RUNNING HEAD: Computer Scoring

Computer Scoring of

The Levels of Emotional Awareness Scale

Kimberly A. Barchard, University of Nevada, Las Vegas, USA

Jane Bajgar, University of Wollongong, Australia

Duncan Ermini Leaf, Purdue University, USA

Richard D. Lane, University of Arizona, USA

SUBJECT TO COPY EDITING

Corresponding Author:

Kimberly A. Barchard, Associate Professor, University of Nevada, Las Vegas

4505 Maryland Parkway, Las Vegas, NV, 89154-5030

Email: barchard@unlv.nevada.edu Phone (702) 895-0758 FAX (702) 895-0195

Author Notes

Portions of these results were presented at the 2004 conference for the Association for Computing Machinery Special Interest Group on Computer Science Education, and the 2004 and 2007 conferences of the Western Psychological Association.

Abstract

The Levels of Emotional Awareness Scale (LEAS; Lane, Quinlan, Schwartz, Walker, & Zeitlan, 1990) is the most commonly used measure of differentiation and complexity in the use of emotion words, and is associated with important clinical outcomes. Hand scoring the LEAS is time-consuming. Existing programs for scoring open-ended responses cannot mimic LEAS hand scoring. Therefore, Leaf and Barchard (2006) developed Program for Open-Ended Scoring (POES) to score the LEAS. This article reports three studies examining the reliability and validity of POES scoring. The studies used three participant types (adult community members, university students, children), three LEAS versions (paper-based, computer-based, and the LEAS for Children), and a diverse set of criterion variables. Across this variety of conditions, the four POES scoring methods had internal consistencies and validities that were comparable to hand-scoring, indicating that POES scoring can be used in clinical practice and other applied settings where hand scoring is impractical.

Key words: Levels of Emotional Awareness Scale, LEAS, Computer Scoring, Computer Assisted Testing, Program for Open-Ended Scoring, POES, free software

Introduction

The Levels of Emotional Awareness Scale (LEAS; Lane, Quinlan, Schwartz, Walker, & Zeitlan, 1990) is an open-ended test of the ability to use emotion words in a complex and differentiated fashion. Respondents describe how they and another person would feel in 20 emotionally evocative situations. Because the LEAS is open-ended, it takes a long time to score. In the first author's lab, training a new LEAS scorer can take up to 10 hours. Once trained, LEAS scorers can take up to 20 minutes for each respondent. Thus, a new scorer may take up to 43 hours to score the LEAS for 100 respondents, and an experienced and efficient LEAS scorer will still take 10 minutes per respondent or roughly 17 hours for 100. The amount of time involved in scoring is likely the primary reason that researchers have not used the LEAS more often. To address this issue, Lane et al. (1990) advocated automating the LEAS. The purpose of this present project was to implement this suggestion: we developed a free computer program to score the LEAS, and this paper assesses the validity of the scores that this program produces. Before describing the scoring program, this paper first describes emotional awareness and the LEAS in more detail.

The Levels of Emotional Awareness Scale

Being aware of emotional experiences in the self and in others is perhaps the most fundamental of emotional intelligence skills (Lane, 2000). Individuals with higher emotional awareness have at their disposal complex and differentiated emotion information, which can be used to support higher level emotion processes such as understanding and managing emotions and using emotions to facilitate cognition.

As would be expected, lower levels of emotional awareness are related to a number of clinical conditions. LEAS scores are lower in people with somatoform disorders (Subic-Wrana,

Bruder, Thomas, Gaus, Merkle, & Köhle, 2002), depression (Berthoz, Ouhayoun, Parage, Kirzenbaum, Bourgey, & Allilaire, 2000; Donges, Kersting, Dannlowski, Lalee-Mentzel, Arolt, & Suslow, 2005), eating disorders (Bydlowski, Corcos, Jeammet, Paterniti, Berthoz, Laurier, et al., 2005), post-traumatic stress disorder (Frewen, Lane, Neufeld, Densmore, Stevens, & Lanius, 2008), multiple addictive behaviors (Jouanne, Edel & Carton, 2005) and borderline personality disorder (Levine, Marziali, & Hood, 1997). LEAS scores increase over the course of treatment for people with somatoform disorders independent of changes in negative affect (Subic-Wrana, Bruder, Thomas, Lane, & Köhle, 2005) and moreover predict improvements among psychosomatically ill patients (Simson, Martin, Schafer, Franz & Janssen, 2006). Finally, higher scores on the LEAS are associated with a stable sense of general well-being despite temporary negative mood (Ciarrochi, Caputi, & Mayer, 2003). Thus, emotional awareness is central to many clinical conditions and a foundational emotional skill for all people.

There is strong evidence for the construct validity of the LEAS as a measure of emotional awareness. LEAS scores are correlated with changes in cerebral blood flow associated with film- and recall-induced emotional experiences (Frewen et al., 2008; Lane, Reiman, Axelrod, Yun, Holmes & Schwartz, 1998), and are related to higher accuracy in the perception of emotion (Lane, Sechrest, Reidel, Weldon, Kaszniak, & Schwartz, 1996; Lane, Sechrest, Riedel, Shapiro, & Kaszniak, 2000; Subic-Wrana et al., 2005). Additionally, several studies (Lane, Sechrest, & Riedel, 1998; Rose, 2004; Walgren, 1996; Waller & Scheidt, 2004) have shown that higher LEAS scores are associated with lower scores on self-reported difficulty expressing feelings.

The LEAS is based upon a five-tiered developmental theory of emotional awareness (Lane & Schwartz, 1987), modeled after Piaget's theory of cognitive development. In this theory, each level represents a greater degree of differentiation and integration of emotional

experiences. In hand scoring, each item on the LEAS receives a score ranging from 0 (no awareness) to 5 (the highest level of awareness). See Table 1 for descriptions and examples of the six levels.

To score the LEAS by hand, the scorer needs to make a number of subjective decisions. First, the scorer needs to interpret the meaning of emotion words based upon context (e.g., "pain" can be scored as level 1 or level 3, depending upon whether it represents physical pain or emotional pain). Second, the scorer needs to determine if words are synonyms of each other, to decide if a person is describing a single emotion (e.g., "I would feel happy and joyous") or a blend (e.g., "I would feel happy and proud"). Third, the scorer needs to distinguish between emotions that are attributed to the self and emotions that are attributed to the other (e.g., "I would feel loved") and determine if the self and other are feeling the same emotions (e.g., "I would be happier than my husband.") These subjective decisions are fundamental to the cognitive-developmental theory upon which the LEAS is based.

The subjective decisions that are involved in hand scoring cannot be precisely duplicated by a computer. The best we can hope for is to develop a computer scoring program that is relatively similar to LEAS hand scoring. Even then, the computer scoring must be considered a new method, and its validity will need to be demonstrated.

Computerized Scoring

There has been substantial work on automated scoring of open-ended responses. However, none of these methods can be adapted to mimic LEAS hand scoring. Automated essay scoring is the largest area (for reviews of specific methods and validity results, see Shermis & Burstein, 2003; Valenti, Neri, & Cucchiarelli, 2003; Yang, Buckendahl, Juszkiewicz, & Bhola, 2002). These methods are typically designed to rate some combination of an essay's content and

writing style. In particular, the *k*-nearest-neighbors method of Larkey (1998) could be easily adapted to score LEAS responses. This approach classifies an essay into a rating category by comparing the words contained in that essay to the words present in a collection of essays that have already been rated by hand. Unfortunately, this method would not mimic the hand scoring of the LEAS, for which we have strong validity evidence. The *k*-nearest-neighbors approach might assign high scores because of the presence of non-emotion words, such as "because" or "I", and might give low scores to responses that provide detailed, clear descriptions of unusual emotional reactions.

Other computational methods have been specifically designed to measure emotional content in open-ended responses, but they, too, cannot mimic LEAS hand scoring. The Dictionary of Affect in Language (DAL; Whissell, 1989) has been used by many researchers to score text for emotional content (e.g., Dubois, 1997; Mossholder, Settoon, Harris, & Armenakis, 1995; Whissel, 2008; Zwaal, Prkachin, Husted, Stones, 2003). Each word in DAL has been rated for pleasantness and activity level, based upon the circumplex model of emotion (Russell, 1980). To score text, DAL searches for words from its Dictionary, and calculates the mean rating for the scored words, for each of the two dimensions. The DAL method is similar to LEAS scoring, in that words are assigned specific scores. However, the criteria for the word scores are different: DAL is based upon pleasantness and activity level whereas LEAS hand scoring is based upon the complexity and specificity of the emotion words. Furthermore, LEAS hand scoring uses criteria not specified in DAL, such as the presence of synonyms and the distinction between self and other. Thus, DAL could be used to score LEAS responses, but the scores would bear little conceptual relationship to LEAS hand scores.

The Linguistic Inquiry and Word Count program (LIWC; Pennebaker, Chung, Ireland, Gonzales, & Booth, 2008) has also been used to score text for emotion content in a wide variety of research areas (e.g., Ali & Levine, 2008; Carroll, 2007; Pennebaker, 1997). LIWC 2007 contains a dictionary of about 4500 words that have been classified as falling into one or more of 64 linguistic categories, including several categories related to emotions. LIWC counts the number of words in a response that match the words in each linguistic category. Additionally, the LEAS glossary could be used with LIWC to obtain a count of the number of times that LEAS glossary words are used. However, LIWC could not be used to approximate LEAS hand scoring. LIWC gives the same score to each word in the list, does not take into account identical words being used repeatedly, and does not distinguish between emotions that were attributed to the self versus the other person in the scenario.

Thus, none of the existing computerized scoring methods can be adapted to mimic LEAS hand scoring. This motivated us to write a computer program, Program for Open-Ended Scoring (POES; Leaf & Barchard, 2006), specifically designed for automated LEAS scoring. This program includes one scoring method (All-Sum) that is similar to the DAL and LIWC programs, but it also goes beyond those methods to more closely mimic LEAS hand scoring. Additionally, because POES is distributed under the GNU General Public License (Free Software Foundation, 1991), it is freely available and can be modified by others to include new scoring methods. POES consists of two files: a DOS-based program which takes 96 KB of disk space, and a Windows-based interface which takes 64 KB of disk space.

The Current Study

The purpose of this present project was to assess the validity of POES scoring. First, we scored three sets of data using hand scoring and POES scoring. Then, we correlated the two

types of scores with each other, and evaluated their internal consistencies and their correlations with criterion variables. The three datasets we used involved different types of participants and different versions of the LEAS, and the criterion variables are diverse. Therefore, any results that are consistent across the three datasets are likely to be generalizable to many other datasets.

Two of the data sets used in this paper have already been used to examine the internal consistency and criterion-related validity of the LEAS. We re-scored the data from those two studies to determine if POES scores produce similar results. First, we re-scored the adult community member data from McRae, Reiman, Fort, Chen, and Lane (2008). That study found that hand scoring was internally consistent (coefficient alpha = .84) and was associated with blood flow in the dorsal anterior cingulate cortex. This area is associated with awareness and monitoring of emotional experiences (Hutcherson, Goldin, Ochsner, Gabrieli, Feldman Barrett, & Gross, 2005; Ochsner, Bunge, Gross, & Gabrieli, 2002; Taylor, Phan, Decker, & Liberzon, 2003). Second, we re-scored the data from Bajgar, Ciarrochi, Lane, and Deane (2005). That study found that hand scoring of the children's version of the LEAS had moderate internal consistency (coefficient alpha = .66) and had a significant relationship with the Vocabulary subtest of the WISC-III (r = .31, p < .05) and nearly significant relationship with Emotion Comprehension. The purpose of the current study was to determine if POES scoring of these two datasets could replicate these internal consistencies and criterion validities.

The third study used new research data from a computerized form of the LEAS. In the previous two data sets, we used three types of POES scores. In the third study, we also included a fourth POES scoring method, which could only be used with the computerized form. The third study used adult university students to assess internal consistency and the correlations between POES scoring and hand scoring.

Method

Participants

Sample 1: Adult Community Members

Forty-four participants (22 female, 22 male) were recruited through advertisements posted at a medical center, at a large local university, and in churches, libraries and other public facilities. Participants were screened by phone and were excluded if they reported a history of neurological abnormalities, head injury, learning disabilities, current psychoactive medication use, current drug or alcohol abuse, current major depressive episode, or lifetime prevalence of a psychotic disorder. By design, participants ranged in age from 19 to 30 (mean 24.6, SD 3.3). Participants identified themselves as follows: 75% White non-Hispanic, 14% Hispanic, 6% Asian, and 5% African-American. All participants spoke English as their native language. Additional information about these participants can be found in McRae et al. (2008).

Sample 2: Children

Fifty-one children (25 female, 26 male) were recruited from two private schools in a regional city with a population of 180,000. They ranged in age from 10 to 11 (mean 10.3, SD .48). All children came from middle class backgrounds with the majority of parents working in professional or semi-professional occupations. Five children were of non-English speaking cultural backgrounds. However, all children were identified as competent English speakers. Additional information about these children can be found in Bajgar et al. (2005).

Sample 3: Adult University Students

Sixty-six university students (52 female, 14 male) participated in return for course credit. They ranged in age from 18 to 46 (mean 24.7, SD 8.1). Participants identified themselves as follows: 62% White/Caucasian, 9% Black/African American, 8% Hispanic, 6% Pacific Islander

and 6% Asian. All participants either spoke English as their first language or had been speaking English for at least 10 years and reported being very comfortable reading and writing in English.

Measures

Levels of Emotional Awareness Scale

Paper-administered LEAS

The adult community members in Sample 1 completed the original paper-administered LEAS. Each of the 20 LEAS items consists of an emotionally evocative scenario involving oneself and another person, and is presented at the top of a new page. Each scenario was designed to evoke one of four different emotions (anger, sadness, fear, happiness), at one of the five levels of Emotion Awareness in the Lane and Schwartz (1987) model. Each scenario is followed by two prompts: "How would you feel?" and "How would the other person feel?"

LEAS for Children

The children in Sample 2 completed the paper-administered Levels of Emotional Awareness Scale for Children (LEAS-C; Bajgar et al., 2005). The LEAS-C consists of 12 scenarios, each involving oneself and another person. Of these 12 scenarios, 10 are modified LEAS items with simplifications of grammar and vocabulary. Two new items reflect school and peer-related concerns. In this study, the LEAS-C scenarios were read aloud while the children read silently. The children then provided written responses to the two questions: "How would you feel?" and "How would the other person feel?"

Computer-administered LEAS

The adult students in Sample 3 completed the 20-item computer-administered LEAS. It consists of two web pages, each containing ten LEAS items. For each item, the scenario description was followed by two prompts: "How would you feel?" and "How would the other

person feel?" After each prompt, a text input box (8 rows by 40 columns) collected participants' responses. These web pages were uploaded to the university server so they could be accessed from university computer labs via the Internet.

Scoring

All three versions of the LEAS were scored in two ways. First, responses were scored using the hand-scoring method described in the manual for the adult version of the LEAS (Lane, 1991). Second, responses were spell-checked and then scored using POES version 1.2.2 (Leaf & Barchard, 2006) and LEAS Wordlist 2.1 (Barchard, 2006).

Hand scoring. When hand scoring the LEAS, item scores are calculated in three stages (Lane, 1991). First, each word in the response is assigned a score. Non-emotion words (e.g., expect) are scored 0. Physiological sensations (e.g., dizzy) are scored 1. Words that indicate emotions but also have non-emotional meanings (e.g., bad) are scored 2, as are actions related to emotions (e.g., cry) and personality traits (e.g., tolerant). Discrete emotion words (e.g., fear) are scored 3. The scoring manual glossary (Lane, 1991) lists words and phrases and their associated levels. However, the glossary is not exhaustive: the scorer often has to look for synonyms or check the rules regarding the types of words included at each level. Also, words often have more than one possible score, and the scorer must use these general scoring rules and the context to decide which score is appropriate.

The second stage is to calculate self and other scores. To do this, the scorer decides if each emotion word is attributed to the self or the other person. For example, in the statement "I would be happy my love was reciprocated", who feels "love" – the self, the other, or both? The scorer also has to decide if emotion words are synonymous: is "love" the same as "care for"? Based upon these two subjective judgments, the scorer calculates the self and other scores. If the

response has two non-synonymous Level 3 emotion words attributed to the self, the self score is 4. If not, the self score is the maximum of the word scores for emotions attributed to the self. The other score is calculated analogously, based upon emotions attributed to the other person.

The third stage is to calculate the item score. If self and other scores are both 4 and the emotions for self and other are different, the item score is 5. Otherwise, it equals the maximum of the self and other scores. This results in item scores that range from 0 (no emotion words) through 5 (combinations of blends of emotions), corresponding to the six levels of emotional awareness in the Lane and Schwartz (1987) model. See Table 1. Total test scores are calculated as the sum of the item scores for the 20 items, and thus range from 0 to 100.

For each of the three samples, hand scoring was completed by experienced LEAS scorers, who had scored at least 1000 responses before beginning scoring of these samples (for example, they had previously scored at least 50 participants who completed all 20 items). Inter-rater reliability was assessed by having one additional person score the responses for each sample. In Sample 1 (adult community members) and Sample 3 (adult university students), the second scorers met the above definition of experienced scorers, and they rated all protocols. In Sample 2 (children), the second scorer was well trained, having received five hours of training, including a detailed explanation of the scoring procedure and intensive trial scoring with feedback. The second scorer rated 22 of the protocols. In all samples, inter-rater reliability was high. In Sample 1, the Pearson correlation was .98; in Sample 2, it was .89; in Sample 3, it was .95.

POES scoring. To score the LEAS, POES 1.2.2 (Leaf & Barchard, 2006) requires two input files: the participants' typed response data and a Wordlist file that specifies the score values to be given for specific words and phrases. Currently, there are two versions of the LEAS

Wordlist. The initial version, LEAS Wordlist 1.0 (Leaf & Barchard, 2002), was based directly upon the LEAS hand-scoring manual (Lane, 1991). The major challenge in translating the LEAS glossary into the LEAS Wordlist 1.0 was handling words with multiple values. When the LEAS glossary gave only a single value for a word, then that value was entered in LEAS Wordlist 1.0. In some cases, however, a glossary word has two values depending upon context. For example, "hurt" could be Level 1 if it refers to a physical sensation or Level 3 if it describes an emotion. Because POES 1.2.2 cannot make distinctions based upon context, each Wordlist entry must be associated with only a single value. For multi-valued glossary words, a team of experienced LEAS scorers discussed which interpretation was most common, and decided which value would be used in the Wordlist. The LEAS Wordlist 1.0 contained 888 word and phrase entries. The main limitation of this original Wordlist was that it did not contain words and phrases that were nearly identical to LEAS glossary entries.

Wordlist 2.1 (Barchard, 2006) expanded the list of words and phrases that received scores in Wordlist 1.0. A team of five expert LEAS scorers considered 467 possible additions and modifications. The changes fell into five categories. First, if an LEAS glossary entry included a pronoun (e.g., "at the end of my rope"), additional entries were added for all remaining pronouns (e.g., "at the end of his rope"). Second, if an LEAS glossary entry did not include the most common way of phrasing an idea, the most common phrasing was added. Third, when an entry included a verb, alternative verb tenses were added. Fourth, when an entry included a word for which there are different forms, the other forms of the word were added (e.g., original entry "joy"; additional entry "joyful"). Fifth, if an entry included parenthetical material to clarify meaning, we separated the parenthetical and non-parenthetical material, and considered both as

possible entries. Changes were made if four of the five expert scorers agreed that the change should be made. The final LEAS Wordlist 2.1 consists of 1242 word and phrase entries.

Once the Wordlist file and participants' response data are specified, POES scores the data in three stages, analogous to the three stages of LEAS hand scoring. First, POES scans each item response for words and phrases given in the Wordlist. These words and phrases are called valuables. Each valuable found in the response, along with its score value, is recorded in the Valuables List for that item. Next, POES calculates item scores using four different scoring methods (described below). Finally, for each scoring method, the 20 item scores are summed to calculate the total test score. Participants receive four total test scores, one for each scoring method.

POES 1.2.2 uses four methods of calculating item scores for the contents of the Valuables List. These methods are summarized in Table 2. The first three methods were used in all samples; the last method was used only in Sample 3 because it requires separate self and other responses to each question. The first method, All-Sum, calculates the sum of all the values in the Valuables List for an item. This method is the most straightforward procedure and incorporates all of the information given in a response. Because of its simplicity and generality, All-Sum is the scoring method that is most applicable to scoring other open-ended tests, besides the LEAS.

The second method, Highest-4, calculates the sum of the four highest values in the Valuables List. For LEAS hand scoring, item scores are based upon four values: the two highest values for self and the two highest values for other. Highest-4 is a simplification of that idea. It looks at the four highest values, but ignores which person an emotion was attributed to.

The third method, 334, is like hand scoring in that it gives higher scores to responses that include multiple, distinct emotion words than to responses that use identical words or phrases

repeatedly. Specifically, the 334 method searches the Valuables List for all valuables with a value of 3. If these are all identical, an item score of 3 is assigned. If any two valuables are not identical, an item score of 4 is given. If there are no valuables with a value of 3, the item score is the maximum value found in the Valuables List.

The final scoring method, 3345, provides the best emulation of LEAS hand scoring. This method uses the separate responses to the questions "How would you feel?" and "How would the other person feel?" to distinguish between emotions attributed to self and other. In this study, the responses were separated by using two text input boxes. On a paper-administered form, there could be separate areas for responses to the two questions. In either case, this separation only approximates the distinction made during hand scoring because LEAS responses often attribute emotions to self and other in the same sentence. There is no guarantee that respondents enter only self-emotions in the "How would you feel?" area and other-emotions in the "How would the other person feel?" area.

The 3345 method calculates an item score in three steps. First, self and other Valuables Lists are created, based upon the responses to the two questions. Next, these lists are scored separately using the 334 method to obtain self and others scores. Finally, the item score is calculated based upon the self and other scores: an item score of 5 is given if the self and other scores are both 4; otherwise, the item score is the maximum of the self and other scores. Because the 334 and 3345 methods are logically closest to hand scoring, we expected them to have higher correlations with hand scoring and more similar correlations with the criterion variables.

The POES 1.2.2 scoring methods follow concrete algorithms and do not attempt to mimic any of the subjective judgments required for hand scoring. In particular, POES does not

determine word meaning based upon context, does not consider synonyms, and does not actively try to decide whether an emotion is attributed to self or other. Thus, even the POES scoring methods that most closely mimic hand scoring only *approximate* human scoring. Thus, the purpose of this study is to evaluate the reliability and validity of these scoring methods. Four different methods were included so we could assess the value of increasing scoring complexity to more closely mimic hand scoring.

Cerebral Blood Flow

In Sample 1, we examined the relationship of POES scores to changes in cerebral blood flow due to emotional arousal. We varied participants' levels of emotional arousal by having each participant complete eight conditions: six picture viewing conditions and two visual fixation conditions. During the six picture viewing conditions, participants looked at pictures from the International Affective Picture System (Lang, Bradley, & Cuthbert, 2005). These pictures have been normatively rated based upon valence (pleasant - unpleasant) and arousal. In this study, four of the conditions consisted of high and low arousal pictures that were pleasant or unpleasant. The two remaining picture viewing conditions were neutral on valence: one contained human faces in each picture and the other consisted mostly of scenes and everyday objects. These six picture viewing conditions were counter-balanced around the first and fifth conditions, which were always visual fixation. During all eight conditions, participants were told "Please look at the screen and allow yourself to feel whatever each picture evokes in you." While participants viewed the screen, functional images were acquired on an Exact HR+ PET scanner (Siemens, Knoxville, TN) operating in 3D mode. See McRae et al. (2008) for details regarding the PET imaging procedures.

To identify those areas of the brain where changes in blood flow are most highly

associated with LEAS scores, image pre-processing and voxel-based analyses were performed using the Statistical Parametric Mapping (SPM2) program. We followed the standard practice of partialling out global blood flow. Details regarding the analyses of LEAS hand scores are given in McRae et al. (2008). In this paper, we repeated those analyses, substituting POES scores where McRae et al. used hand scores.

Emotion Expressions

The Emotion Expressions task assesses recognition of emotion cues in faces (Fine, Izard, Mostow, Trentacosta, & Ackerman, 2003). The children in Sample 2 viewed 18 facial photos of adults posing one of six emotions (anger, surprise, sadness, disgust, joy, or fear), taken from Izard's I-M series (Izard, 1971) and the Glenn pictures (Glenn, 1974). These 18 photos were recommended by C. Izard (personal communication, April, 2000). For each photo, children wrote down what the person was feeling. The response was scored 2 if the correct emotion or a synonym was given; 1 if the emotion was inaccurate but had the correct positive or negative valence (e.g., anger instead of fear); and 0 if both the emotion and valence were incorrect. Internal consistency of the Emotion Expressions task has been reported as $\alpha = .52$ (Schultz & Izard, 1998). In this study, the internal consistency was $\alpha = .52$.

Emotion Comprehension

The children in Sample 2 completed an Emotion Comprehension task (Cermele, Ackerman & Izard, 1995). This task assesses recognition of emotion cues in situations (Garner, Jones, & Miner, 1994; Ribordy, Camras, Stefani, & Spaccarelli, 1988). Children listened to two series of emotionally evocative scenarios, while reading along silently. In the 18 scenarios in the first series, children selected the emotional response of the protagonist from the following list: happy, sad, mad, interested, or ashamed. In the nine scenarios in the second series, children

selected from a slightly different list of responses: happy, mad, proud, guilty, ashamed, or looking down on someone. Similar to the Emotion Expression task, responses were scored according to accuracy and valence. Internal consistency of the Emotion Comprehension task has been reported as $\alpha = .58$ (Schultz & Izard, 1998). In this study, internal consistency was $\alpha = .59$.

Vocabulary

The children in Sample 2 completed the vocabulary subtest of the Wechsler Intelligence Scale for Children – 3rd version (WISC-III; Wechsler, 1991). In this study the task was group administered (Barrett, Lane, Sechrest & Schwartz, 2000; Subic-Wrana, Thomas, Huber, & Koehle, 2001). Fifteen words were read aloud to students, the starting point being the one recommended for the lowest age of the participants (Sattler, 1992). As each word was presented, children wrote down its meaning. The average internal consistency for the vocabulary subtest of the WISC-II is $\alpha = .87$ (Kamphaus, 2005). In this study, the internal consistency of the subtest was $\alpha = .71$.

Results

Internal Consistency (All Samples)

For all three samples, the four POES scoring methods had acceptable internal consistencies. See Table 3. For the adult samples, internal consistencies ranged from good ($\alpha = .79$) to excellent ($\alpha = .92$). For the child sample, the POES 334 method had barely acceptable internal consistency ($\alpha = .60$). However, hand scoring also had only acceptable levels of internal consistency in this sample, and the internal consistency of the 334 method was not significantly lower than the internal consistency for hand scoring.

The POES methods differed in their internal consistencies. The Highest-4 and All-Sum methods had the highest internal consistencies in all three samples. The internal consistencies

for the 334 and 3345 methods were the most similar to hand scoring, and in most cases, coefficient alpha for these two POES methods was not significantly different from coefficient alpha for hand scoring.

Correlations of the Four POES Scores with Hand Scoring (All Samples)

For all three samples, each of the four POES scores had high correlations with hand scoring. This indicates that all POES scoring methods are tapping the same general construct as hand scoring, thus providing evidence of concurrent validity. See Table 4.

Next, we wanted to determine which POES methods had the highest correlations with hand scoring. We had hypothesized that the 3345 and 334 methods – the methods that were logically most similar to hand scoring – would have higher correlations with hand scoring than the All-Sum and Highest-4 methods. However, this was not always true and sometimes the reverse was true. The 334 method and Highest-4 methods appear to be roughly comparable in terms of their correlation with hand scoring.

The 3345 method had a significantly higher correlation with hand scoring than any of the other POES scoring methods, in the one sample (adult students) in which it was calculated. Because it was possible to use this scoring method only in the sample that used the computer-administered LEAS, this result requires replication before we can be confident that the 3345 method has the highest correlation with hand scoring.

To further examine the similarity of the constructs underlying hand scoring and computer scoring, we corrected these correlations for attenuation due to lack of internal consistency. See Table 4. These corrected correlations were often quite high, and for Sample 3 (adult students) the corrected correlations for the 334 and 3345 methods were close to 1. For the LEAS-C used

in Sample 2 (children), the corrected correlations were somewhat low. It may be that the lower emotional awareness of children results in restriction of range, which reduces these correlations.

Associations with Criterion Variables (Sample 1)

McRae et al. (2008) showed that LEAS hand scores are associated with changes in blood flow in the dorsal anterior cingulate cortex (dACC) between the high arousal and low arousal conditions, using the data in Sample 1. The purpose of the current analyses was to determine if a similar area of the dACC would be associated with POES scores. We therefore repeated McRae et al.'s analyses, but in the analyses we replaced the hand-scored LEAS with the Highest-4, All-Sum, and 334 scores. All three POES scoring methods were significantly related to changes in blood flow between the low and high arousal conditions. Moreover, for all three methods, the area with the highest association with POES scores was in Broadman's area 24 at coordinates 0, 12, 32. See Table 5. As mentioned previously, this cerebral area is associated with awareness and monitoring of emotional experiences (Hutcherson, Goldin, Ochsner, Gabrieli, Feldman Barrett, & Gross, 2005; Ochsner, Bunge, Gross, & Gabrieli, 2002; Taylor, Phan, Decker, & Liberzon, 2003).

Associations with Criterion Variables (Sample 2)

Bajgar et al. (2005) correlated hand scores with Emotion Expressions, Emotion Comprehension, and Vocabulary, using the data in Sample 2. The purpose of the current study was to determine if the correlations for POES scoring would be similar to the correlations reported by Bajgar et al. for hand scoring. We therefore correlated the three POES scores with the three criterion variables. Table 6 shows that all three POES methods had significant (or nearly significant) correlations with each of the three criterion variables. The correlations with Emotion Expressions and Emotion Comprehension are moderate, as would be expected because

these are related but not identical concepts. The correlations with Vocabulary are also moderate, as would be expected because emotional experiences are conveyed through language. These results provide evidence for the validity of all three POES methods.

We wanted to determine which of the POES scoring methods had the highest correlations with the criterion variables. Because the LEAS-C was not computer administered, it was not possible to calculate the 3345 method. Of the remaining three methods, the correlations for the 334 method were the most similar to the correlations for hand scoring, as expected. The Highest-4 method had consistently higher correlations than the other POES methods, but these correlations were not significantly greater than the correlations for the other POES methods.

Discussion

The Levels of Emotional Awareness Scale can assist in clinical practice and research. Awareness of emotional experiences is fundamental to individuals' ability to be self reflective and to relate harmoniously with others. Emotional dysfunctions – such as restricted expression of emotions, difficulties modulating the intensity and duration of emotions, and predominance of emotions such as anger, sadness, or anxiety – are central to a number of Axis I and Axis II disorders (American Psychiatric Association, 1994). Because of this, many mental health interventions increase awareness of emotions (Greenberg, 2002; Linehan, 1993) to lay the groundwork for understanding more complex processes, such as the source of emotions, appropriateness of emotions, and impact of emotions on functioning. Because the LEAS measures the degree of complexity in the cognitive processing of emotion information, it has clinical relevance to understanding individual mental health needs, and can facilitate research on the etiology and treatment of clinical disorders.

Research and clinical practice using the LEAS is hampered, unfortunately, because this measure is time-consuming to score. Lane et al. (1990) recommended automating the LEAS, which would greatly reduce scoring time and would eliminate the need to evaluate inter-rater reliability. The purpose of this paper was to examine the usefulness of our attempt to automate the LEAS. The results were highly encouraging. The four computer scoring methods resulted in high internal consistencies and high correlations with hand scoring in the adult samples, and moderately high internal consistencies and correlations with hand scoring in the child sample. POES scores also demonstrated validity by having moderate correlations with Emotion Expressions, Emotion Comprehension, and Vocabulary, and by being associated with changes in cerebral blood flow in the exact same location as hand scoring. These results suggest that POES scoring could be used instead of hand scoring, in situations for which the validity of hand scoring is known. The paper-based adult LEAS has significant validity evidence in a variety of contexts. Research on the validity of the computer-based adult LEAS and the LEAS-C has just begun, but could fruitfully use POES scoring.

Two primary differences were found between the four computer scoring methods we examined. First, the 334 and 3345 methods had slightly higher correlations with hand scoring, as expected. Therefore, if researchers or applied psychologists are trying to select a computer scoring method that is as similar as possible to the well-validated hand-scoring method, then the 334 or 3345 methods should be used. Of these two, the 3345 method had a significantly higher correlation with hand scoring and slightly higher internal consistency and is therefore preferred if both scoring methods are an option (i.e., if respondents provided separate answers to the two questions for each item).

The second primary difference was in the reliability and validity evidence. The Highest-4 and All-Sum methods had significantly higher internal consistencies and slightly higher correlations with Emotion Expressions. Thus, if a test user is trying to select the computer scoring method with the greatest validity evidence, these two methods are recommended. Of these two, the Highest-4 method had significantly higher correlations with hand scoring and Emotion Expressions, and is therefore preferred. Additional validity studies are needed to replicate and extend these findings to conclude more definitely which POES method is the most reliable and valid in any particular context. However, the diversity of samples, LEAS versions, and criterion-related variables examined in this paper leads us to predict the generalizability of these conclusions.

This paper was limited by the small sample sizes in each of the three samples. Small sample sizes limit statistical power. Although the Type I error rate is controlled by alpha (in this paper, we used alpha = .05), the Type II error rate can become high. For example, although the Highest-4 method had significantly higher internal consistency than hand scoring in each of the three studies, hand scoring had higher internal consistency than the 334 method in each of the three studies but none of those differences reached statistical significance. Similarly, the differences between the correlations for the four POES methods rarely reached statistical significance. Because of the small samples sizes, this paper cannot definitively state which scoring method is the best, even though it is able to provide compelling evidence for the internal consistency, concurrent validity, and criterion-related validity of the POES scoring methods. Future research should strive to incorporate larger sample sizes so that we can more clearly determine which scoring methods are preferred. Until future research definitively indicates which POES method is the best, all four POES methods should be used when feasible.

Future research should use POES to score other constructs. The LEAS Wordlist was designed to score open-ended responses for emotional awareness. In this Wordlist, higher scores are given to emotion words that indicate greater specificity and complexity. By using other Wordlists, POES can be used to score other constructs. Thus, POES goes beyond the functionality of LIWC and DAL. LIWC counts the number of words that fall into each linguistic category and allows user-defined lists, but all words within those categories receive identical scores. DAL looks up the scores for each word, but is limited to the two dimensions of pleasantness and activity level. POES combines the best features of each program, by allowing the researcher to use their own Wordlists. Moreover, POES goes beyond LIWC and DAL because it includes more than one scoring method. All-Sum is similar to the methods used by LIWC and DAL, but Highest-4 is also likely to be relevant to many testing circumstances. Finally, POES is more flexible than the other programs because the detailed score report from POES can be used to produce descriptive statistics, such as the number of times certain valuables are used or the number of times a certain value is used; Combined with custom-created Wordlists, this makes POES adaptable to scoring the LEAS for additional constructs and to scoring new open-ended tests. The next revolution in computerized scoring may be incorporation of user-defined scoring methods. Because POES is distributed under the GNU General Public License, researchers are welcome to modify it to include new scoring methods or user-defined scoring methods. Users are also encouraged to offer suggestions, if they do not have the programming skills to make the changes themselves.

Finally, future research should explore other ways of using POES to assist scoring. For example, researchers could use POES as an aid in hand scoring, rather than using it as a substitute for hand scoring as we did here. A fully trained LEAS scorer could use the detailed

score report available from POES to identify words and phrases that need to be scored. Words and phrases that have multiple values in the LEAS Glossary could be flagged by being assigned a score of 99 in the Wordlist, and the researcher could determine the appropriate score based upon context. However, we cannot recommend the use of POES as an aid in hand scoring until research demonstrates the reliability and validity of scores derived in that manner. Although it is true that scorers sometimes overlook words that occur in the LEAS Glossary (and the use of POES would likely prevent that), hand scoring is also based upon words that do not occur in the LEAS Glossary, and it is possible that the use of POES would make it more likely that those words would be overlooked unless the Wordlist was made to be truly comprehensive. Research in this area would be valuable.

Taken as a whole, the results of this paper indicate that LEAS computer scoring and hand scoring are tapping the same general construct and that computer scoring may be used instead of hand scoring in some contexts. Such use could enlighten and facilitate research on the etiology, nature, and treatment of clinical disorders. In addition, computer scoring could facilitate the measurement of emotional awareness in applied settings, where clinicians, educators, and Human Resource personnel may not have time to score by hand. Research on LEAS-like scenarios that are specialized for specific disorders (e.g., playground bullying, eating disorders, post-traumatic stress disorder) may further stimulate applied use, and will be facilitated by the speed and flexibility of POES scoring.

References

Ali, M. & Levine, T. (2008). The language of truthful and deceptive denials and confessions. *Communication Reports, 21*, 82-91. doi:10.1080/08934210802381862

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Bajgar, J., Ciarrochi, J., Lane, R., & Deane, F. (2005). Development of the Levels of Emotional Awareness Scale for Children (LEAS-C). *British Journal of Developmental Psychology*, *23*, 569-586. doi:10.1348/026151005X35417

Barchard, K. A. (2006). *LEAS Wordlist 2.1.* File to be used with Program for Open-Ended Scoring to allow scoring of the Levels of Emotional Awareness Scale. Available from Kim Barchard, Department of Psychology, University of Nevada, Las Vegas, 4505 Maryland Parkway, Las Vegas, NV, 89154-5030, barchard@unlv.nevada.edu

Barrett, L. F., Lane, R. D., Sechrest, L., & Schwartz, G. E. (2000). Sex differences in emotional awareness. *Personality & Social Psychology Bulletin, 26,* 1027-1035. doi:10.1177/01461672002611001

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

Carroll, D. W. (2007). Patterns of student writing in a critical thinking course: A quantitative analysis. *Assessing Writing, 12*, 213-227. doi:10.1016/j.asw.2008.02.001

Cermele, J. A., Ackerman, B. P., & Izard, C. E. (1995). *Children's Emotion Situation Knowledge*. Unpublished manuscript, University of Delaware, Newark, DE.

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

Donges, U., Kersting, A., Dannlowski, U., Lalee-Mentzel, J. Arolt, V. & Suslow, T. (2005). Reduced awareness of others' emotions in unipolar depressed patients. *Journal of Nervous & Mental Disease, 193*, 331-337. doi:10.1097/01.nmd.0000161683.02482.19

Dubois, S. L. (1997). Gender differences in the emotional tone of written sexual fantasies. *Canadian Journal of Human Sexuality, 6*, 307-315.

Feldt, L. S. (1965). The approximate sampling distribution of Kuder-Richardson reliability coefficient twenty. *Psychometrika, 30,* 357-370. doi:10.1007/BF02289499

Feldt, L. S. (1980). A test of the hypothesis that Cronbach's alpha reliability coefficient is the same for two tests administered to the same sample. *Psychometrika, 45*, 99-105. doi:10.1007/BF02293600

Fine, S. E., Izard, C. E., Mostow, A. J., Trentacosta, C. J., & Ackerman, B. P. (2003). First grade emotion knowledge as a predictor of fifth grade self-reported internalizing behaviors in children from economically disadvantaged families. *Development and Psychopathology, 15*, 331-342. doi:10.1017/S095457940300018X

Free Software Foundation. (1991). *GNU General Public License (version 2).* Retrieved from http://www.gnu.org/licenses/old-licenses/gpl-2.0.html

Frewen, P., Lane, R. D., Neufeld, R. W. J., Densmore, M., Stevens, T., & Lanius, R. (2008). Neural correlates of Levels of Emotional Awareness during trauma script-imagery in posttraumatic stress disorder. *Psychosomatic Medicine, 70,* 27-31. doi:10.1097/PSY.0b013e31815f66d4

Garner, P. W., Jones, D. C., & Miner, J. L. (1994). Social competence among low-income preschoolers: Emotion socialization practices and social cognitive correlates. *Child Development, 65,* 622-637. doi:10.2307/1131405

Glenn, R. (1974). *The Glenn pictures: Afro American faces of emotion*. Bloomington, IN: Indiana University.

Greenberg, L. (2002). *Emotion-focused therapy: Coaching clients to work through their feelings*. Washington, DC: American Psychological Association.

Howell, D. C. (2007). *Statistical methods for psychology (6th edition).* Belmont, CA: Thompson Wadsworth.

Hutcherson, C. A., Goldin, P. R., Ochsner, K. N., Gabrieli, J. D. E., Feldman Barrett, L., & Gross, J. J. (2005). Attention and emotion: Does rating emotion alter neural responses to amusing and sad films? *NeuroImage, 27,* 656-668. doi:10.1016/j.neuroimage.2005.04.028

Izard, C. E. (1971). *The face of emotion.* New York: Appelton-Century-Crofts.

Jouanne, C., Edel, Y., & Carton, S. (2005). Emotional deficits among multiple substance abusers. [French]. *Annales Medico-Psychologiques, 163*, 625-630.

Kamphaus, R. W. (2005) *Clinical Assessment of Child and Adolescent Intelligence, 2nd Ed*. Springer, US.

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. (2000). Levels of Emotional Awareness: Neurobiological, psychological and social perspectives. In R. Bar-On, & J. Parker (Eds.), *Handbook of Emotional Intelligence: Theory, development, assessment and application at home, school, and in the workplace* (pp. 171-191). San Francisco: Jossey-Bass.

Lane, R. D., Quinlan, D. M., Schwartz, G. E., Walker, P. A., & Zeitlan, 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., Reiman, E. M., Axelrod, B., Yun, L. S., Holmes, A., & Schwartz, G. E. (1998). Neural correlates of levels of emotional awareness: Evidence of an interaction between emotion and attention in the anterior cingulate cortex. *Journal of Cognitive Neuroscience, 10,* 525-535. doi:10.1162/089892998562924

Lane, R. D., & Schwartz, G. E. (1987). Levels of emotional awareness: A cognitive-developmental theory and its application to psychopathology. *American Journal of Psychiatry, 144,* 133-143.

Lane, R. D., Sechrest, L., Riedel, R., Shapiro, D. E., & Kaszniak, A. W. (2000). Pervasive emotion recognition deficit common to alexithymia and the repressive coping style. *Psychosomatic Medicine, 62,* 492-501.

Lane, R. D., Sechrest, L., & Reidel R. (1998). Sociodemographic correlates of alexithymia. *Comprehensive Psychiatry, 39,* 377-385. doi:10.1016/S0010-440X(98)90051-7

Lane, R. D., Sechrest, L., & Reidel, R., Weldon, V., Kaszniak, A., & Schwartz, G. E. (1996). Impaired verbal and nonverbal emotion recognition in alexithymia. *Psychosomatic*

Medicine, 58, 203-210.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2005). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-6*. University of Florida, Gainesville, FL.

Larkey, L. S. (1998). Automatic essay grading using text categorization techniques. *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, 90-95, Australia. doi:10.1145/290941.290965

Leaf, D. E. & Barchard, K. A. (2002). *LEAS Wordlist 1.0.* File to be used with Program for Open-Ended Scoring to allow scoring of the Levels of Emotional Awareness Scale. Available from Kim Barchard, Department of Psychology, University of Nevada, Las Vegas, 4505 Maryland Parkway, Las Vegas, NV, 89154-5030, barchard@unlv.nevada.edu

Leaf, D. E. & Barchard, K. A. (2006). *Program for Open-Ended Scoring [POES] version 1.2.2.* [Unpublished computer program]. Available from Kim Barchard, Department of Psychology, University of Nevada, Las Vegas, 4505 Maryland Parkway, Las Vegas, NV, 89154-5030, barchard@unlv.nevada.edu

Levine, D., Marziali, E., & Hood, J. (1997). Emotion processing in borderline personality disorders. *Journal of Nervous and Mental Diseases, 185*, 240-246. doi:10.1097/00005053-199704000-00004

Linehan, M. M. (1993). *Skills training manual for treating borderline personality disorder.* New York: Guilford Press.

McRae, K., Reiman, E. M., Fort, C. L., Chen, K., & Lane, R. D. (2008). Association between trait emotional awareness and dorsal anterior cingulate activity during emotion is arousal-

dependent. *Neuroimage, 41,* 648-655. doi:10.1016/j.neuroimage.2008.02.030

Mossholder, K. W., Settoon, R. P., Harris, S. G., & Armenakis, A. A. (1995). Measuring emotion in open-ended survey responses: an application of textual data analysis. *Journal of Management, 21*, 335-355. doi:10.1016/0149-2063(95)90061-6

Ochsner, K. N., Bunge, S. A., Gross, J. J., & Gabrieli, J. D. E. (2002). Rethinking feelings: An fmri study of the cognitive regulation of emotion. *Journal of Cognitive Neuroscience, 14,* 1215-1229. doi:10.1162/089892902760807212

Pennebaker, J. W. (1997). Writing about emotional experiences as a therapeutic process. *Psychological Science*, *8*, 162-166. doi:10.1111/j.1467-9280.1997.tb00403.x

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. (2008). *The development and psychometric properties of LIWC2007.* Retrieved Dec 29, 2009, from http://www.liwc.net/LIWC2007LanguageManual.pdf

Ribordy, S. C., Camras, L. A., Stefani, R., & Spaccarelli, S. (1988). Vignettes for emotion recognition research and affective therapy with children. *Journal of Clinical Child Psychology, 17,* 322-325. doi:10.1207/s15374424jccp1704_4

Rose, D. (2004). Theory and treatment of alexithymia: An affect theory perspective. *Dissertation Abstracts International: Section B: The Sciences & Engineering, Vol 64(8-B),* 4060.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39,* 1161-1178. doi:10.1037/h0077714

Sattler, J. M. (1992). *Assessment of children: WISC-III and WPPSI-R supplement.* La Mesa, CA: Jerome M. Sattler, Publisher, Inc.

Schultz, D., & Izard, C. E. (1998). The Assessment of Children's Emotion Skills: Facial Expressions subtest. *Unpublished raw data*. University of Delaware, Newark, DE.

Shermis, M. D., & Burstein, J. (Eds.). (2003). *Automated essay scoring: A cross-disciplinary perspective*. Mahwah, NJ: Lawrence Erlbaum Associates.

Simson, U., Martin, K., Schafer, R., Franz, M., & Janssen, P. (2006). Veränderung der Wahrnehmung von Emotionen im Verlauf stationärer psychotherapeutischer Behandlung [German]. = Changes in alexithymia and emotional awareness during psychotherapeutic inpatient treatment. *Psychotherapie Psychosomatik Medizinische Psychologie, 56,* 376-382. doi:10.1055/s-2006-940127

Subic-Wrana, C., Thomas, W., Huber, M., & Koehle, K. (2001). Levels of Emotional Awareness Scale (LEAS): The German version of a new alexithymia test. *Psychotherapeut, 46*, 176-181. doi:10.1007/s002780100157

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. doi:10.1055/s-2002-35281

Subic-Wrana, C., Bruder, S., Thomas, W., Lane, R.D. & Köhle, K. (2005). Emotional Awareness Deficits in Inpatients of a Psychosomatic Ward: A Comparison of Two Different Measures of Alexithymia. *Psychosomatic Medicine, 67,* 483-489. doi:10.1097/01.psy.0000160461.19239.13

Taylor, S. F., Phan, K. L., Decker, L. R., & Liberzon, I. (2003). Subjective rating of emotionally salient stimuli modulates neural activity. *NeuroImage, 18,* 650-659. doi:10.1016/S1053-8119(02)00051-4

Valenti, S., Neri, F., & Cucchiarelli, A. (2003). An overview of current research on automated essay grading. *Journal of Information Technology Education, 2*, 330.

Walgren, M. B. (1996). Alexithymia and levels of emotional development in a substance abuse population. *Dissertation Abstracts International Section A: Humanities & Social Sciences,* Vol 56(8-A), 3065.

Waller, E. & Scheidt, C. E. (2004). Somatoform disorders as disorders of affect regulation: A study comparing the TAS-20 with non-self-report measures of alexithymia. *Journal of Psychosomatic Research, 57*, 239-247. doi:10.1016/S0022-3999(03)00613-5

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children – Third Edition*. Harcourt Assessment, Inc.

Whissell, C. M. (1989). The dictionary of affect in language. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research and experience, Vol. 4, The measurement of emotions* (113-131). New York: Academic Press.

Whissell, C. (2008). Emotional fluctuations in Bob Dylan's lyrics measured by the Dictionary of Affect accompanies events and phases in his life. *Psychological Reports, 102*, 469-483. doi:10.2466/PR0.102.2.469-483

Williams, E. J. (1959). The comparison of regression variables. *Journal of the Royal Statistical Society, Series B, 21,* 396-399.

Yang, Y., Buckendahl, C. W., Juszkiewicz, P. J., & Bhola, D. S. (2002). A review of strategies for validating computer-automated scoring. *Applied Measurement in Education*, *15*, 391-412. doi:10.1207/S15324818AME1504_04

Zwaal, C., Prkachin, K. M., Husted, J., Stones, M. (2003). Components of hostility and verbal communication of emotion. *Psychology & Health, 18,* 261-273. doi:10.1080/0887044021000058067

Table 1

The Six Levels of Emotional Awareness in Hand Scoring

Level	Description	Example Response
0	cognitions	I would expect him to help me.
1	bodily sensations	I would feel pain.
2	action tendencies	I would cry.
3	single emotions	I would feel angry.
4	blends of emotion	I would feel happy but guilty.
5	combinations of blends	I would feel sad and frightened. My friend would feel sympathetic and relieved.

Table 2

POES 1.2.2 Scoring Methods

Name	Description	Example Score[a]
All-Sum	Calculates the sum of all the values in the Valuables List for an item	11
Highest-4	Calculates the sum of the four highest values in the Valuables List	10
334	Searches the Valuables List for all valuables with a value of 3. If these are all identical, an item score of 3 is assigned. If any two valuables are not identical, an item score of 4 is given. If there are no valuables with a value of 3, the item score is the maximum value found in the Valuables List.	4
3345	Self and other Valuables Lists are scored separately using the 334 method. An item score of 5 is given if the self and other scores are both 4; otherwise, the item score is the maximum of the self and other scores. This method requires separate responses for self and other.	3

a. Score for the response "I would be in pain (1) and would probably start crying (2), but I would also be angry (3)." "My friend would feel a bit guilty (3) about making me cry (2), but would not understand (0) my reaction." The value for each valuable that appears in the LEAS Wordlist is given in parentheses.

Table 3

Internal Consistency

Scoring Method	Coefficient Alpha+	95% Confident Interval++
Sample 1 (Adult Community Members), Paper LEAS		
POES Highest-4	.89**[a]	[.84, .93]
POES All-Sum	.89**[a]	[.83, .93]
POES 334	.81**[b]	[.72, .88]
Hand Scoring	.84**[b]	[.77, .90]
Sample 2 (Children), Paper LEAS-C		
POES Highest-4	.78**[a]	[.68, .86]
POES All-Sum	.76**[b]	[.65, .84]
POES 334	.60**[c]	[.41, .74]
Hand Scoring	.66**[bc]	[.51, .78]
Sample 3 (Adult Students), Computerized LEAS		
POES Highest-4	.91**[a]	[.88, .94]
POES All-Sum	.92**[a]	[.89, .94]
POES 334	.79**[b]	[.71, .86]
POES 3345	.86**[c]	[.81, .90]
Hand Scoring	.88**[c]	[.83, .92]

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

+ The values of coefficient alpha were compared to each other using the Pitman technique recommended by Feldt (1980). Coefficients with different superscripted letters were significantly different using $p < .05$.

++ 95% confidence intervals were calculated using the Feldt (1965) technique.

Table 4

Correlations of POES Total Scores with Hand Scoring Total Scores

POES Method	Correlation[+]	CI for Correlation[++]	Correlation when Corrected for Attenuation[+++]	CI for Corrected Correlation[++++]
Sample 1 (Adult Community Members), Paper LEAS				
Highest-4	.93**[a]	[.87, .96]	.91**	[.66, 1.07]
All-Sum	.89**[b]	[.81, .94]	.91**	[.66, 1.07]
334	.92**[ab]	[.86, .96]	.90**	[.60, 1.10]
Sample 2 (Children), Paper LEAS-C				
Highest-4	.61**[a]	[.40, .76]	.65*	[.18, 1.02]
All-Sum	.61**[a]	[.40, .76]	.65*	[.18, 1.04]
334	.69**[a]	[.51, .81]	.84*	[.23, 1.34]
Sample 3 (Adult Students), Computerized LEAS				
Highest-4	.80**[a]	[.68, .87]	.87**	[.69, 1.01]
All-Sum	.76**[b]	[.64, .85]	.82**	[.61, .96]
334	.80**[a]	[.70, .88]	.97**	[.74, 1.13]
3345	.86**[c]	[.79, .92]	.98**	[.81, 1.10]

** $p < .001$.

+ Correlations with different superscripted letters are significantly different using $p < .05$ using Williams' (1959) T2 statistic.

++ 95% confidence intervals were calculated using the Fisher r-to-z transformation as described in Howell (2007).

+++ We corrected these correlations for attenuation due to lack of internal consistency. Because POES scoring and hand scoring shared the same items, to correct for lack of internal consistency we used a procedure recommended by Leonard Feldt (personal communication, Feb 2006). For each POES method, we first calculated the correlation between Form A for hand scoring and Form B for POES scoring, and corrected the correlation using the internal consistencies of those two half-length tests. This provides an unbiased estimate of the correlation when it has been corrected for attenuation due to lack of internal consistency in the two tests. Second, we calculated the correlation for Form B for hand scoring and Form A for POES scoring, and corrected the correlation using the internal consistencies of those two half-length tests. Finally, we averaged the two corrected correlations.

++++ 95% confidence intervals were calculated using a similar procedure. First, we calculated the confidence interval for the correlation between Form A for hand scoring and Form B for POES scoring, using the Fisher r-to-z method recommended by Howell (2007). We then dis-attenuated the end-points. Next, we calculated the confidence interval for the dis-attenuated correlation between Form B for hand scoring and Form B for POES scoring. Finally, we averaged the endpoints of the two confidence intervals.

Table 5

Associations with Cerebral Blood Flow

Scoring Method	Area with Highest Association	
	Coordinates	z-score
Highest-4	0, 12, 32	3.41**
All-Sum	0, 12, 32	3.59**
334	0, 12, 32	3.59**
Hand Scoring	0, 12, 32	4.04**

** $p < .001$.

Table 6

Correlations between LEAS-C and Criterion Variables – Sample 2 (Children)

Criterion Variable	Scoring Method	Correlation	95% Confidence Interval
Emotion Expressions			
	POES Highest-4	.46**[a]	[.21, .65]
	POES All-Sum	.40*[b]	[.14, .61]
	POES 334	.29*[abc]	[.02, .53]
	Hand Scoring	.15[c]	[-.13, .41]
Emotion Comprehension			
	POES Highest-4	.29*[a]	[.01, .52]
	POES All-Sum	.27+[a]	[-.01, .51]
	POES 334	.28*[a]	[.00, .51]
	Hand Scoring	.28+[a]	[.00, .51]
Vocabulary			
	POES Highest-4	.46**[a]	[.21, .65]
	POES All-Sum	.44*[a]	[.19, .64]
	POES 334	.38*[a]	[.12, .60]
	Hand Scoring	.31*[a]	[.03, .54]

+ $p < .10$. * $p < .05$. ** $p < .001$.

Note. Correlations with different superscripted letters were significantly different using $p < .05$ using Williams' (1959) T2 statistic. Confidence intervals for the correlations were calculated using the Fisher r-to-z method recommended by Howell (2007).