43**	.67**	.32**
334	.42**	.48**	.33**
<i>Gives credit only the first time an emotion word is used</i>			
AllSum-AllinOne	.46**	.78**	.37**
Highest80-AllinOne	.46**	.76**	.38**
Highest60-AllinOne	.48**	.72**	.39**
Highest40-AllinOne	.52**	.62**	.44**
Highest20-AllinOne	.50**	.39**	.44**
Response Length	.30**		

\*\*  $p < .001$ .

## DISCUSSION

Both studies provide evidence for the internal consistency, convergent validity, and discriminant validity of the various LEAS scoring methods. This success provides evidence for the validity of the fundamental idea behind the LEAS: the use of some words indicates more Emotional Awareness than other words.

The ten LEAS scoring methods incorporate two radically different strategies for calculating a total score based upon the item scores. The first strategy is based upon the Lane et al. (1990) developmental model of Emotional Awareness, in which higher scores for each item are assigned if a respondent indicates that the self and other will have complex emotional reactions that are different from each other. This strategy was used in hand-scoring, and to greater or lesser extents in the Highest-4, the 334 method, and the 3345 method. The usefulness of this scoring strategy was demonstrated by Barchard, Bajgar, Leaf, and Lane (2010). The second strategy is to focus upon the breadth of emotion words used across all items, by only counting each word or phrase the first time it is used. This strategy will usually give higher scores to respondents who say that the self and other will have different emotional reactions, but will also give higher scores to people who say that they will have different emotional reactions in different situations. The balance of the evidence from these two studies suggest that this second strategy is more powerful when assessing non-psychiatric samples of adults, even though it does not distinguish between emotions attributed to the self and other in any particular situation.

The usefulness of the HighestN-AllinOne method should be explored in other populations. Eliminating duplicates is likely to be beneficial in groups that have average or high levels of Emotional Awareness, but may not be helpful in groups that have lower levels of Emotional Awareness. In particular, researchers should not assume this method will be valid when assessing children and clinical groups. For those groups, it is likely that higher validities will be obtained by staying closer to the original hierarchical model of Emotional Awareness proposed by Lane et al. (1990). In that model, high scores are obtained by acknowledging that a person can feel more than one emotion at a time, and by distinguishing between the emotions of the self and other. This current paper suggests there may be another level of Emotional Awareness, beyond the ones proposed in the original model. This is the ability to make fine distinctions between similar emotional states, such as angry, furious, and annoyed. The HighestN-AllinOne method may be capturing this new level of Emotional Awareness.

## REFERENCES

- Bagby, R. M., Taylor, G. J., & Parker, J. D. A. (1994). The twenty-item Toronto Alexithymia Scale. II. Convergent, discriminant, and concurrent validity. *Journal of Psychosomatic Research*, 38, 33-40. doi:10.1016/0022-3999(94)90006-X
- Barchard, K.A. (2004a). *Las Vegas Vocabulary Test*. Unpublished test. Available from Kim Barchard, Department of Psychology, University of Nevada, Las Vegas, 4505 Maryland Parkway, Las Vegas, NV, 89154-5030, barchard@unlv.nevada.edu
- Barchard, K.A. (2004b). *Metaphors Test*. Unpublished Maximum-performance test of Emotional Intelligence. Available from Kim Barchard, Department of Psychology, University of Nevada, Las Vegas, 4505 Maryland Parkway, Las Vegas, NV, 89154-5030, barchard@unlv.nevada.edu
- Barchard, K.A. (2009, May). *Evaluating Score Consistency through Score Change Probabilities*. Paper presented at the Association for Psychological Science convention in San Francisco, CA.
- Barchard, K.A. (2010, Feb). *LEAS Wordlist 2.4*. File to be used with POES to allow scoring of the Levels of Emotional Awareness Scale.
- Barchard, K.A., Bajgar, J., Leaf, D.E., & Lane, R. (2010). Computer scoring of the Levels of Emotional Awareness Scale. *Behavior Research Methods*, 42(2), 586-595. doi:10.3758/BRM.42.2.586
- Berthoz, S., Ouhayoun, B., Parage, N., Kirzenbaum M., Bourgey M., & Allilaire J.F. (2000). Étude préliminaire des niveaux de conscience émotionnelle chez des patients déprimés et des contrôles. = Preliminary study of the levels of emotional awareness in depressed patients and controls. *Annales Medico-Psychologiques*, 158, 665-672.
- Bydlowski, S., Corcos, M., Jeammet, P., Paterniti, S., Berthoz, S., Laurier, C., Chambry, J., & Consoli, S.M. (2005). Emotion-processing deficits in eating disorders. *International Journal of Eating Disorders*, 37, 321-329. doi:10.1002/eat.20132
- Ciarrochi, J., Caputi, P., & Mayer, J.D. (2003). The distinctiveness and utility of a measure of trait emotional awareness. *Personality and Individual Differences*, 34, 1477-1490. doi:10.1016/S0191-8869(02)00129-0
- Free Software Foundation. (1991). *GNU General Public License (version 2)*. Retrieved from <http://www.gnu.org/licenses/old-licenses/gpl-2.0.html>
- Lane, R.D. (1991). *LEAS Scoring Manual and Glossary*. Unpublished manual for the Levels of Emotional Awareness Test. Available from Richard D. Lane, General Clinical Research Center, University of Arizona, PO Box 245002, Tucson, AZ 85724-5002.
- Lane, R. D., Quinlan, D. M., Schwartz, G. E., Walker, P. A., & Zeitlin, S. B. (1990). The Levels of Emotional Awareness Scale: A cognitive-developmental measure of emotion. *Journal of Personality Assessment*, 55, 124-134. doi:10.1207/s15327752jpa5501&2\_12
- Lane, R.D., Sechrest, L., Reidel, R., Weldon, B.S., Kasniak, A., & Schwartz, G.E. (1996). Impaired verbal and nonverbal emotion recognition in alexithymia. *Psychosomatic Medicine*, 58, 203-210.
- Leaf, D.E. & Barchard, K.A. (2010, Jan). *Program for Open-Ended Scoring [POES] version 1.4.1*. Windows-based program that scores open-ended tests according to the criteria given in the selected Wordlist. Available from Kim Barchard, Department of Psychology, University of Nevada, Las Vegas, 4505 Maryland Parkway, Las Vegas, NV, 89154-5030, barchard@unlv.nevada.edu
- Mayer, J.D., Salovey, P., & Caruso, D.R. (2002). *Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) User's Manual*. North Tonawanda, NY: Multi-Health Systems.
- O'Sullivan, M., & Guilford, J. P. (1976). *Four factor tests of social intelligence (behavioral cognition): Manual of instructions and interpretations*. Orange, CA: Sheridan Psychological Services.
- Rau, J.C. (1993). Perception of verbal and nonverbal affective stimuli in complex partial seizure disorder. *Dissertation Abstracts International B* 54, 506B.
- Sifneos, P.E. (1973). The prevalence of "alexithymia" characteristics in psychosomatic patients. *Psychotherapy & Psychosomatics*, 22, 255-262.
- Subic-Wrana, C., Bruder, S., Thomas, W., Gaus, E., Merkle, W., & Köhle, K (2002). Verteilung des Persönlichkeitsmerkmals Alexithymie bei Patienten in stationärer psychosomatischer Behandlung--gemessen mit TAS-20 und LEAS. = Distribution of alexithymia as a personality-trait in psychosomatically ill in-patients--measured with TAS-20 and LEAS. *Psychotherapie Psychosomatik Medizinische Psychologie*, 52, 454-460.