Human vs. Computer: Scoring a Test of Emotional Awareness

Kimberly A. Barchard[a], Dalton J. Terry[a], and Duncan Ermini Leaf[b]
a. University of Nevada, Las Vegas
b. Purdue University

UNLV

Reference: Barchard, K.A., Terry, D.J., & Ermini Leaf, D. (2014, April). *Human vs. computer: Scoring a test of emotional awareness.* Poster presented at the Western Psychological Association annual convention in Portland, OR.
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

Emotional intelligence is the ability to perceive, understand, use, and mange emotions (Mayer et al, 2003). One key aspect of emotional intelligence is emotional awareness: the ability to describe feelings in oneself and others (Lane & Schwartz, 1987). The Levels of Emotional Awareness Scale (LEAS; Lane et al., 1990) includes 20 open-ended items. For each, participants read a short scenario that involves themselves and another person. Then they describe how they and the other person would feel. These written responses are scored based upon the type and number of emotion words used.

Traditionally, the LEAS is scored by hand. However, hand scoring is time-consuming. Therefore, Program for Open-Ended Scoring (POES) was created to score the LEAS (Leaf, 2003). The purpose of this study was to compare the validity of hand scoring and the eight POES scoring methods, by correlating these scores with the 141-item Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT; Mayer et al., 2003). A total of 356 undergraduates completed the LEAS and MSCEIT online.

All correlations were significant, thus supporting the validity of the LEAS. HighestN and HighestNUnique had the highest correlations, making these methods a viable alternative to hand scoring. The new AllSumUnique method had a higher correlation than the AllSum method, and the new HighestNUnique method had a higher correlation than the HighestN method, thus indicating that duplicate words add little to our knowledge of someone's emotional awareness. The new CountCat method counts the number of distinct categories used. It deserves further exploration, perhaps with revised categories that more precisely examine the richness of one's emotional descriptions.

Introduction

Emotional intelligence is the ability to perceive, understand, and manage emotions and to use emotions to facilitate thinking (Mayer, Salovey, & Caruso, 2004). One key aspect of emotional intelligence is emotional awareness. Emotional awareness is a cognitive ability that allows individuals to identify and describe feelings in themselves and others (Lane & Schwartz, 1987). Emotional awareness is important because recognizing emotion and understanding emotion is something that we do in our day to day functioning. Moreover, emotional awareness may be a prerequisite to learning higher order emotional skills.

The Levels of Emotional Awareness Scale (LEAS; Lane, Quinlan, Schwartz, Walker, & Zeitlin, 1990) is an open-ended measure of emotional awareness. The LEAS contains 20 emotionally evocative scenarios that involve the self and another person. For each scenario, participants describe how they would feel and how the other person would feel. Participants' scores on the LEAS depend upon the emotion words they use. Participants get higher scores when they use a larger number of emotion words and more specific emotion words.

The LEAS can be scored by hand or by computer. Scoring a LEAS response by hand requires expert subjective judgments based upon the meaning of words, the context in which they appear, and the presence of synonyms. Learning to make these subjective judgments takes approximately 10 hours (Barchard, Bajgar, Leaf, & Lane, 2010). Once a person is fully trained, doing the scoring itself is also time-consuming. Even the most experienced hand scorer will take about 10 minutes to score each participant (Barchard et al., 2010). Thus, scoring 100 participants (a reasonable sample size for most psychological studies) would take about 17 hours.

Computer scoring of the LEAS is also possible. Automated scoring was suggested when the LEAS was first developed in 1990 (Lane, Quinlan, Schwartz, Walker, & Zeitlin, 1990) but it wasn't until 2003 that a computer program was developed to score the LEAS automatically (Leaf, 2003). Subsequent research has demonstrated that the initial Program Open-Ended Scoring (POES) scoring methods had strong correlations with hand scoring, high internal consistency, and validity that was comparable to or higher than the validity of hand scoring (Barchard et al., 2010). Recently, a new version of POES has been developed (Ermini Leaf & Barchard, 2013). The AllSum and HighestN techniques remain the same as in the previous version of POES. However, the remaining scoring methods have been revised to be more flexible. First, the AllSum-AllinOne and HighestN-AllinOne methods were replaced by the AllSumUnique and HighestNUnique methods, which allow scores to be calculated at the subpart and item level, a feature that was not available in past versions. Second, the 334 and 3345 methods have been replaced with the UniqueMaximums method. Finrally, POES 2.0.1 incorporates three new scoring methods (CountWords, CountCat, and CountFreq). The

purpose of this study is to compare the validity of hand scoring and the eight POES 2.0.1 scoring methods of the LEAS. To do so, we correlated the LEAS scoring methods with the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT; Mayer, Salovey, Caruso & Sitarenios, 2003).

Method

Participants

A total of 356 university students (204 females, 152 males) participated in this study in return for course credit. They ranged from 18 to 50 years (mean 19.95, SD 3.53). Participants identified themselves as follows: Caucasian (57.9%), Hispanic (12.4%), Asian (12.4%), African American (7.0%), Pacific Islander (6.2%), Native American (0.3%), and other (3.9%).

Measures

The Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT; Mayer, Salovey, Caruso & Sitarenios, 2003) is a 141-item test.

The Levels of Emotional Awareness Scale (LEAS; Lane et al. 1990) consists of 20 items that present emotionally-evocative scenarios. Each item has two subparts. The first subpart is the answer to the question, "How would you feel?" The second subpart is the answer to the question, "How would the other person feel?" The LEAS was scored both by hand or by POES computer scoring.

Hand scoring the LEAS

The LEAS hand scoring manual (Lane, 1991) contains a glossary of emotion words that are scored from 0 to 3 based on the complexity of the emotion word. LEAS scorers first search the two subparts of the item for words and phrases that indicate emotions. Words that are indicative of a thought rather than a feeling (e.g., "confused") receive a score of 0. Words that describe a physical sensation (e.g., "dizzy") receive a score of 1. Words that include positive or negative connotations, personality traits, or actions that are part of an emotional response (e.g., "rude," "honorable," "grudge," respectively) receive a score of 2. Finally, words that identify a specific emotion (e.g., "happy") receive a score of 3.

Next, the scorer determines whether each emotion word was attributed to the self or to the other person in the scenario and calculates the Self and Other scores as the maximum of the values of words and phrases attributed to that person. When a participant gives a response for either Self or Other with two separate, non-synonymous level 3 emotion words, the Self or Other score is raised to 4.

Third, the scorer calculates the item score as the maximum of the Self and Other scores. When both Self and Other scores are 4 and the responses are not the same as each other, the item score is raised to 5.

To calculate total scores across all items, the scores for the 20 items are summed. Because there are 20 items, each with a maximum score of 5, the highest possible total score on the LEAS is 100.

POES Computer Scoring the LEAS

Computer scoring of the LEAS was done using POES 2.0.1 (Ermini Leaf & Barchard, 2013). POES 2.0.1 uses eight scoring methods. Most of the POES scoring methods require the use of a wordlist. A wordlist contains key words and phrases and the values that are assigned to each of those keys. For example, the key word "happy" receives a value of 3. This study used LEAS Wordlist 2.5 (Barchard, 2013), which was based upon the LEAS hand-scoring glossary, but has been adapted for computer scoring.

CountWords. The CountWords method counts the number of words in a response. This score can be calculated at the level of the subpart, the item, or the subject. For this study, we used subject-level scores.

AllSum. The AllSum method calculates the sum of all key values. First, POES searches each response for the key words and phrases that appear in the Wordlist. It then notes the values of each of those keys. To calculate the AllSum score for a particular response, POES sums the values of the keys in that response. POES calculates AllSum scores for each subpart, item, and subject. In this study, we used subject-level scores.

AllSumUnique. The AllSumUnique method uses the same scoring technique as the AllSum method but with one major distinction. This method eliminates repetitions from being scored twice. For example, imagine that a response says, "I would feel happy. I would feel so happy I would cry." To score this response, POES creates a *unique keys list*. The unique keys list contains each key that appears in the response, but it only contains each key once. In the example above, the unique keys list would contain the keys "happy" and "cry". The unique keys list also contains the values of each of these keys. Once the unique keys list is compiled, the AllSumUnique score is calculated as the sum of the values for each key that occurs on that list. AllSumUnique can be calculated at the level of subparts, items, or subjects. For this study, we used the subject-level scores, so that if a subject used the same word in more than one response, they only received credit once.

HighestN. The HighestN score is calculated as the sum of the highest N values of the keys that appear in a response. HighestN scores can be calculated at the subpart, item, or subject level. The user can specify the value of N. When N is set to 4 (the default value), the item-level scores are identical to the Highest-4 method in previous versions of POES (Leaf & Barchard, 2010). In this study, we used subject-level scores with N equal to 40 to reduce the influence of long responses on the scores.

HighestNUnique. To score a response using the HighestNUnique method, POES first creates a unique keys list for that response. Then the HighestNUnique score is calculated as the sum of the highest N values of the keys on that unique keys list. HighestNUnique scores can be calculated at the subpart, item, or subject level. When scores are calculated at the subject level and N is set to 40, HighestNUnique scores are identical to the Highest40-AllinOne scores in previous versions of POES (Leaf & Barchard, 2010); That method had the highest validity of all scoring methods examined (Alsaid-Habia, Tomme, Ladrazo, Craun, & Barchard, 2012). Therefore, for this study, we used subject-level scores with N equal to 40, so that if a subject used the same word in more than one response, they only received credit once.

UniqueMaximums. The UniqueMaximums method requires the user to set maximum values for each level of scoring. The user sets values for KeyMax, SubpartMax, ItemMax, and SubjectMax. The scoring starts at the subpart level. For each subpart, an

initial subpart score is calculate as being equal to KeyMax or the highest key value in the response, whichever is lower. If there are multiple distinct keys that have the highest possible value, then the subpart score is incremented by one for each additional key. For example, if the response said, "I would be happy, jealous, and angry, all at the same time" and KeyMax was set to 3, then the initial subpart score would be 5 because "happy," "jealous", and "angry" each have a key score of 3. The final subpart score is calculated as the minimum of this initial subpart score and SubpartMax. For example, if SubpartMax was set to 4, then the subpart score would be 4.

Next, the item score is calculated. The initial item score is calculated as ItemMax or the highest subpart score, whichever is lower. This initial item score is incremented by one for each subpart that has the maximum score. For example, if the score for subpart 1 was 4 and the score for subpart 2 was 4, then the initial item score is 5. The final item score is calculated as the minimum of this initial item score and ItemMax.

Finally, the subject score is calculated. The initial subject score is calculated as SubjectMax or the highest item score, whichever is lower. This initial subject score is incremented by one for each item that has the maximum score. For example, if the score for item 1 was 5 and the score for item 2 was 5, then the initial subject score is 6. The final subject score is calculated as the minimum of this initial subject score and SubjectMax.

UniqueMaximums can be used to calculate two scores that were produced by previous versions of POES. To calculate 334 scores, set KeyMax to 3 and SubpartMax to 4, and then calculate 334 scores as the sum of the subpart scores across all items and subparts. To calculate 3345 scores, set KeyMax to 3, SubpartMax to 4, and ItemMax to 5, and then calculate 3345 scores as the sum of the item scores. For this study, we used the default settings for UniqueMaximums (KeyMax = 3, SubpartMax = 4, ItemMax = 5), and then calculated the sum of the item scores, so that these scores are equal to 3345 scores.

CountCat. The CountCat method counts the number of distinct values that occur in a response. For example, if a response said, "I feel so sad I would cry", then "sad" has a value of 3 and "cry" has a value of 2, so that two distinct values (2 and 3) occur in this response. The CountCat score would be 2. Thus, CountCat is treating each value as a category and is counting the number categories that are represented by the keyed responses. The CountCat score can be calculated at the level of the subpart, the item, or the subject. For this study, we calculated the total of the item-level scores.

CatFreq. The CatFreq method counts the number of times each value occurs in a response. CatFreq produces as many scores as there are values. For example, imagine a response said, "I would be worried about her. She would be unhappy and might be thinking about suicide." This response contains four keys: "worried" with a value of 3, "unhappy" with a value of 3, "thinking" with a value of 0, and "suicide" with a value of 2. CatFreq0 would equal 1 because there was one key with a value of 0. CatFreq1 would equal 0 because no level 2 keys were used. Similarly, CatFreq2 and CatFreq3 would both equal 1. The CatFreq score can be calculated at the level of the subpart, the item, or the subject. In this study, we used the subject-level scores.

Procedures

Participants completed the MSCEIT and the LEAS as a part of a larger online study.

Results

All LEAS scoring methods had significant positive correlations with total scores on the MSCEIT (see Table 1), thus demonstrating their convergent validity. The HighestN and HighestNUnique methods (both with N = 40) had the highest correlations with the MSCEIT. Most methods had higher correlations than CountWords: The only exceptions were the count of the number of level 0 emotion words (CatFreq0) and the count of the number of level 1 emotion words (CatFreq1). The UniqueMaximums method (with KeyMax = 3, SubpartMax = 4, and ItemMax = 5) had a slightly higher correlation than the hand scoring method.

Table 1
Correlations of LEAS Scoring with the MSCEIT (Total Score)

LEAS Score	Correlation
Hand Scoring	.33***
POES Scoring	
CountWords	.21***
AllSum	.29***
AllSumUnique	.36***
HighestN†	.39***
HighestNUnique†	.40***
UniqueMaximums††	.34***
CountCat	.32***
CatFreq0	.20***
CatFreq1	.15**
CatFreq2	.25***
CatFreq3	.31***

** $p < .01$. *** $p < .001$. † N = 40.
†† KeyMax = 3, SubpartMax = 4, ItemMax = 5.

Discussion

The Levels of Emotional Awareness Scale (LEAS) can be scored by hand or by using Program for Open-Ended Scoring (POES 2.0.1). The purpose of this study was to compare the validity of hand scoring and the eight POES scoring methods by correlating each of these nine scores with total scores on the MSCEIT. All scoring methods had significant positive correlations with the MSCEIT, thus reinforcing the validity of the LEAS as a measure of emotional awareness. The highest correlations were for the HighestN (N = 40) and HighestNUnique (N = 40) methods. These correlations were larger than the correlations for hand scoring, thus making them a viable alternative to the time-consuming hand-scoring method.

One of the new scoring methods was CatFreq. It counts how often respondents use words that receive scores of 0, 1, 2, and 3. As might be expected, the frequency of level 3 emotion words (happy, sad, angry, etc.) had a higher correlation with the MSCEIT than the frequency of emotion words that receive lower scores. It appears that level 3 emotion words are driving the correlation between the LEAS and the MSCEIT.

Another new method in POES 2.0.1 is the UniqueMaximums method. This method is a generalization and replacement of the previous 334 and 3345 methods that were the closest possible approximation to hand-scoring. In this study, UniqueMaximums was set up with KeyMax = 3, SubpartMax = 4, and ItemMax = 5, so that it is equivalent to the 3345 method. Like previous research (Barchard et al., 2010), this method had slightly higher validity (r = .34) than the hand-scoring method it approximates (r = .33).

Another innovation in this version of POES is the AllSumUnique and HighestNUnique scoring methods. These methods give credit for an emotion word only the first time it appears. AllSumUnique had a higher correlation (r = .36) with the MSCEIT than AllSum did (r = .29). Similarly, HighestNUnique had a higher correlation (r = .40) with the MSCEIT than HighestN did (r = .39), although here the difference was trivially small (.01). Thus, it appears that duplicate words add little to our knowledge of someone's emotional awareness. What is most important is whether a person can describe the differences between various emotional experiences.

The last new method in POES 2.0.1 is CountCat. This method counts the number of distinct categories that were used in the responses, and thus indexes the richness of the emotional descriptions. In this dataset, CountCat had relatively high validity (r = .32), similar to hand scoring (r = .33), but not as good as the HighestN and HighestNUnique methods. This method deserves further exploration, perhaps with revised categories that more precisely examine the breadth of one's descriptions.

In conclusion, this study has demonstrated the validity of computer scoring of the LEAS. Some POES methods had higher correlations with the MSCEIT than hand-scoring. Moreover, having a computer do the scoring allows psychologists to use scoring methods that would be unwieldy if attempted by hand (AllSumUnique, HighestNUnique, UniqueMaximums), and thus allows us to measure new aspects of emotional awareness. In particular, the elimination of duplicate words across the entire set of 20 items allows us to measure the breadth of emotion knowledge, and counting the number of unique emotion categories allows us to measure the richness of emotional descriptions.

References

Alsaid-Habia T., Tomme A. K., Ladrazo L. M., Craun E. A., & Barchard K. A. (2012, April). *Examining the convergent validity of computerized scoring of the Levels of Emotional Awareness Scale.* Poster presented at the Western Psychological Association Annual Convention, San Francisco, CA.

Barchard, K. A. (2013, March). *LEAS Wordlist 2.5.* File to be used with POES to allow scoring of the Levels of Emotional Awareness Scale. Available from 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, kim.barchard@unlv.edu

Barchard, K. A., Bajgar, J., Leaf, D., & Lane, R. D. (2010). Computer scoring of the Levels of Emotional Awareness Scale. *Behavior Research Methods*, *42*, 586-595. doi:10.3758/BRM.42.2.586

Ermini Leaf, D., & Barchard, K. A. (2013). *Program for Open-Ended Scoring version 2.0.1*. [Unpublished computer program]. Available from 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, kim.barchard@unlv.edu

Lane, R. D. (1991). *LEAS scoring manual and glossary.* Unpublished manuscript. Available from R. 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., & Schwartz, G. E. (1987). Levels of emotional awareness: A cognitive-developmental theory and its application to psychopathology. *The American Journal of Psychiatry, 144*, 133-143.

Leaf, D. E. (2003). *Computerized Scoring Program for the LEAS.* [Unpublished computer program]. Available from 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, kim.barchard@unlv.edu

Leaf, D. E. & Barchard, K. A. (2010). *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.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2004). Emotional intelligence: Theory, findings, and implications. *Psychological Inquiry*, *15*, 197-215. doi:10.1207/s15327965pli1503_02

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT V2.0. *Emotion*, *3*, 97-105. doi:10.1037/1528-3542.3.1.97

Schutte, N. S., Malouff, J. M., Bobik, C., Coston, T. D., Greeson, C., Jedlicka, C., . . . Wendorf, G. (2001). Emotional intelligence and interpersonal relations. *The Journal of Social Psychology*, *141*(4), 523-536. doi:10.1080/00224540109600569