Exploration of a new tool for assessing emotional inferencing after traumatic brain injury

Barbra Zupan, Ph.D.[1]

Dawn Neumann, Ph.D.[2]

Duncan R. Babbage, Ph.D.[3]

Barry Willer, Ph.D.[4]

[1]Associate Professor, Brock University, Department of Applied Linguistics
St. Catharines, ON L2S 3A1 ORCID: 0000-0002-4603-333X
Email: bzupan@brocku.ca
Phone: 905-688-5550 ext. 3704

[2] Assistant Research Professor, Indiana School of Medicine, Department of Physical Medicine and Rehabilitation
Clinical Research Associate, Rehabilitation Hospital of Indiana
Indianapolis, IN 46254

[3] Associate Professor, Auckland University of Technology, Centre for eHealth
Auckland, New Zealand. ORCID: 0000-0002-9259-9246

[4] Professor, State University of New York at Buffalo, School of Medicine, Department of Psychiatry
Buffalo, NY

Key Words: emotional inferencing, context, traumatic brain injury

Word Count: 6924

Abstract

Objective: To explore validity of an assessment tool under development—the Emotional Inferencing from Stories Test (EIST). This measure is being designed to assess the ability of people with traumatic brain injury (TBI) to make inferences about the emotional state of others solely from contextual cues.

Methods and Procedures: *Study 1*: 25 stories were presented to 40 healthy young adults. From this data, two versions of the EIST (EIST-1; EIST-2) were created. *Study 2*: Each version was administered to a group of participants with moderate-to-severe TBI—EIST 1 group: 77 participants; EIST-2 group: 126 participants. Participants also completed a facial affect recognition (DANVA2-AF) test. Participants with facial affect recognition impairment returned two weeks later and were re-administered both tests.

Main Outcomes: Participants with TBI scored significantly lower than the healthy group mean for EIST-1, $F(1,114)=68.49$, $p <.001$, and EIST-2, $F(1,163)=177.39$, $p <.001$. EIST scores in the EIST-2 group were significantly lower than the EIST-1 group, $t=4.47$, $p <.001$. DANVA2-AF scores significantly correlated with EIST scores, EIST-1: $r=.50$, $p<.001$; EIST-2: $r=.31$, $p<.001$. Test-retest reliability scores for the EIST were adequate.

Conclusions: Both versions of the EIST were found to be sensitive to deficits in emotional inferencing. After further development, the EIST may provide clinicians valuable information for intervention planning.

Introduction

The context of a situation influences emotion perception during every day social interactions [1–3]. To successfully interact with others, we continually integrate available contextual cues with language to make more accurate inferences about the thoughts, feelings and intentions of others [4–6]. In addition to comprehension of the words and sentences used, this pragmatic skill requires the perceiver to consider more than one possible meaning of the message, continually checking that the inference generated is consistent with the current communicative context [4,7,8].

Social communication is a complex process that requires considerable attention, working memory, information processing, and executive function capabilities (i.e., decision making, problem solving, goal formulation, behaviour monitoring)[9]. It should not be surprising then, that despite the heterogeneity of their injuries, people with traumatic brain injury (TBI) are frequently known to have social skill deficits, including problems recognizing nonverbal emotion cues (i.e., facial and vocal expressions), and interpreting the intended meaning of a message [6,10–20]. Deficits occur even when language comprehension remains relatively intact [5]. Simply understanding the literal meaning provided in a speaker's message is insufficient for successful social interaction; perceivers must be able to additionally integrate a range of available social and emotional cues that allow them to generate appropriate inferences about the situation, thoughts and feelings of the speaker, and the intended meaning of the message [5].

Few studies have specifically examined the ability of people with TBI to infer emotion using contextual information. To date, studies have focused primarily on tasks designed to assess 'theory of mind' (ToM) − the ability to understand and predict the behaviour of others based on inferences made about their intentions and beliefs [21]. One common paradigm used to assess ToM

is the false-belief task. False-belief stories generally describe an object being moved or changed without the character's knowledge. Participants are then asked to explain one of the character's actions or respond to a question that measures their understanding that one character in the story holds a mistaken belief [9,22]. Bibby and McDonald (2005) have shown that people with TBI perform worse than Controls on false belief tasks, suggesting that they have difficulty making general inferences when given contextual information. The task did not assess how well they were able to make inferences about other's emotions.

The ability to recognize a social faux pas is considered a more complex ToM task. A social faux pas occurs when a speaker says something that should not have been said because it may be perceived as hurtful or offensive by the listener [23]. The Faux Pas test consists of 20 vignettes, half of which describe a social faux pas. The questions that follow each vignette are aimed to assess understanding of the intentions and beliefs of the characters in those vignettes (e.g., Did someone say something they shouldn't have said? Why shouldn't they have said it?)[24]. As such, the test is more directly aimed at assessing the ability to make inferences about verbal behaviour within a social context, rather than the ability to infer the emotional state of others. Numerous studies have used the Faux Pas test with people with TBI and have consistently reported poorer understanding and recognition of social faux when compared to Controls [9,21,25,26].

The Awareness of Social Inference Test (TASIT) [15,27,28] was specifically developed for people with TBI to measure different aspects of social perception using ecologically valid stimuli. The TASIT is a video-based measure that includes a series of short vignettes that show actors interacting in everyday settings. Part 1 of the TASIT assesses the ability to recognize emotions using primarily nonverbal (i.e., facial and vocal expressions) information in context-neutral situations. Parts 2 and 3 are focused on conversational inference abilities, specifically the

ability to differentiate sincere versus sarcastic interactions, and differentiate lies versus sarcasm, respectively[15]. Part 3 is the only subtest that provides contextual information; however, participants are not asked to infer feelings. The TASIT has been shown to be a sensitive and reliable measure for differentiating people with TBI from their uninjured counterparts on ToM and conversational inference (i.e., sarcasm, lies) tasks [27,29]. However, similar to other ToM measurements, the TASIT does not specifically measure the ability to make inferences about how others may be feeling within a specific context.

In an attempt to meet the need for a test that specifically measures social inference in context, Turkstra [6] developed the Video Social Inference Test (VSIT). This test was modeled after the TASIT and includes 16 short vignettes ($\leq$ 30 seconds) that portray actors engaged in a conversation exchange. The situations were designed to be relevant to adolescents and young adults, so portray only peer relationships. Vignettes are paired, allowing social inference ability to be assessed in two stages; the first vignette provides important context for the second. Thus, the VSIT aims to assess one's ability to generate an inference given social contextual cues, but also one's ability to use that inference to explain behaviours that occur later [6]. Turkstra [6] administered the VSIT to 19 people with TBI and 19 matched Controls. Her results showed this test appeared to be sensitive to deficits in social inference abilities for people with TBI.

Although Turkstra's [6] VSIT appears to successfully identify difficulties in social inference abilities by people with TBI, the video-based format of this test intrinsically adds additional cues. Turkstra [6] hypothesized that the addition of body, facial, and vocal cues supplied by the video format would positively contribute to performance on the VSIT task. To test this hypothesis, she compared performance on the VSIT to performance on the 'Reading the Mind in the Eyes' Test [30], a measure that assesses one's ability to infer someone's thoughts using only

their eyes. Indeed, participants were better at inferring information on the context-enriched VSIT task compared to the minimalistic Eyes Test. Turkstra [6] concluded that the increased performance was due to the additional cues provided through the video format of the VSIT task. However, despite this improvement, participants with TBI still scored significantly lower on the VSIT compared to Controls, indicating that they still had a disadvantage even with the additional cues. Because this study did not run separate conditions to solely test participants' ability to make inferences using context alone, it could not be determined if the disadvantage was due to problems recognizing nonverbal cues (e.g. facial and/or vocal affect), or problems interpreting the context, or both.

Given the previous discussion, it appears that further exploration of emotion inference abilities by people with TBI is needed. Few tools exist that specifically measure the ability to make inferences when given social context. The VSIT is reported to be sensitive to social inference impairments of people with TBI, but questions remain regarding the source of impairments: whether from inferring the emotion from nonverbal cues, inferring the emotion from context, integrating all available cues (i.e., body, face, voice, verbal language, situation), or some combination. In order to determine this, it is important to have an instrument that assesses emotion inference from contextual information that is separate from nonverbal and verbal cues. Results of such a test have important implications for intervention, since having a better understanding of which cues most contribute to emotion inference impairments, would allow for more focused treatment.

The current paper describes the development and evaluation of a measure designed to assess a person's ability to make inferences about others' emotions when given contextual cues, but no nonverbal cues. Study 1 describes the development and initial testing of the EIST

instrument with a normal subject sample of young adults. Study 2 describes trials with people

with TBI using two separate versions of the test, comparing their performance to the normative

sample to provide some preliminary evidence of the sensitivity of each version. In our TBI

sample, we examined concurrent validity with a test of facial affect recognition (emotional

inference from nonverbal cue). We also examined the test-retest reliability of each version of the

EIST.

Study 1 – Stimuli Development

Methods

Participants

40 participants, eight males and 32 females, were recruited at the first author's host

university. All participants reported they had no history of speech and/or language delays, nor

diagnosed head injury. The average age of participants was 21.86 years (s.d. = 4.24; range =17-

44). All participants used English as their primary language.

Measures

Emotion inference from stories test (EIST).

Stimuli included a total of 25 written short stories describing contexts that targeted *happy*

(6), *sad* (6), *angry* (6), and *fearful* (7) emotions (see table 1). Stories were between 70 and 256

words (mean=170.36; s.d.=48.08) and averaged just below a sixth grade reading level

(mean=5.97; range=4.3-8.6) using the Flesch-Kincaid Grade Level rating system [31]. All stories

were written about topics that were considered relevant for adults (e.g., starting a new job; a

friend canceling plans; struggling to make ends meet financially). Each story was followed by a

single question that inquired about how one of the characters in the story was feeling given the

situation described. Accurate interpretation of the targeted emotion experienced by the character

required integration of the available contextual cues, including the situation or event, the

character's wants, beliefs and expectations as these related to the situation described, and the

character's behaviour in response to that situation or event (see appendix for an example).

Reponses were scored as either correct or incorrect and each item was analysed separately.

Procedure

Participants were presented with the 25 short stories and follow-up questions regarding

how a character in that story was feeling on a computer. The story was presented first, with the

follow-up question presented afterwards on the next screen; participants were not allowed to re-

read the story to answer the question. Stories were presented in a random order using Cedrus

SuperLab software [32]. The text for each story was displayed on a single screen in 23-point Arial

font. The follow-up question was displayed at the top of the screen, with the five response

options (*happy, sad, angry, fearful, I don't know*) listed directly below the question. After

reading the question, participants entered a response using a forced-choice response key-pad.

<div align="center">

Results

</div>

Summary data for the emotion inference stories are presented in table 1. The percent of

participants who selected the target emotion for each of the 25 stories ranged from 62 to 100%

(mean=88.3; s.d.=0.06). Overall, *happy* (95%) was the emotion category most consistently

identified by participants, followed by *angry* (90%), *fearful* (86%), and *sad* (84%).

<div align="center">-- INSERT TABLE 1 ABOUT HERE --</div>

Although all stories created for the EIST targeted the intended emotion well above

chance level (i.e., 25%), we were concerned that presenting such a lengthy test to people with

TBI would negatively impact responses due to fatigue and attention issues. Instead, we created

two versions of the EIST using item-level data from the healthy group. Each version contained

twelve stories, three stories from each emotion category. Stories chosen within each version represented a wide array of themes. As discussed previously, little is known about the ability of people with TBI to make inferences about others' emotions when given only contextual cues. Even less is know about their ability to do this when given narrative scenarios that extend beyond a few brief sentences. Given this, we decided to vary the difficulty of the two test versions, creating one version in which the target emotion should be more easily inferred, and a second version in which the target emotion should be more difficult to infer.

The first version of the Emotion Inference from Stories Test (EIST-1) included only stories identified as the target emotion more than 90% of the time. The twelve stories selected for EIST-1 are shaded in gray in Table 1. The second version, EIST-2, included stories that had been correctly identified between 70 and 100% of the time (see table 1). The resulting healthy group mean for stories selected for the EIST-1 was 11.75 (s.d.=0.49) and the average reading grade level was 6.05. The healthy group mean for stories selected for the EIST-2 was 11.28 (s.d.=0.90) which was significantly lower, $t=4.00$, $p < .001$, indicating that this version was more challenging. The average reading grade level for the EIST-2 (5.89) did not differ significantly from EIST-1, $t=0.32$, $p=.75$.

Study 2 –Performance of People with TBI for Emotional Inferencing from Stories

The purpose of Study 2 was to collect preliminary data concerning the validity (discriminant and concurrent) and test-retest reliability of two alternative versions of the Emotion Inference for Stories Test (EIST) with two groups of people with TBI. People with TBI were recruited as part of a larger randomized control trial of treatment approaches to target affect recognition deficits following [33]. Only measures relevant to the current study are reported here. The hypotheses for Study 2 were as follows:

1) Participants with TBI would have lower scores on each version of the EIST compared to the typically developing participants included in Study 1. This finding would provide support for the discriminant validity of the EIST.

2) A significant difference between EIST-1 participants with TBI compared to EIST-2 participants with TBI. Since EIST-2 was designed to be more challenging, we expected that EIST-2 participants with TBI would have significantly lower scores than EIST-1 participants with TBI. We expected similar performance on all remaining measures.

3) A significant relationship between facial affect recognition abilities and EIST scores. This finding would lend support for convergent validity for the EIST since facial affect recognition impairment occurs frequently following moderate-severe TBI [11,34].

4) A significant correlation between scores on both versions of the EIST for Session 1 and Session 2, demonstrating test-retest reliability.

5) Since the overall purpose of this study was to explore the validity of the EIST as a potential measure for assessing the ability of people with TBI to make accurate inferences about how others are feeling when given only contextual cues, classical test theory was also used to assess item quality. We predicted that the majority of items would fall within a moderate range of difficulty.

Methods

Participants

As part of a larger randomized control trial of treatment approaches targeting affect recognition deficits following TBI[33], participants were recruited from out-patient brain injury rehabilitation centers and local brain injury support groups in the southern regions of Ontario, Canada; North and South Carolina, USA; and in Wellington and Palmerston North, New

Zealand. Ethics approval was received from all participating institutions/regions and all participants were required to provide written consent prior to participation. Participants were recruited on the basis of injury type and severity, not on the basis of reported social or emotion recognition difficulties.

To participate in the study, participants had to be between 18 and 65 years of age, and must have sustained a moderate to severe TBI at least one year prior. Injury severity was determined by one of the following: Glasgow Coma Scale (GCS) score of ≤ 12 at the time of injury; loss of consciousness (LOC) of ≥ 24 hours, or posttraumatic amnesia (PTA) of ≥ 24 hours. Exclusion criteria: 1) current substance dependence; 2) premorbid developmental affective or acquired (nontraumatic) neurological disorder (e.g., autism spectrum disorder, stroke); 3) premorbid major psychiatric disorder (e.g., schizophrenia); or 4) uncorrected or impaired vision/hearing that would prohibit full participation in the experimental tasks.

The EIST-1 group consisted of 77 participants with moderate to severe TBI, the majority of whom were Caucasian (81.8%). The study sample included 53 males and 24 females ranging in age from 21-62 years. On average, participants had achieved 13.29 years of education and were 10.95 years post injury. Based on GCS, LOC, and/or PTA information, the majority of participants were classified as having a severe TBI (98.6%). Motor vehicle accidents were the most common cause of injury (68%). A total of 27 participants from group one returned for a second session.

The EIST-2 group included 126 participants, almost all of whom were classified as having severe TBI (97.6%). The sample was primarily Caucasian (81%) and included more males (n=97) than females (n=29). The EIST-2 group was between 21 and 65 years of age and had achieved an average of 12.9 years of education. Participants were on average, 9.88 years

post-injury and motor vehicle accidents were the most common cause of injury (58.7%). A total

of 46 participants in the EIST-2 group returned for a second session.

Table 2 details information on participant characteristics for EIST-1 and EIST-2 groups

(sessions one and two).

-- INSERT TABLE 2 ABOUT HERE --

Measures and Procedures

As part of the larger clinical trial, participants attended a screening session in which a

series of measures were administered. Participants with impaired affect recognition returned for

a second session two weeks later, when selected tests were re-administered. For the measures

relevant to the current study, tests were administered on a computer and participants indicated

their responses orally and/or via pointing to printed text. Responses were recorded by the

examiner. All participants were seen individually.

Discourse comprehension task.

Discourse comprehension was assessed in the first session using the Discourse

Comprehension Test (DCT)[35]. This measure was included to ensure that participants had

adequate discourse comprehension of non-emotional language to complete the EIST task. The

standardized form of the DCT includes two sets of five stories (~200 words in length), each

followed by eight yes/no questions. The questions assess comprehension of stated and implied

main ideas and details. For the current study, participants were administered two stories from the

DCT. Participants read one of these stories to themselves (i.e., self-read) and received the other

in auditory format (pre-recorded); the order of administration for self-read versus auditory format

was counterbalanced. Participants were not allowed to return to the story to answer the yes/no

questions that followed. The number of overall errors (maximum=8), errors regarding

information stated in the story, (maximum=4), and errors for implied information (maximum=4) were calculated for each story. Participants had to achieve 75% on either the self-read or auditory version of the story to be eligible to return for the second testing session. All participants met this criterion.

Facial affect recognition task.

Facial affect recognition abilities were assessed in both sessions using the Adult-Faces subtest of the Diagnostic Assessment of Nonverbal Accuracy–2 (DANVA2-AF)[36]. Twenty four faces were presented via computer for 15 seconds each and participants were instructed to select which emotion was being expressed from a list of four options: *happy, sad, angry*, or *fearful*. Faces in the DANVA2-AF vary in sex, race and intensity of expression. The DANVA2-AF is a standardized measure that has been used with people with TBI in previous studies [13,20,33,37–39]. Each item was scored as either correct or incorrect to a maximum score of 24. Participants who scored below the normative age mean for the DANVA2-AF were re-administered this test in a second session two weeks later.

Emotion inference from stories test (versions 1 and 2).

Participants were assigned to one of two groups. The first group of participants were administered the EIST-1 items and the second group were administered EIST-2 items. Participants who scored below the normative mean on the DANVA2-AF and met the DCT criteria during the first session, returned for a second session and were re-administered the EIST. Participants received the same version of the EIST for both sessions.

To ensure that participant responses were not negatively affected by challenges in reading and processing written text, the stories were presented in both auditory and written format. For the auditory version, a female speaker was recorded while reading each of the stories. Stories

were recorded at a moderate speaking rate. In order to make the reading sound natural, some

intonation was used. However, in recording the stories, the speaker did not emphasize key words

or content in the story, nor did she use intonation that would add nonverbal cues about how the

character was feeling (e.g., the speaker did not raise the volume of her voice if the situation was

meant to portray anger). The follow-up question for each story was similarly recorded.

During the EIST administration, the text for each story was displayed on a single screen

in 23-point Arial font. In the corner of the screen was a small icon that played the audio for the

story when clicked. After participants finished reading and listening to the story, they were

brought to the next screen where the follow-up question was displayed at the top of the screen in

the same fashion as the story. Participants were asked to select a response from five response

options (*happy, sad, angry, fearful, I don't know*) that were listed directly below the question.

Each version of the EIST presented stories in the same restricted randomized order (by

targeted emotion) to ensure that no stories targeting the same emotion occurred consecutively. A

different restricted randomized order was created for the second administration in Session 2.

Data Analyses

Descriptive statistics were calculated for all demographics and injury related variables to

identify any important covariates using Chi-square tests and two-tailed independent *t*-tests. No

significant group differences were found (see table 2).

We compared mean percentage accuracy scores for EIST-1 and EIST-2 TBI groups to the

healthy group in Study 1 using one-way analysis of variance ($\alpha < .05$). A chi-square was

conducted to compare the number of participants in the EIST-1 group who scored below 2

standard deviations of the normative mean on the EIST to participants in the EIST-2 group.

Between group analyses were conducted for all three tests administered at Session 1:

DCT ($\alpha < .017$); DANVA2-AF ($\alpha < .05$); EIST ($\alpha < .05$) using mean percentage accuracy scores.

Within group paired samples *t*-tests were also conducted to compare response accuracy for the

auditory versus self-read versions of the DCT ($\alpha < .017$). Spearman correlations ($\alpha = .017$) were

calculated between DANVA2-AF scores and the EIST for each group. Additionally, Spearman

correlations were conducted using the normative scores for EIST-1 and EIST- 2 ($\alpha = .05$).

For participants who had significant facial affect recognition impairment and returned for

a second session, Spearman correlations ($\alpha <.05$) were conducted to determine the relationship

between their EIST scores from Session 1 to Session 2.

Finally, Classical Test Theory was applied to the data collected in Session 1 in order to

analyse item difficulty, item discrimination and reliability for each version of the EIST for

people with TBI.

Results

TBI Performance on EIST Compared to Healthy Group

The first study hypothesis was that participants with TBI would show significantly poorer

performance on each version of the EIST compared to the typically developing participants in

Study 1. This hypothesis was supported (see figure 1). People with TBI had significantly lower

scores than the healthy group from Study 1 for both EIST-1, $F(1,114)=68.49$, $p < .001$; and

EIST-2, $F(1, 163)=177.39$, $p <.001$.

-- INSERT FIGURE 1 ABOUT HERE --

Using the healthy group data from Study 1, all participants were identified as having

either impaired or unimpaired performance on the EIST. Any participant who scored more than 2

standard deviations from the healthy group mean was categorized as impaired. Only one

participant (2%) in the healthy group fell into the impaired category. However, 78% of

participants with TBI in the EIST-1 group (i.e., 59 participants) and 93% of participants in the

EIST-2 group (i.e., 116 participants) fell more than 2 standard deviations below the healthy

group mean. The difference in identification of impairment via EIST-1 and EIST-2 was

significant, χ^2 = 9.6, p = .004.

<center>-- INSERT TABLE 3 ABOUT HERE --</center>

EIST-1 versus EIST-2

  The second study hypothesis was that participants in the EIST-1 group would have

significantly higher scores on the EIST than participants in the EIST-2 group, but performance

on all other measures would be equal. This hypothesis was only partially supported. Summary

data from Session 1 are shown in table 3 for EIST-1 and -2 groups. On average, participants in

the EIST-1 group scored 8.49 (s.d.=2.46; range= 0-11), and participants in the EIST-2 group

scored 7.01 (s.d.=1.96; range=0-10). The difference between group scores was significant,

t=4.47, p <.001, indicating that EIST-2 items were significantly more difficult than EIST-1

items.

  For the DCT, no significant group difference was found in response to the self-read story

but Group 2 had significantly higher scores than Group 1 for the auditory version. To determine

whether these scores may be due to better auditory than reading comprehension for people in

Group 2, paired samples t-tests were conducted. Group 2's scores for the auditory version of the

DCT story were not significantly higher than the self-read version: Total score: t=.18, p=.86;

Direct questions: t=1.93, p=.06; Implied questions: t=1.57, p=.12. Paired samples t-tests

conducted for Group 1 also showed that the format in which the DCT story was delivered did not

significantly affect total score, $t=1.25$, $p=.21$; responses to direct questions, $t=1.57$, $p=.12$; or responses to indirect questions, $t=.12$, $p=.90$.

As hypothesized, facial affect recognition (DANVA2-AF) was not significantly different between EIST-1 and EIST-2 groups. Using the normative age data specified in the DANVA2-AF manual [36], 27 of the 77 participants (35%) in the EIST-1 group and 46 of the 126 participants (36%) in the EIST-2 group were identified as having facial affect impairment and returned for a second session two weeks later. The number of participants identified as having facial affect recognition impairment did not significantly differ between groups, $\chi^2 = .043$, $p = .881$.

EIST Convergent Validity with Facial Affect Recognition

The third study hypothesis was that there would be a significant relationship between EIST performance and facial affect recognition. Performance on the DANVA2-AF was significantly correlated with scores on the EIST for participants in the EIST-1 group, $r = .50$, p < .001 and the EIST-2 group, $r = .31$, p < .001.

Test-Retest Reliability of the EIST.

Participants in the EIST-1 and -2 groups identified as having impaired facial affect recognition returned two weeks later for a second session in which the EIST was re-administered. We hypothesized that this sub-set of participants would show similar scores on the EIST between Sessions 1 and 2; this hypothesis was supported. Participants in the EIST-1 group who returned for Session 2 (n=27) had an average score of 7.41 (s.d. = 2.29) at retest. Although this score was higher than the score obtained from this same subset of participants in Session 1 (mean=7.07, s.d.=2.73), the test scores for the EIST-1 group for Sessions 1 and 2 were significantly correlated, $r = .76$; $p < .001$. The 46 participants from the EIST-2 group who completed the second session also increased their scores from Session 1, achieving a mean score

of 6.53 (s.d.=2.30) in Session 2 compared to their previous score of 6.33 (s.d.=2.23). EIST-2

scores for Session 1 and Session 2 were also found to be significantly correlated, $r = .72$,

$p < .001$. The significant between group difference for EIST-1 versus EIST-2 found in Session 1

was not found in Session 2. Although participants in the EIST-1 group continued to have lower

scores than participants in the EIST-2 group, the difference between versions was no longer

significant, $t = 0.98$, $p = .33$.

Classical Test Theory Analyses of the EIST-1 and EIST-2.

In order to evaluate the quality of the items included in each version of the EIST, data

from Study 2, Session 1 was used to compute item difficulty (p), discrimination index (D), item-

test correlation and reliability using classical test theory. The proportion of participants who

correctly identified the target emotion (*p*) for items included in each version of the EIST is

shown in table 4. Table 4 shows that the proportion of items considered low in difficulty (i.e.,

$P \geq .80$) for EIST-1 (33%) was greater than EIST-2 (17%). All but one of the items targeting

happy were low in difficulty and only one *sad* item was ranked easy. The remainder of items on

both versions of the test were considered moderate in difficulty (i.e., $P = .40-.80$), and thus

acceptable items [40,41]. Neither version of the EIST included items that were considered difficult

(i.e., $P \leq .40$). To determine the extent to which items discriminated between participants with

high versus low emotional inferencing abilities, a discrimination index was calculated using the

upper and lower 25% of participants for each version of the test. As shown in table 4, all items

had a positive discriminating power and the majority of items had a discrimination index of at

least .20 [41]. The two items found to have poor discriminating power (Stories 2 and 19) were in

EIST-2 and were also found not to correlate with the overall test score ($\alpha < .004$). Cronbach's

alpha coefficient for EIST-1 (.761) indicated acceptable internal consistency of the items.

However, the internal consistency for EIST-2 was more questionable, Cronbach's alpha = .631

[42].

Discussion

The primary aim of this study was to evaluate the effectiveness of a novel measure for emotional inferencing for use with people with TBI. Both versions of the Emotional Inferencing from Stories Test (EIST) aim to assess the ability of people with TBI to make inferences when given social context with no accompanying nonverbal emotion cues. The focus of Study 1 was to evaluate the stories we had written for inclusion in the EIST and ensure that those stories were indeed targeting the intended emotion. Twenty of the 25 stories were easily identified by our healthy group (≥75%). The remaining five stories were correctly identified 62-72% of the time. Since these rates of identification were still well above chance (25%), it appeared that the contextual cues in these stories were still accurate but more subtle, making the task of inferring how others were feeling within that context more challenging. Two versions of the EIST were created; one in which the targeted emotion was correctly inferred by our healthy group at least 90% of the time and a second, more challenging version, in which the target emotion was correctly identified at least 70% of the time.

The purpose of Study 2 was to characterize performance of people with TBI on each version of the EIST and to explore the quality of the items included in each version. As hypothesized, participants with TBI scored significantly lower than the healthy group for both EIST-1 and EIST-2. Thus, while typically developing young adults were able to easily make the

required inferences, particularly for stories in EIST-1, participants with TBI had some trouble

with inferring emotions from the stories. These results are consistent with studies that have

indicated that people with TBI need more direct questions and prompts when engaging in social

interaction [43,44], suggesting that the EIST is sensitive to emotional inferencing deficits in people

with TBI. Using the healthy group mean for each version of the EIST, we identified participants

as having either impaired or unimpaired performance on the EIST. A greater proportion of

participants who completed EIST-2 were identified as impaired (i.e., $\geq$ 2 standard deviations

below normative mean). Since participants in EIST-1 and EIST-2 groups did not significantly

differ in demographics, these findings suggest that EIST-2 may be a more sensitive measure.

Our results comparing EIST-1 and 2 supported our hypotheses that EIST-2 was indeed

more challenging than EIST-1, as demonstrated by significantly lower scores on EIST-2. This

suggests that at least some of the stories in EIST-2 included more subtle contextual cues. It is

important to note that this difference was not due to the fact that the language used in the stories

was more complex since the reading level of the stories in each version of the EIST did not differ

from one another. In fact, the reading level of EIST-2 was slightly lower than EIST-1.

Participants in EIST-1 and 2 groups did not significantly differ from one another in their

facial affect recognition abilities (DANVA2-AF) and scores for both groups were found to

significantly correlate with EIST performance, supporting the convergent validity of the EIST.

Similar to reports in the literature [11], we found that 35% of participants in the EIST-1 group and

36% of participants in the EIST-2 group were impaired in their facial affect recognition abilities.

Moreover, participants identified as having facial affect recognition impairment had significantly

lower scores on the EIST than participants with TBI without facial affect recognition

impairment, thus supporting the convergent validity of the EIST. It is possible that people with

TBI have a general dysfunction in emotion recognition [14,17] that extends to contextual cues. Given that emotion recognition deficits have been reported to potentially underlie the social and behavioural changes often reported following TBI [45], this is an important finding because it suggests that in addition to assessing recognition of nonverbal emotion cues following injury, we should also be evaluating how well people with TBI are able to use of verbal and contextual cues to make emotional inferences within social settings.

Finally, we applied classical test theory to each version of the EIST to explore item difficulty, item discrimination and reliability. Overall, this data showed that the majority of items on each version of the EIST fell within an acceptable difficulty level (.40 to .80), and had good discriminating power ($\geq$ .20)[41]. According to Downing[41], items are considered high in quality if they have both an item difficulty between .45 and .75 and an item discrimination of .20 or higher. Five items in EIST-1 met this criteria and six items in EIST-2 (see table 4). While these items included at least one story targeting *sad*, *angry*, or *fearful*, none of these items targeted *happy*. In reviewing the item difficulty analysis, it seems that for all but one of the stories targeting *happy*, participants could easily infer what the character in that story was feeling. Some literature suggests that this may be because the identification of happy remains relatively intact following TBI [15,17,46–51]. However, one recent study has reported that people with TBI who have impaired affect recognition are also impaired for their recognition of happy [39]. Another potential explanation lies in the valence of the stimuli. Happy was the only positively-valenced emotion targeted in this study. While this is typical of many studies in emotion-based tasks, it is possible that participants are first assessing the contextual cues on the basis of valence. Since there was only one positive emotion targeted in the EIST, participants were able to easily infer this emotion

through the process of elimination. However, if they found the story to be negatively-valenced in content, the task of inferring a specific emotion was more difficult.

Both EIST-1 and EIST-2 were found to have good test-retest reliability as indicated by the strong correlations found in scores from Session 1 to Session 2. This finding suggests that both versions of the test are stable. While the majority of data in Study 2 seem to suggest that EIST-2 was a more sensitive measure than EIST-1, the internal reliability for EIST-2 was more questionable. This was likely due to the fact that two items included in the EIST-2 had both poor discriminating power and no correlation with the overall test score. Given that the test includes only 12 items overall, three targeting each of the emotion categories, two poor items is certainly problematic. In addition, when we considered only scores for people found to be impaired in facial affect recognition, the increased difficulty of EIST-2 over EIST-1 was no longer evident.

Limitations and Future Directions

The healthy group from Study 1 were recruited to ensure that the stories included in the EIST were successfully targeting the intended emotion. As a result, we did not set out to specifically match these participants for age and gender to our participants with TBI. Future exploration of the EIST may consider including an age and gender matched control group. In addition, we did not have gender balance for our participants. Given that the literature shows that females score higher on tests of nonverbal decoding of affect [52], emotional awareness [53], and are more efficient in processing multisensory emotion expressions [54], collecting normative information for male versus female performance on the EIST will certainly be a priority in future development of this tool. In addition, since we did not have gender balance in our healthy group, we were unable to examine gender differences between this group and our participants with TBI. Again, such comparisons should be conducted as this tool is developed further.

An additional limitation of the present study was the inclusion of only one positively

valenced emotion in our stimuli. Although many studies of emotion focus on only *happy*, *sad*,

angry, and *fearful* [39,47,55,56], further development of the EIST should consider the inclusion of at

least one more positively valenced emotion. Moreover, creating a test with a greater number of

items within each emotion category would allow for more detailed analyses, potentially

providing important information about whether deficits in emotion inferencing are specific to

negatively valenced items, or whether the deficit is more general in nature. Since the literature

seems to suggest that people with TBI have more difficulty with negatively valenced stimuli, this

is an important question to explore.

The third limitation of the present study was found when completing analyses of the

items themselves using data collected with our participants with TBI. As discussed previously,

the majority of items were found to be of average difficulty, an appropriate outcome. However,

the remaining items were all easy—no difficult items were identified. Future development of the

EIST needs to consider including items with a higher difficulty level in order to better

discriminate between people with and without emotional inferencing deficits. Practice effects

over time should also be explored so that interpretation of results, particularly results pre- and

post-intervention, can be confidently interpreted.

Finally, the EIST was presented to participants in a combined format so participants

could listen and/or read along for each story. Given the common motor and sensory impairments

reported after TBI [57], our goal in presenting both text and pre-recorded audio was to ensure that

participants had full access to the content of the stories regardless of their auditory and visual

processing skills. However, since the frontal areas of the brain are vulnerable to damage

following TBI [58], it is possible that by combining the format we inadvertently increased the

cognitive load of the task, forcing participants to either attempt to attend to all available cues or

to sustain attention to only one while the other was still present. Since people with TBI who

show executive deficits have also been reported to have emotion recognition deficits [59], this is a

reasonable possibility. However, there was no way for us to determine this in the present study.

Future use of the EIST should examine whether test format (i.e., pre-recorded; self-read;

combined) impacts overall performance by presenting the stories in each of the three modalities,

across different testing sessions, to determine if one format consistently leads to increased

scores. Additionally, sensory and attention tests should be conducted to explore their influence

on performance and relationship with EIST scores. If and when the measure is fully standardized

for wider use in the future, clinicians who choose to use EIST in a clinical setting may wish to

consider presenting practice stories in each of the three modalities (i.e., auditory, self-read,

combined) and asking participants to specify their preference.

Conclusions

People with moderate-severe TBI have been reported to have nonverbal affect

recognition impairment (e.g., facial emotion expressions) as well as difficulty using verbal and

situational cues to determine how someone might be feeling. To date, no tool exists that

independently assesses the ability to determine how someone else is feeling when given only

contextual cues. If emotional inferencing is only being assessed using stimuli that also requires

accurate interpretation of nonverbal emotion cues, there is a risk that someone may be identified

as having deficits in this area when if fact the deficit is only for one or more nonverbal channels.

Thus, the creation of a tool that allows us to independently consider how well people with TBI

can make emotional inferences has important implications for intervention. Access to

intervention is limited so we need to ensure that as clinicians, we are not spending intervention time on a skill that is not impaired.

In sum, preliminary use of the EIST with a large number of people with moderate–severe TBI suggests that this tool has met our initial goals and has sufficient promise to warrant further development. The EIST appears to be sensitive to deficits in emotion inferencing abilities as evidenced in the significantly lower scores on each version by people with TBI compared to people without. In addition, while some items were found to be too easy, many items included were of acceptable difficulty and the majority of items showed good discrimination. Further development will require the creation of new, more difficult items and analysis of the influence of test format on performance. On the basis of this initial study, we are optimistic that a future, fully developed version of the EIST may provide clinicians valuable information for intervention planning and additionally lead to advanced study in the cues that contribute to emotion perception for people with TBI. On the basis of this research, further development of the EIST will draw on the highest quality items from the EIST-1 and EIST-2 as well as additional, more difficult items and items that target a broader range of emotion categories. To develop these additional items, we anticipate forming focus groups to identify relevant contexts for people with TBI to inform the writing and validation of new stories. More pilot work investigating the influence of format on test performance and relationship to attention and sensory processing also will assist with ongoing development of this and other similar measures.

Declaration of Interest

References

1. Barrett LF, Mesquita B, Gendron M. Context in Emotion Perception. *Curr Dir Psychol Sci*. 2011;20(5):286-290. doi:10.1177/0963721411422522.

2. Barrett LF, Lindquist KA, Gendron M. Language as context for the perception of emotion. *Trends Cogn Sci*. 2007;11(8):327-32. doi:10.1016/j.tics.2007.06.003.

3. Lindquist KA, Gendron M. What's in a Word? Language Constructs Emotion Perception. *Emot Rev*. 2013;5(1):66-71. doi:10.1177/1754073912451351.

4. Martin I, McDonald S. Weak coherence, no theory of mind, or executive dysfunction? Solving the puzzle of pragmatic language disorders. *Brain Lang*. 2003;85(3):451-466. doi:10.1016/S0093-934X(03)00070-1.

5. Mcdonald S. Editorial Putting communication disorders in context after traumatic brain injury. *Aphasiology*. 2000;14(4):339-347.

6. Turkstra LS. Conversation-based assessment of social cognition in adults with traumatic brain injury. *Brain Inj*. 2008;22(5):397-409. doi:10.1080/02699050802027059.

7. Ferstl EC, Rinck M, von Cramon DY. Emotional and temporal aspects of situation model processing during text comprehension: an event-related fMRI study. *J Cogn Neurosci*. 2005;17(5):724-39. doi:10.1162/0898929053747658.

8. Channon S, Watts M. Pragmatic language interpretation after closed head injury: relationship to executive functioning. *Cogn Neuropsychiatry*. 2003;8(4):243-60. doi:10.1080/135468000344000002.

9. Muller F, Simion A, Reviriego E, Galera, C, Mazaux, J, Barat, M, Joseph, P. Exploring theory of mind after severe traumatic brain injury. *Cortex*. 2010;46(9):1088-99. doi:10.1016/j.cortex.2009.08.014.

10. Milders M, Ietswaart M, Crawford JR, Currie D. Social behavior following traumatic brain injury and its association with emotion recognition, understanding of intentions, and cognitive flexibility. *J Int Neuropsychol Soc*. 2008;14(2):318-26. doi:10.1017/S1355617708080351.

11. Babbage DR, Yim J, Zupan B, Neumann D, Tomita MR, Willer B. Meta-analysis of facial affect recognition difficulties after traumatic brain injury. *Neuropsychology*. 2011;25(3):277-285. doi:10.1037/a0021908.

12. Zupan B, Neumann D, Babbage DR, Willer B. The importance of vocal affect to bimodal processing of emotion: implications for individuals with traumatic brain injury. *J Commun Disord*. 2009;42(1):1-17. doi:10.1016/j.jcomdis.2008.06.001.

13.   Radice-Neumann D, Zupan B, Babbage DR, Willer B. Overview of impaired facial affect recognition in persons with traumatic brain injury. *Brain Inj*. 2007;21(8):807-16. doi:10.1080/02699050701504281.

14.   Ietswaart M, Milders M, Crawford JR, Currie D, Scott CL. Longitudinal aspects of emotion recognition in patients with traumatic brain injury. *Neuropsychologia*. 2008;46(1):148-59. doi:10.1016/j.neuropsychologia.2007.08.002.

15.   McDonald S, Flanagan S, Rollins J, Kinch J. TASIT: A new clinical tool for assessing social perception after traumatic brain injury. *J Head Trauma Rehabil*. 2003;18(3):219-38. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12802165.

16.   McDonald S, Saunders JC. Differential impairment in recognition of emotion across different media in people with severe traumatic brain injury. *J Int Neuropsychol Soc*. 2005;11(4):392-9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16209419.

17.   Williams C, Wood RL. Impairment in the recognition of emotion across different media following traumatic brain injury. *J Clin Exp Neuropsychol*. 2010;32(2):113-22. doi:10.1080/13803390902806543.

18.   McDonald S. Exploring the process of inference generation in sarcasm: a review of normal and clinical studies. *Brain Lang*. 1999;68(3):486-506. doi:10.1006/brln.1999.2124.

19.   Bird J, Parente R. Recognition of nonverbal communication of emotion after traumatic brain injury. *NeuroRehabilitation*. 2014;34:39-43. doi:10.3233/NRE-131006.

20.   Zupan B, Babbage DR, Neumann D, Willer B. Recognition of facial and vocal affect following traumatic brian injury. *Brain Inj*. 2014;Early Onli:1-9. doi:10.3109/02699052.2014.901560.

21.   Martín-Rodríguez JF, León-Carrión J. Theory of mind deficits in patients with acquired brain injury: a quantitative review. *Neuropsychologia*. 2010;48(5):1181-91. doi:10.1016/j.neuropsychologia.2010.02.009.

22.   Bibby H, McDonald S. Theory of mind after traumatic brain injury. *Neuropsychologia*. 2005;43(1):99-114. doi:10.1016/j.neuropsychologia.2004.04.027.

23.   Shamay-Tsoory SG, Tomer R, Berger BD, Goldsher D, Aharon-Peretz J. Impaired "affective theory of mind" is associated with right ventromedial prefrontal damage. *Cogn Behav Neurol*. 2005;18(1):55-67. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15761277.

24.   Stone VE, Baron-Cohen S, Knight RT. Frontal lobe contributions to theory of mind. *J Cogn Neurosci*. 1998;10(5):640-56. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15701227.

25. Milders M, Ietswaart M, Crawford JR, Currie D. Impairments in theory of mind shortly after traumatic brain injury and at 1-year follow-up. *Neuropsychology*. 2006;20(4):400-8. doi:10.1037/0894-4105.20.4.400.

26. Milders M, Fuchs S, Crawford JR. Neuropsychological impairments and changes in emotional and social behaviour following severe traumatic brain injury. *J Clin Exp Neuropsychol*. 2003;25(2):157-72. doi:10.1076/jcen.25.2.157.13642.

27. McDonald S, Flanagan S, Martin I, Saunders C. The ecological validity of TASIT: A test of social perception. *Neuropsychol Rehabil*. 2004;14(3):285-302. doi:10.1080/09602010343000237.

28. McDonald S, Bornhofen C, Shum D, Long E, Saunders C, Neulinger K. Reliability and validity of The Awareness of Social Inference Test (TASIT): a clinical test of social perception. *Disabil Rehabil*. 2006;28(24):1529-42. doi:10.1080/09638280600646185.

29. McDonald S. New Frontiers in Neuropsychological Assessment: Assessing Social Perception Using a Standardised Instrument, The Awareness of Social Inference Test. *Aust Psychol*. 2012;47(1):39-48. doi:10.1111/j.1742-9544.2011.00054.x.

30. Baron-Cohen S, Wheelwright S, Hill J, Raste Y, Plumb I. The "Reading the Mind in the Eyes" Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *J Child Psychol Psychiatry*. 2001;42.

31. Kincaid J., Fishburne R., Rogers R., Chissom B. *Derivation of new readability formulas (Automated readability index, fog count, and Flesch reading ease formula) for Navy Enlisted Personnel*. Millington, TN: Naval Technical Training, US Naval Air Station; 1975:8-75.

32. SuperLab. SuperLab Version 5. 2012.

33. Neumann D, Babbage D, Zupan B, Willer B. A Randomized Controlled Trial of Emotion Recognition Training After Traumatic Brain Injury. *J Head Trauma Rehabil*. 2014;(Published. Available at: http://pdfs.journals.lww.com/headtraumarehab/9000/00000/A_Randomized_Controlled_Trial_of_Emotion.99768.pdf. Accessed May 29, 2014.

34. Biszak AM, Babbage DR. Facial affect recognition difficulties in traumatic brain injury rehabilitation services. *Brain Inj*. 2014;28(1):97-104. doi:10.3109/02699052.2013.856475.

35. Brookshire RH, Nicholas LE. *Discourse Comprehension Test*. Tuscon, AZ: Communication Skills Builder, A division of the Psychological Corporation; 1993.

36. Nowicki S. The Manual for the Receptive Tests of the Diagnostic Analysis of Nonverbal Accuracy 2 (DANVA2). 2008.

37.   Spell LA, Frank E. Recognition of Nonverbal Communication of Affect Following Traumatic Brain Injury. *J Nonverbal Behav*. 2000;24(4):285-300.

38.   Neumann D, Zupan B, Babbage DR, Radnovich, AJ, Tomita, M, Hammond, F, Willer, B. Affect recognition, empathy, and dysosmia after traumatic brain injury. *Arch Phys Med Rehabil*. 2012;93(8):1414-20. doi:10.1016/j.apmr.2012.03.009.

39.   Zupan B, Neumann D. Affect Recognition in Traumatic Brain Injury: Responses to Unimodal and Multimodal Media. *J Head Trauma Rehabil*. 2013. doi:10.1097/HTR.0b013e31829dded6.

40.   Osterlind SJ. *Constructing Test Items: Multiple-choice, Constructed-Response, Performance, and Other Formats*. Boston, MA: Kluwer Academic Publishers; 1998.

41.   Downing SM. Statistics of testing. In: *Assessment in Health Professions Education*. New York, NY: Routledge; 2009:93-118.

42.   George D, Mallery P. *SPSS for Windows Step by Step Reference, 11.0 Update*. 4th ed. Boston, MA: Allyn & Bacon; 2003.

43.   Bond F, Godfrey HP. Conversation with traumatically brain-injured individuals: A controlled study of behavioral changes and their impact. *Brain Inj*. 1997;11:319-329.

44.   Galski T, Tomkins C, Johnston MV. Competence in discourse as a measure of social integration and quality of life in persons with traumatic brain injury. *Brain Inj*. 1998;12(769-782).

45.   Benedictus MR, Spikman JM, van der Naalt J. Cognitive and behavioral impairment in traumatic brain injury related to outcome and return to work. *Arch Phys Med Rehabil*. 2010;91(9):1436-1441.

46.   Croker V, McDonald S. Recognition of emotion from facial expression following traumatic brain injury. *Brain Inj*. 2005;19(10):787-799. doi:10.1080/02699050500110033.

47.   Hopkins MJ, Dywan J, Segalowitz SJ. Altered electrodermal response to facial expression after closed head injury. *Brain Inj*. 2002;16(3):245-57. doi:10.1080/02699050110103346.

48.   Jackson HF, Moffat NJ. Impaired emotional recognition following severe head injury. *Cortex*. 1987;23(2):293-300.

49.   Kucharska-Pietura K, Phillips ML, Gernand W, David AS. Perception of emotions from faces and voices following unilateral brain damage. *Neuropsychologia*. 2003;41(8):1082-1090.

50.   Dethier M, Blairy S, Rosenberg H, McDonald S. Emotional regulation impairments following severe traumatic brain injury: An investigation of the body and facial feedback effects. *J Int Neuropsychol Soc*. 2013;19:367-379.

51.   Prigatano GP, Pribram KH. Perception and Memory of Facial Affect Following Brain Injury. *Percept Mot Skills*. 1982;54:859-869.

52.   Rosip JC, Hall J a. Knowledge of Nonverbal Cues, Gender, and Nonverbal Decoding Accuracy. *J Nonverbal Behav*. 2004;28(4):267-286. doi:10.1007/s10919-004-4159-6.

53.   Barrett LF, Lane RD, Sechrest L, Schwartz GE. Sex Differences in Emotional Awareness. *Personal Soc Psychol Bull*. 2000;26:1027-1035. doi:10.1177/01461672002611001.

54.   Collignon O, Girard S, Gosselin F, Saint-Amour D, Lepore F, Lassonde M. Women process multisensory emotion expressions more efficiently than men. *Neuropsychologia*. 2010;48(1):220-5. doi:10.1016/j.neuropsychologia.2009.09.007.

55.   Green R. Deficits in facial emotion perception in adults with recent traumatic brain injury. *Neuropsychologia*. 2004;42(2):133-141. doi:10.1016/j.neuropsychologia.2003.07.005.

56.   Radice-Neumann D, Zupan B, Tomita M, Willer B. Training emotional processing in persons with brain injury. *J Head Trauma Rehabil*. 2001;24(5):313-323. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19858965.

57.   Kinsella GJ. Assessment of Attention Following Traumatic Brain Injury: A Review. *Neuropsychol Rehabil*. 1998;8(3):351-375. doi:10.1080/713755576.

58.   Stuss D, Gow AD. Frontal dysfunction after traumatic brain injury. *Neuropsychiatry, Neuropsychol Behav Neurol*. 1992;5:272-282.

59.   Spikman JM, Boelen DHE, Lamberts KF, Brouwer WH, Fasotti L. Effects of a multifaceted treatment program for executive dysfunction after acquired brain injury on indications of executive functioning in daily life. *J Int Neuropsychol Soc*. 2010;16(1):118-129.

Table 1. Summary data for each of the 25 stories administered to typically developing participants in Study 1.

Story #	Story Description	Target Emotion	Words	Reading grade level	Mean** (max=1)	s.d.
1 *	After being kicked out of her father's home, Nicole's boss offers to assist her in finding a place to live.	Happy	256	4.3	1	.00
2 *	Dave is given a promotion at work.	Happy	152	6.3	.97	.16
3	Tom and his dad build a model plane together.	Happy	78	***	1	.00
4	John and Lisa buy a new home in a safer area of town.	Happy	70	***	1	.00
5 *	Brandon realizes he doesn't have to pay for a dinner he can't afford.	Happy	150	5.1	.72	.45
6	Doug and Stacy reminisce about their marriage on their 25th wedding anniversary.	Happy	242	6.2	1	.00
7 *	Kevin finds himself alone after years of working in a demanding job.	Sad	172	6.6	.85	.36
8 *	Maria comes home to find that her daughter has spilled paint all over an item she personally values.	Sad	250	5.3	.70	.46
9	Angela and Jeremy have been unable to conceive the baby they desperately want.	Sad	204	8.6	1	.00
10	Susan is disappointed when her best friend is unable to attend her 40th birthday celebration at the last minute.	Sad	195	5.1	.62	.49
11	Debra loses her aunt's ring, a family heirloom.	Sad	151	5.6	.97	.16
12 *	Carolyn's daughter is moving over 500 miles away to attend college.	Sad	134	6.7	.93	.27
13 *	After nearly saving enough money for a down payment on a home, Susan and George's car stops running and they must use the money to replace it.	Angry	148	6.3	.95	.22
14	Steven, a very hard worker, observes his colleagues spending much of their day talking around the water cooler instead of working.	Angry	188	4.8	.97	.16
15	While reviewing reports at work, Kate notices numerous errors in a	Angry	155	7.2	.97	.16

	report submitted two days late by one of her employees.					
16	Emma's parents tell her that they are going to move a new home, away from her current school and friends.	Angry	157	5.1	.70	.46
17 *	Jason's best friend cancels their weekly get together to attend a social function at work.	Angry	138	5.1	.90	.30
18 *	Lynn has been married to Brian for two years, but his ex-wife continually stops by unannounced.	Angry	197	7.9	.92	.27
19 *	Beth and her daughter, Sarah, are evicted from their home and have two months to find a place to live.	Fearful	156	4.3	.85	.36
20	After making a rude comment to a customer, Elizabeth's boss asks her to meet with him.	Fearful	174	7.1	.95	.22
21 *	Monica, a widow, is attempting to find ways to fill her time but is nervous about doing things alone.	Fearful	114	6.4	.85	.36
22	After being separated from his wife for one year, Derek's friends encourage him to try to move on with his life.	Fearful	172	6.9	.65	.48
23 *	Matthew is dared to steal answers to a test from the teacher's desk.	Fearful	254	6.4	.88	.33
24	Jeremy is leaving a job he has worked at for many years and starting over somewhere else.	Fearful	191	5.5	.92	.27
25	Even though he knows he isn't supposed to touch it, Steve plays with his father's model airplane and breaks it.	Fearful	161	4.6	.95	.22

Note: Gray shading indicates stories selected to create the first version of the Emotion Inference from Stories Test (EIST-1).

* Indicates stories selected for the second version of the Emotion Inference from Stories Test (EIST-2).

** Mean proportion of participants who correctly identified the target emotion.

*** Stories were not of sufficient length (i.e., ≥ 100 words) to be analysed by the Flesch-Kincaid grade level rating system.

Table 2. Demographics and injury characteristics for each group of participants with TBI.

	EIST-1 Group		EIST-2 Group		
	Session 1	Session 2	Session 1	Session 2	Difference (p)*
Participants (n)	77	27	126	46	
Gender (n)					
Males	53	20	97	39	$\chi^2 = 1.65$ (.25)
Females	24	7	29	7	
Age (years)					
Mean (s.d.)	39.14 (11.27)	39.76 (10.54)	40.19 (12.40)	40.70 (11.70)	$t = .606$ (.54)
Range	21-62	23-60	21-65	21-61	
Ethnicity (%)					
Caucasian	82	85	81	76	
African American	8	7	13	20	$\chi^2 = 1.56$ (.46)
Other	10	7	6	4	
Education (years)					
Mean (s.d.)	13.29 (2.52)	12.29 (1.90)	12.90 (2.40)	12.84 (2.54)	$t = .98$ (.33)
Years Post Injury					
Mean (s.d.)	10.95 (9.36)	10.94 (9.72)	9.88 (8.64)	9.72 (7.57)	$t = .81$ (.42)
Range	1-41	1-36	1-42	1-29	
Severity (%)					
Moderate	1	0	2	4	$\chi^2 = .25$ (1.0)
Severe	99	100	98	96	
GCS					
Mean (s.d.)	5.56 (3.24)	6.20 (3.29)	5.35 (3.21)	4.94 (2.70)	$t = .259$ (.80)
	(n=25)	(n=10)	(n=48)	(n=17)	
LOC in days					
Mean (s.d.)	51.28 (78.89)	74.96(111.38)	40.14 (51.53)	64.75 (69.65)	$t = .85$ (.40)
	(n=37)	(n=13)	(n=62)	(n=36)	
PTA in days (%)					
1-6	8	8	15	6	
≥ 7	92	92	83	94	$\chi^2 = 1.82$ (.23)
	(n=62)	(n=24)	(n=107)	(n=36)	
Cause (%)					
MVA	68	65	59	65	
Fall	12	15	15	15	$\chi^2 = 1.64$ (.65)
Assault	5	12	8	8	
Other	15	8	18	11	

* Groups 1 and 2 were compared for Session 1 only using either chi-square testing, or independent *t*-tests.

Abbreviations: %, Percentage of sample; GCS, Glasgow Coma Scale score; LOC, Loss of consciousness; PTA, posttraumatic amnesia; MVA, Motor vehicle accident.

Table 3. Summary data for Session 1 for participants with TBI in Study 2 (Group 1 and Group 2).

Participants with TBI	DCT Story 1 (Auditory)			DCT Story 2 (Self-read)			DANVA2-AF	EIST
	Total (max=8)	Direct (max=4)	Implied (max=4)	Total (max=8)	Direct (max=4)	Implied (max=4)	(max=24)	(max=12)
Group 1 (n=77)								
Mean	5.54	2.73	2.85	5.28	2.54	2.76	16.19	8.49
s.d.	1.25	.79	.76	1.25	.72	.84	3.86	2.46
Group 2 (n=126)								
Mean	6.59	3.49	3.23	5.52	2.67	2.90	16.46	6.98
s.d.	1.47	.70	.83	1.16	.72	.80	3.23	1.98
Group Difference								
t	4.88	6.42	2.97	1.19	.994	1.03	.53	4.47
p	<.001*	<.001*	.003*	.24	.32	.30	.60	<.001*

Table 4. Classical Test Theory Item Statistics for EIST 1 (n=76).

EIST 1				
Story #	**Target Emotion**	**Item _p_**	**DI**	**Item-test Correlation*** **(r, p)**
1	Happy	.86	.53	_r_ =.65, _p_ <.001
4	Happy	.86	.35	_r_ =.46, _p_ <.001
6	Happy	.87	.53	_r_ =.60, _p_ <.001
9	Sad	.72	.71	_r_ =.56, _p_ <.001
11	Sad	.88	.41	_r_ =.41, _p_ <.001
12	Sad	.53	.65	_r_ =.35, _p_ <.001
13	Angry	.75	.65	_r_ =.55, _p_ <.001
14	Angry	.79	.59	_r_ =.67, _p_ <.001
18	Angry	.71	.53	_r_ =.42, _p_ <.001
20	Fearful	.78	.41	_r_ =.47, _p_ <.001
24	Fearful	.74	.53	_r_ =.50, _p_ <.001
25	Fearful	.79	.53	_r_ =.54, _p_ <.001
Avg		**.77**	**.53**	
EIST 2				
Story #	**Target Emotion**	**Item _p_**	**DI**	**Item-test Correlation*** **(r, p)**
1	Happy	.90	.29	_r_ =.48, _p_ <.001
2	Happy	.89	.13	_r_ =.25, _p_ =.005
5	Happy	.42	.74	_r_ =.57, _p_ <.001
7	Sad	.58	.61	_r_ =.46, _p_ <.001
8	Sad	.61	.61	_r_ =.43, _p_ <.001
12	Sad	.71	.52	_r_ =.42, _p_ <.001
13	Angry	.76	.65	_r_ =.58, _p_ <.001
17	Angry	.78	.32	_r_ =.40, _p_ <.001
18	Angry	.66	.35	_r_ =.35, _p_ <.001
19	Fearful	.48	.13	_r_ =.05, _p_ =.56
21	Fearful	.75	.22	_r_ =.13, _p_ =.13
23	Fearful	.67	.61	_r_ =.37, _p_ <.001
Avg		**.68**	**.43**	

Item p = Item difficulty; DI = Discrimination index

* Pearson correlations significant at p = .004

Figure Captions.

Figure 1. Responses to EIST (Version 1 and 2) by people with and without TBI.

Appendix. Sample Story from EIST.

Story 18 (This story appeared in both EIST-1 and EIST-2)

Lynn was a 32 year old woman who had been married for two years to a 45 year old divorcee with a ten year old daughter. In their first year of marriage, Lynn often found it difficult to deal with the fact that her husband, Brian, had two other women in his life. Brian and Lynn only lived a few towns over from his daughter and ex-wife. Surprisingly, Lynn really got along with Brian's daughter, and she tried to get along with his ex-wife. Brain tried to remain on somewhat friendly terms with his ex for the sake of his daughter. As far as Lynn was concerned it was too friendly. For whatever reason, his ex would always stop over Brian and Lynn's house, without even calling! To top it off, Brian didn't even seem to mind. Lynn wasn't threatened by his ex, but she felt it was an invasion of their privacy. Lynn remembered trying to calmly explain to Brian how she felt about the situation and that he needed to make some boundaries with his ex-wife. Her communication paid off and after one year of marriage, Brian and Lynn eventually worked out a reasonable solution.

Question: How do you think Lynn felt about Brian's ex-wife making frequent visits to their home without calling first?

Answer: Angry