WHY BULLYING VICTIMS ARE NOT BELIEVED:
DIFFERENTIATING BETWEEN CHILDREN’S TRUE AND FABRICATED REPORTS
OF STRESSFUL AND NON-STRESSFUL EVENTS.

by

Megan K. Brunet

A thesis submitted in conformity with the requirements
for the degree of Master of Arts
Graduate Department of Human Development and Applied Psychology
Ontario Institute for Studies in Education
University of Toronto

©Copyright by Megan K. Brunet 2009
WHY BULLYING VICTIMS ARE NOT BELIEVED:
DIFFERENTIATING BETWEEN CHILDREN’S TRUE AND FABRICATED
REPORTS OF STRESSFUL AND NON-STRESSFUL EVENTS
Master of Arts 2009
Megan K. Brunet
Department of Human Development and Applied Psychology
University of Toronto

Abstract
To date, limited research has been conducted to identify differences in children’s truthful and
deceptive statements concerning stressful events. The present study uses automated linguistic
software to detect linguistic patterns and objectively differentiate between the true and false
stressful reports of bullying and non-stressful reports of sports events 7- to 14-year-olds. Results
revealed that children displayed different linguistic patterns when reporting true and false stories,
and between stressful and non-stressful stories. A discriminant analysis reliably differentiated
between true and false stressful and non-stressful stories, though the veracity of non-stressful
stories was more accurately classified than stressful stories. Experiment 2 revealed that adults
were below chance levels in accurately identifying children’s true and false reports of stressful
events (bullying), with confidence ratings and experience with children failing to improve
accuracy scores. Taken together, results reveal that children are able to fabricate emotional and
stressful stories that closely replicate their true reports.
Acknowledgements

There are a number of people who I would like to thank for their support and assistance throughout the process of completing my thesis. First of all, I would like to thank my thesis supervisor, Kang Lee, for his patience, guidance, and for challenging me to work to my utmost potential. He is a talented and creative researcher, and I feel fortunate to have had such a rich research experience under his mentorship. Further, I would like to acknowledge the assistance of my thesis committee member, Victoria Talwar, for her valuable input throughout this process and for introducing me to the world of research. I would also like to thank Angela Evans for her helpful suggestions, invaluable editing skills, and continuous support over the past two years.

I would like to thank my parents, Dave and Monique Brunet, as well as my sisters, Celine and Danielle, for their constant love and support during my academic pursuits. I also want to thank Oliver Costa for encouraging and believing in me when I needed it most. This process was also aided through the amazing support from my friends and colleagues at O.I.S.E., and in the Child Development Research Group.

Finally, I would like to thank the many people who were also involved in the project through data collection, transcribing, and editing from Queen’s University, McGill University and the University of Toronto. Thanks to Nicholas Bala, Cindy Arruda, Shanna Williams, Hannah Lo and Jeta Haxhimanka for all of their hard work.
Table of Contents

Title Page………………………………………………………………………………… i
Abstract…………………………………………………………………………………..   ii
Acknowledgements………………………………………………………………………  iii
Table of Contents………………………………………………………………………...   iv
List of Tables, Figures and Appendices………………………………………………….   v
1.0 Introduction.................................................................................................   1
  1.1 – Children’s Lie-Telling Behaviours....................................................... 1
  1.2 – Adult Lie Detection Accuracy.............................................................. 2
  1.3 – Linguistic Lie Detection Methods....................................................... 3
  1.4 – Automated Linguistic Software Programs............................................ 4
  1.5 – The Current Study................................................................................ 5
2.0 Methods ..................................................................................................... 7
  2.1 – Participants........................................................................................... 7
  2.2 – Procedure and Materials.................................................................... 8
  2.3 – Linguistic Pattern Detection............................................................... 9
  2.31 – LIWC Analysis.................................................................................. 10
  2.32 – Connexor FDG Parser Analysis......................................................... 10
3.0 Results and Discussion................................................................................ 11
  3.1 – LIWC Analysis...................................................................................... 11
  3.11 – Length of Stories.............................................................................. 11
  3.12 – Word Usage....................................................................................... 12
  3.2 – Connexor FDG Parser Analysis............................................................ 18
  3.3 – Discriminant Analysis.......................................................................... 20
  3.31 – Non-Stressful Stories......................................................................... 20
  3.32 – Stressful Stories................................................................................ 21
4.0 Experiment 2 Introduction.......................................................................... 22
  4.1 – Experience and Lie Detection.............................................................. 22
  4.2 – The Current Study................................................................................ 22
5.0 – Methods.................................................................................................. 23
  5.1- Participants............................................................................................ 23
  5.2 – Materials and Procedure.................................................................... 23
6.0 – Results and Discussion............................................................................. 25
  6.1 – Overall Accuracy Rates........................................................................ 25
  6.11 – Signal Detection Analysis................................................................. 25
  6.2 – Experience and Classification Accuracy.............................................. 27
  6.21 – Signal Detection Analysis............................................................... 27
  6.3 – Judgement Confidence......................................................................... 28
  6.4 – Linguistic Judgements......................................................................... 28
7.0 General Discussion....................................................................................... 30
  7.1 – Developmental Linguistic Differences................................................ 30
  7.2 – True vs. Fabricated Stories................................................................... 31
  7.3 – Stressful versus Non-stressful Stories................................................ 33
8.0 References................................................................................................... 36
List of Tables, Figures, and Appendices

Tables

Table 1. The Frequency of the Linguistic Inquiry Word Count Categories by Story Type… 42

Figures

Figure 1. Frequency of LIWC Categories………………………………………….............. 43

Figure 2. Adult Truth and Lie Classification Accuracy Rates………………………………. 44

Appendices

Appendix A: Child Bullying Questionnaire………………………………………………... 45
Why Bullying Victims are Not Believed: Differentiating between Children’s True and Fabricated Reports of Stressful and Non-stressful events

The assessment of the credibility of witnesses is a central challenge for the police and the justice system. Whether a witness is honestly recounting his or her memory of an event is pivotal to the outcome of most cases. The great increase in the number of children appearing as witnesses over the past quarter century has added new dimensions to the assessment of witness credibility and reliability. Many are concerned that children may not be capable of making accurate accounts or that children may be coerced into falsely reporting incidents (Bala, Ramakrishnan, Lindsay, & Lee, 2005). While some studies have revealed that children can give highly accurate accounts and can make competent witnesses (e.g., Bruck & Ceci, 1999; Quas, Goodman, Ghetti, & Redlich, 2000), other studies have indicate that children can be coached into telling convincing fabricated accounts of events (e.g., Lyon, Malloy, Quas, & Talwar, 2008; Orcutt, Goodman, Tobey, Batterman-Faunce, & Thomas, 2001; Talwar, Lee, Bala & Lindsay, 2006). As such, further research is needed to assess the ability of adults to accurately distinguish children’s truthful statements and lies.

1.1 Children’s Lie-Telling Behaviours

Extensive research has revealed that after committing a minor misdeed (e.g., peeking at a toy), the majority of children can and will lie in an effort to conceal their transgression (Lewis, 1993; Peskin, 1992; Talwar, Gordon, & Lee, 2007; Talwar & Lee, 2002; Talwar, Lee, Bala & Lindsay, 2002; Talwar, Lee, Bala & Lindsay, 2004; for review see Talwar & Lee, 2008). Furthermore, some children are even willing to lie to conceal another person’s transgression (Talwar et al., 2006; Tye, Amato, Honts, Devitt, & Peters, 1999). While children in the preschool years are capable of telling simplistic lies and creating fabricated reports of events, the ability to
tell sophisticated and convincing lies improves with age (Gordon, Talwar, & Lee, 2005; Lewis, Stanger, & Sullivan, 1989; Talwar & Lee, 2002; Talwar et al., 2002, 2004). Specifically, young children often have difficulty maintaining consistency between statements and incriminate themselves by leaking critical information about their own deceit (Talwar & Lee, 2002; Talwar et al., 2007). However, between 6 and 10 years of age, the ability to elaborate on lies and to maintain consistency across statements increases (Talwar et al., 2007).

1.2 Adult Lie Detection Accuracy

Children’s true and false reports have also been found to increase in length, complexity, and descriptive detail with age (Craig, Sheibe, Raskin, Kircher, & Dodd, 1999; Goodman & Reed, 1986; Pipe, Lamb, Orbach, & Esplin, 2004; Vrij, 2005), suggesting that older children’s lies may be more difficult to identify. Indeed, extensive research reveals that when adults attempt to differentiate between true and deceptive statements of children from various ages, they are highly inaccurate and rarely above chance (Crossman & Lewis, 2006; Edelstein, Luten, Ekman, & Goodman, 2006; Orcutt et al., 2001; Strömwell, Bengtsson, Leander, & Granhag, 2004; Strömwell, Granhag, & Landstrom, 2007; Talwar & Lee, 2002; Talwar et al., 2006; Tye et al. 1999; Vrij, Akehurst, Brown, & Mann, 2006). Experience with children, or with lie-telling situations in general, does not appear to impact lie detection rates; teachers, police officers, lawyers, and social workers perform at chance levels in differentiating between true and false statements from children and adults (Bala et al., 2005, Ekman & O’Sullivan, 1991; Leach, Talwar, Lee, Bala, & Lindsay, 2004; Strömwell et al., 2007; Vrij, 2005). However, there are some exceptions as studies have found judges, Secret Service agents, CIA agents, and forensic clinical psychologists to perform slightly above chance levels (Bala et al., 2005; Crossman & Lewis, 2006; Ekman & O’Sullivan, 1991; Ekman, O’Sullivan, & Frank, 1999). Thus, there are
some indications that forensic experience or training specific to lie detection may improve veracity classification accuracy, though general experience with a population does not.

1.3 Linguistic Lie Detection Methods

As adults are generally unable to see through children’s deceptive strategies (Strömwell et al., 2007), more systematic methods of lie-detection have been explored. Researchers over the last five decades have developed systematic and objective methods to analyze linguistic differences between statements in an effort to increase veracity classification efficiency and accuracy. Statement Validity Analysis is a method developed in the 1960s to evaluate child testimony in sexual abuse cases (Vrij, 2005). The main component of Statement Validity Analysis is Criteria Based Content Analysis, where 19 criteria are evaluated and higher frequencies of each criterion indicate that the statement is likely to be truthful. Criteria Based Content Analysis has been accepted as evidence in some trials in Continental Europe, as this procedure has been found to correctly categorize true reports 65%-90% of the time (Strömwell et al., 2004; Vrij, 2005). An alternative linguistic analysis entitled Reality Monitoring has also been developed in an attempt to correctly classify truths and lies, with some studies showing it to be more effective than Criteria-Based Content Analysis (Vrij, 2005). Reality Monitoring is based on the idea that true statements will contain more sensory, contextual, and temporal information, while false statements will contain more references to cognitive mechanisms.

While both of these methods have had some success with detecting true and false reports, they have been shown to have limited utility with children and adults. For example, methods such as Criteria-Based Content Analysis and Reality Monitoring are drastically affected by the length of the statement. This is problematic as younger children’s reports tend to contain fewer details, which can impact the length of the statement (Goodman & Reed, 1986; Pipe et al., 2004;
Vrij, 2005; Vrij, Akehurst, Soukara, & Bull, 2004a). Further, both methods of classification require extensive training, are labour intensive to complete, and are typically subjective.

1.4 Automated Linguistic Software Programs

Advances in technology may assist with credibility assessment through the practical application of more objective methods to detect differences in language between truths and lies. One such development in the field is the Linguistic Inquiry Word Count (LIWC) computer software program. LIWC is a tool that detects semantic patterns of speech by analyzing text and calculating the frequencies of word use in relation to the total word count (Pennebaker, Francis, & Booth, 2001). Semantic analysis using LIWC has been applied to adult accounts of true and false reports and several linguistic patterns have been discovered. Specifically, LIWC analysis has revealed that adults’ false reports tend to contain fewer first person pronouns, more simple verbs (such as motion terms and spatial terms), more negative emotional words, and fewer exclusive terms (Bond & Lee, 2005; Newman, Pennebaker, Berry, & Richards, 2003). It is suggested that these linguistic patterns decrease cognitive demands, allowing lie-tellers to maintain consistencies with their accounts and to simultaneously separate themselves from the lie (Bond & Lee, 2005; Newman et al., 2003).

In addition to semantics and word usage, differences between true and fabricated reports in terms of syntax, or the way word are combined to form phrases, must be considered. For example, the Connexor Functional Dependency Grammar (FDG) software program is a syntax tagger program that identifies part-of-speech syntactical classes (e.g., nouns, adjectives, or verbs) and in doing so, establishes syntactic patterns of speech, such as sentence complexity (Järvinen et al., 2004). Connexor FDG software has been used in contexts such as children’s language development (Parisse & Le Normand, 2000) and courtroom questioning (Evans, Lee, & Lyon,
revealing notable and reliable patterns. However, this technology has not yet been used to investigate syntactical differences in true and false statements.

While these computer based methods appear to increase the speed and objectivity in the linguistic analysis of witness’s statements, few studies have applied such forensic technology to children’s reports of events. Furthermore, studies considering differences in children’s true and fabricated statements focus on neutral events that may not be ecologically relevant or applicable to the legal settings (Blandon-Gitlin, Pezdek, Rogers, & Brodie, 2005; Newcombe & Bransgrove, 2007; Strömwell et al., 2007; Talwar & Lee, 2002; Talwar et al., 2006; Tye et al., 1999; Vrij, Akehurst, Soukara, & Bull, 2004b). However, often when children’s reliability is called into question, the situation is serious and the topic can be stressful or emotional to discuss. Further, there is some debate regarding children’s reports of stressful events as memory accuracy may be impacted by the emotional nature of the event (Deffenbacher, 1993), though other studies have found that stress actually improves memory and thus, the ability to recount an event accurately (e.g., Goodman, Bottoms, Schwartz-Kenney, & Rudy, 1991; see Pipe et al., 2004 for a review). As such, it is important to consider how a child’s true and false reports differ for both stressful and non-stressful events.

1.5 The Current Study

The present study was conducted to examine 7- to 14-year-old children’s true and fabricated reports of both stressful and non-stressful events using automated linguistic software programs to detect reliable differences in the linguistic patterns of their reports. A follow-up study was also conducted to determine whether adults could accurately detect children’s false reports of stressful and non-stressful events. In this way, classification of true and false statements according to computer-generated linguistic patterns and human detectors can be
compared to establish more accurate and efficient methods of identifying the veracity of children’s reports.

Reports of stressful events in the current study required children to discuss a true and fabricated situation where they had been bullied. Bullying that involves a threat of harm or the application of force is a form of criminal assault and thus a crime where victims may suffer physically and emotionally. As the harmful effects of bullying are increasingly recognized, victims of bullying are becoming more involved with police and may testify in court. Further, a child who has been bullied potentially experiences stress, embarrassment, shame, or feelings of responsibility, which may somewhat replicate some feelings evident in case of child abuse or neglect, and these children often testify in court (Bottoms & Goodman, 1994; Goodman, et al., 1991; Smith, 2008). However, bullying is a naturally occurring phenomenon, which allows for an ethical investigation into how stressful events are reported and fabricated.

Bullying has become extremely common and estimates of school-aged children who have been bullied range from 15% to 45% (Craig, 1998; Veenstra et al., 2005). However, less than half of bullying victims report the incident (Theriot, Dulmas, Sowers, & Johnson, 2005). When incidents of bullying are reported, parents, teachers, school administrators, and increasingly police and youth court judges must determine if children’s reports are accurate and subsequently establish an appropriate course of action. An understanding of linguistic differences between true and false reports of such stressful bullying events would assist in accurate classification and detection of such statements.

As a comparison to the stressful bullying stories, children described a true and fabricated non-stressful sporting event. This control condition was intended to provide a point of comparison and to evaluate whether stressful and non-stressful events truly differ in how they are
delivered. Children’s verbal statements concerning both stressful and non-stressful events were run through two automatic linguistic software programs, LIWC and Connexor FDG. Patterns in children’s truthful and untruthful reports were analyzed. We aimed to detect semantic and syntactic differences in children’s accounts of true and fabricated events and to determine the effects of stress on children’s language usage. Based on the existing findings regarding a developmental trajectory in the sophistication of lies, younger participants are expected to have shorter and less complex narratives. It is also expected that subtle linguistic differences will also surface between true and false reports; similarly to adult reports (Bond & Lee, 2005; Newman et al., 2003), false reports are expected to be shorter, have fewer self references, use more simplistic terms (such as motion terms and spatial terms) and display linguistic patterns that will decrease cognitive demands (such as fewer discrepancy terms, fewer exclusive terms, and more tentative terms), when compared to true reports. Furthermore, differences between the stressful bullying and non-stressful sport stories are expected; it is hypothesized that children’s stressful bullying reports will be longer, less complex and contain more emotional terms than the non-stressful sport reports due to the nature of the story. It is also expected that children will make fewer self references in the stressful bullying stories compared to non-stressful sport stories in an attempt to distance themselves from the potential embarrassment of the story topic.

Experiment 1

Methods

2.1 Participants

A total of 42 participants between the ages of 7 to 14 ($M = 10.52$, $SD = 2.30$; 16 females) were interviewed in a single half-hour session in two large urban cities in Canada. Participants were divided into four age groups: 7- to 8- year olds ($N = 9$), 9- to 10- year olds ($N = 13$), 11- to
12-year olds (N = 9), and 13- to 14-year olds (N = 11). Participants were recruited through newspaper advertisements and flyers in the community that specifically called for child participants who had been bullied. Parents were instructed to discuss the study with their children prior to participating, but asked not to prepare specific stories with their children in advance. Assent was obtained from all children prior to participating in the study.

2.2 Procedure and Materials

Upon arrival, participants discussed the study with a research assistant. Participants were informed that they would be telling four stories, two of which would be true and two would be fabricated. One true and one fabricated story would be concerning a sporting event while one true and one fabricated story would be concerning a bullying event. Participants discussed their true sports story with a research assistant and notes were made concerning the main points of the story. The research assistant then helped the participant create a story concerning a sporting event that had never happened to them before. These stories generally contained non-stressful content. Following this, the research assistant and the child filled out a questionnaire regarding bullying incidents that the child had experienced (Appendix A). The questionnaire was an adaptation of the Revised Olweus Bullying/Victim questionnaire and asked children to indicate which types of bullying they had experienced (Olweus, 1996). The questionnaire included 8 different types of bullying including physical assault, verbal threat of assault, extortion, exclusion, gossip, robbery, racism/sexism, threats, and sexual harassment. Children were also asked to indicate the frequency at which they experienced each type of bullying on a 5-point Likert scale ranging from ‘Never’ to ‘Several times a week’. Following this, the child discussed a true bullying event with the researcher. The research assistant ensured that the child was comfortable discussing the material and informed participants that they could stop at any point if
necessary. No children stated that they wanted to stop or that they were uncomfortable with the procedure. A fabricated bullying event was created based on a type of bullying the child indicated they had never experienced from the questionnaire. Each story was created in this manner to control for practice effects.

The participant was then taken to the testing area with a second research assistant where they were seated directly in front of a camera on a tripod. The research assistant sat behind the camera and discussed the videotaping procedure with the child. The research assistant had notes from the ‘practice’ session with the first research assistant. After discussing the procedure with the child, the camera was turned on and the child was asked a general prompt so that they would know when to begin (e.g. “Tell me about the time you played baseball”). The child was then given an opportunity to recount details about the event. When the child was finished, the research assistant asked an open-ended prompt, “Is there anything else you remember?” or “What happened next?” The same process was repeated for each of the four stories. In order to increase the participant’s comfort, story order was held constant; true followed by false non-stressful sport stories were told first to allow for rapport to build before the child disclosed their true and then false stressful bullying stories. Children received $10 for participating in the study.

2.3 Linguistic Pattern Detection

Each story was transcribed word-for-word and included both participant and interviewer statements. Prior to considering any linguistic analysis, all interviewer statements were removed. All transcripts were edited to ensure the appropriate punctuation (periods, commas, and quotations) was included. All transcripts were then run through LIWC and Connexor FDG linguistic software programs.
2.31 LIWC Analysis. In addition to the aforementioned preparations, transcripts were prepared according to the requirements identified by the LIWC 2001 Manual (Pennebaker et al., 2001), which included identifying filler words (such as ‘like’ and ‘you know’). The LIWC analysis classifies each word within a specific linguistic category. When using all 72 categories, LIWC accurately categorizes 80% of all words (Pennebaker et al., 2001). For the purposes and content of the present study, 29 categories were initially used for the LIWC analysis. A significant portion of the statements were categorized, with no significant differences between story types (True Sport: $M = 43.28\%, SD = 9.38$; True Bully: $M = 45.12\%, SD = 11.56$; False Sport: $M = 42.50\%, SD = 10.38$; False Bully: $M = 44.43\%, SD = 9.10$). In line with the procedure used by Newman et al. (2001), categories with low frequencies (less than 0.2% of the time), or categories left up to the discretion of the transcriber (e.g. assent, non-fluencies and fillers, such as ‘umm’ or ‘uh’) were excluded. In total, 19 variables were included in the subsequent analysis (see Table 1). The frequency of each category was obtained as a percentage for each story type.

2.32 Connexor FDG parser analysis. Connexor FDG parser software was used to obtain a measure of sentence complexity and words per sentence of the children’s statements. Transcripts were also prepared according to the Connexor FDG parser manual guidelines, with few text corrections being made beyond punctuation. Connexor FDG parser is a linguistic software program that extracts linguistic information from natural language text. Connexor programs have a corpus size of 88 million and recognizes 242,000 unique word forms from the English language. The categorization accuracy of program is 99.3%. The software program analyzes the syntactic elements of text using morphosyntactic tags to identify each element of a sentence (Järvinem et al., 2004). The program analyzes each sentence (as denoted by a period) and each
sentence is parsed into noun and verb phrases, each of which are further parsed into noun and verb phrases until the phrases can no longer be parsed. Thus, the parsing creates a multi-layered syntactical tree. The number of layers per sentence provides a measure of complexity of a particular sentence and averaged over all sentences produced by a child provides a measure of an overall sentence complexity produced by the child. Similarly, the parser also produces a count of words used in a sentence, which, when averaged over all sentences spoken by the child, produces an overall measure of mean words per sentence for the child. The mean scores for complexity and words per sentence were thus calculated for each child for the four stories.

Results and Discussion

Preliminary analysis revealed no significant differences between genders for any of the LIWC categories or for the Connexor FDG results. Thus, the data of both genders was combined for the subsequent analyses. The analyses were also conducted with children’s entire transcripts, which included the initial prompt and subsequent open-ended prompts. This was done in order to obtain a more comprehensive estimate of the child’s actual linguistic patterns. The results section will begin by examining the findings from the LIWC software program followed by the Connexor FDG parser software program.

3.1 LIWC Analysis

3.11 Length of Stories

The length of each story, indexed by LIWC’s Word Count, was investigated using a 2 (Story Content: bully vs. sports) x 2 (Veracity: true vs. false) x 4 (Age Group: 7-8 year olds, 9-10 year olds, 11-12 year olds, and 13-14 year olds) repeated measures Analysis of Variance (ANOVA) on children’s mean word count scores. A significant main effect of Story Content was
found, $F(1,38) = 16.37, p = .05, \eta^2 = .30$, indicating that more words were spoken in the stressful stories ($M = 286.96, SD = 201.09$) compared to non-stressful stories ($M = 216.21, SD = 138.67$).

Consistent with our hypothesis, a significant main effect of Veracity was also found, $F(1,38) = 6.67, p < .05, \eta^2 = .15$, indicating that children used significantly more words in the true stories ($M = 536.00, SD = 344.51$) compared to the false stories ($M = 470.36, SD = 329.33$). However, the main effect of Veracity was qualified by an interaction with Age Group, $F(1,38) = 5.95, p < .05, \eta^2 = .32$. To further examine this interaction, four paired-sample t-tests were performed between the true and false stories of each age group. While 9- and 10-year olds were found to use significantly more words in their true stories ($M = 742.15, SD = 395.15$) compared to their false stories ($M = 552.54, SD = 333.93$), $t(12) = 3.82, p < .05$, no significant differences were found between true and false stories for any of the other age groups.

Overall, children were able to provide lengthier reports dealing with emotionally charged events than reports of non-stressful events. This finding is consistent with literature indicating that children’s memory for emotional incidents tends to be more accurate than mundane incidents and thus their descriptions are more detailed and potentially longer (Goodman, Batterman-Faunce, Shaaf & Kenney, 2002; see Pipe et al., 2004 for a review). However, contrary to our original hypothesis that children’s true statements would be longer than their false statements, only 9- and 10-year-olds were found to use more words in their true compared to their false statements. While the differences in statement length appear to be driven by the performance of 9- and 10-year olds, the reason for this finding is currently unclear.

3.12 Word Usage

To evaluate whether children’s true and fabricated stressful and non-stressful reports differ in the types of words used, a series of 2 (Story Content: bully vs. sports) x 2 (Veracity: true
repeated measures ANOVAs were conducted for the 18 remaining LIWC categories. Results revealed significant differences between the story types for nine linguistic variables: frequency of Six-Letter words (includes words with six or more letters), Spatial terms (e.g., down, in), Motion terms (e.g., go, arrive, car), Positive Emotional words (e.g., happy, pretty, good), Negative Emotional words (e.g., references to anxiety, sadness, anger), Sensory and Perceptual Processes (e.g., see, touch, listen), Self References (e.g., I, me, my), and Tentative statements (e.g., guess, maybe, perhaps) (see Figure 1). Each variable will be discussed in turn.

**Six-Letter Words.** The 2 (Story Content) x 2 (Veracity) x 4 (Age Groups) repeated measures ANOVA on six-letter words yielded a significant main effect of Story Content, $F(1,38) = 9.80, p < .05, \eta^2 = .21$. Specifically, significantly more six-letter words were spoken in the stressful stories ($M = 9.50, SD = 1.96$) compared to non-stressful stories, ($M = 8.32, SD = 1.73$). These results suggest that children are able to use longer words and arguably use a more sophisticated vocabulary when speaking about stressful events compared to non-stressful events.

**Spatial Terms.** The 2 (Story Content) x 2 (Veracity) x 4 (Age Groups) repeated measures ANOVA on spatial terms revealed a significant interaction between Story Content and Veracity, $F(3,38) = 5.14, p < .05, \eta^2 = .12$. To further investigate this interaction, four paired sample t-tests were made between the two true stories, the two false stories, the two non-stressful stories and the two stressful stories. The paired samples t-test between the two true stories revealed that significantly more spatial terms were used in the true non-stressful stories ($M = 3.28, SD = 1.62$) compared to the true stressful stories, ($M = 2.61, SD = 1.49$), $t(41) = 2.32, p < .05$. No other significant results were found for spatial terms. Contrary to findings that use Reality Monitoring as a veracity detection technique with adults (Bond & Lee, 2005; Vrij, 2005), the present study
did not find a significant difference in spatial terms with regards to the veracity of children’s statements. This is most likely due to the content differences of the stories, as spatial terms were found to significantly vary as a function of emotional content. It is possible that the stressful nature of the bullying stories detracted from focusing on spatial descriptors of the report, resulting in fewer spatial words in the stressful stories overall.

*Motion Terms.* The 2 (Story Content) x 2 (Veracity) x 4 (Age Groups) repeated measures ANOVA on motion terms revealed a significant interaction between Veracity and Story Content, $F(1,38) = 4.49, p < .05, \eta^2 = .11$. To further investigate this interaction, four paired sample t-tests comparing the two true stories, the two false stories, the two non-stressful stories and the two stressful stories were conducted on motion terms for each story type. False stressful stories ($M = 1.95, SD = 1.34$) were found to contain significantly more motion terms than true stressful stories ($M = 1.55, SD = 1.14$), $t(41) = 1.82, p < .05$. No other paired samples t-tests were found to be significant.

The ANOVA also yielded a significant interaction between Veracity and Age Group, $F(3,38) = 3.56, p < .05, \eta^2 = .22$. Four follow-up paired sample t-tests between the true and false statements for each age group revealed that 9- and 10-year-olds used significantly more motion terms in their false stories ($M = 1.99, SD = .84$) compared to their true stories, ($M = 1.21, SD = .83$), $t(12) = 2.71, p < .05$. No other significant differences by age group were found. This finding for the 9- and 10-year-olds is consistent with previous studies with adults (e.g., Newman et al., 2003) where motion terms were found to be more prevalent in false stories. Newman and colleagues suggested that motion terms are generally simplistic and tangible, making them easily accessible descriptions to implement when creating fabricated stories. Thus, the use of motion terms may decrease the cognitive demands of fabricating a story while providing a more detailed
description of the event. While this pattern has been repeatedly found for adults (Bond & Lee, 2005; Newman et al., 2003), in the present investigation, this strategy was only evident in the 9- and 10-year olds. Thus, these results suggest that younger children’s fabricated reports do not necessarily follow the same pattern as adults’ and they do not exhibit the same behaviours when giving a false report.

Emotional Words. Two 2 (Story Content) x 2 (Veracity) x 4 (Age Groups) repeated measures ANOVAs were performed on positive emotional words and negative emotion words separately. Only a main effect of Story Content was found for both positive and negative emotional words, $F(1,38) = 32.09, p < .05, \eta^2 = .49$, and $F(1,38) = 19.71, p < .05, \eta^2 = .34$, respectively. Specifically, there were significantly more positive emotional words in non-stressful stories ($M = 3.53, SD = 1.60$) than in stressful stories ($M = 1.70, SD = 1.03$). Conversely, significantly more negative emotional words were used in the stressful stories ($M = 2.01, SD = .95$) compared to the non-stressful stories ($M = 1.30, SD = .82$). These findings confirm that story topics were sufficiently different and appropriate for the present study, as the stressful bullying stories elicited significantly fewer positive emotions and more negative emotions than the non-stressful sports stories and vice versa.

Sensory and Perceptual Processes. The 2 (Story Content) x 2 (Veracity) x 4 (Age Groups) repeated measures ANOVA on sensory and perceptual processes yielded only a significant main effect of Story Content, $F(1, 40) = 109.10, p < .05, \eta^2 = .74$. Specifically, there were significantly more sensory and perceptual process words used in the stressful stories ($M = 3.51, SD = 1.04$) than in the non-stressful stories ($M = 1.50, SD = 1.03$). Contrary to previous research in Reality Monitoring with adults, our results did not reveal significant variation with regards to veracity. However, children were significantly more likely to use these words when
discussing stressful stories compared to non-stressful stories. Therefore, while sensory and perceptual details in adult statements may be contingent upon truthfulness, story content impacts the amount of sensory detail included in children’s statements. This finding may be caused by heightened awareness of sensory stimulation during emotional situations and as a result details are recounted in children’s stressful reports.

**Self References.** The 2 (Story Content) x 2 (Veracity) x 4 (Age Groups) repeated measures ANOVA on self references revealed significant main effects for Veracity, \( F(1,38) = 155.21, p < .05, \eta^2 = .80 \) and Story Content, \( F(1,38) = 49.14, p < .05, \eta^2 = .56 \), which was qualified by a significant interaction between Veracity and Story Content, \( F(1,38) = 107.14, p < .05, \eta^2 = .74 \). Four follow-up paired sample t-tests comparing the two true stories, the two false stories, the two non-stressful stories and the two stressful stories revealed that more self references were made in false stressful stories (\( M = 10.74, SD = 2.47 \)) than in true stressful stories (\( M = 9.75, SD = 2.09 \)), \( t(41) = 2.58, p < .05 \). Further, there were significantly fewer self references made in the true non-stressful stories (\( M = 2.82, SD = 2.74 \)) than in false non-stressful stories (\( M = 11.13, SD = 3.08 \)), \( t(41) = 14.23, p < .05 \), and true stressful stories, \( t(41) = 13.10, p < .05 \). Overall, these results suggest that contrary to previous findings with adults (Bond & Lee, 2005; Newman et al., 2003), children use more self references in false reports than in true reports. While previous studies suggested that adults use fewer self references when creating a fabricated account in an attempt to distance themselves from the lie (Bond & Lee, 2005; Newman et al., 2003), children may be using more self references as these terms may be more readily available to them. By referring to themselves more, children may be reducing cognitive demands and making it easier to produce a fabricated report. Further, self references were more prevalent in the stressful stories than the non-stressful stories, which may be related to reducing
cognitive demands but may also be a result of the story topics themselves. Specifically, the non-stressful sport stories often included descriptions of teams and more general group events while the stressful bullying stories focused more on the individual child.

*Tentative Word Use.* The 2 (Story Content) x 2 (Veracity) x 4 (Age Groups) repeated measures ANOVA on tentative word use revealed only a significant main effect of Story Content, $F(1,38) = 10.61, p < .05, \eta^2 = .22$. Specifically, significantly more tentative terms were used in the stressful stories ($M = 2.40, SD = 1.18$) compared to the non-stressful stories ($M = 1.72, SD = 1.44$). This finding suggests that the stressful nature of the stories may be causing children to be more tentative when describing these emotional incidents. The use of tentative words could also be interpreted as a lack of confidence or a sign that the child is fabricating the story. As such, the finding regarding tentative terms is important as it highlights how emotional content can impact how a child recounts an event and it may impact whether or not a child is believed.

Further, a paired samples t-test was performed to explore whether there was a difference in tentative words used between true and false stressful bullying stories. Results revealed that children used marginally more tentative words in true bullying stories ($M = 2.75, SD = 1.68$) compared to false bullying stories ($M = 2.06, SD = 1.21$), $t(41) = 2.54, p < .05$. This finding is concerning as children seem capable of telling fabricated emotional accounts in a more confident manner than their true accounts, which may result in children’s true bullying stories being less believed than their false stories. Previous studies have noted that nonverbal indicators such as fidgeting, perceived nervousness, and blinking suggest a discomfort or a lack of confidence and are common cues when identifying a person who is lying (DePaulo, 1994; Ekman, O’Sullivan,
Friesen, & Scherer, 1991; Vrij et al., 2006). Similarly, verbal indicators of hesitancy or a lack of confidence may impact if child’s account of an emotional event is believed.

To summarize, the LIWC analysis discovered subtle yet detectable semantic differences between children’s true and false stressful and non-stressful reports. In considering Veracity differences, Word Count and Motion terms were found to be significantly different for the 9- and 10-year olds. Further, more self references were also found to indicate a fabricated account. The story topics of bullying and sporting events were found to be linguistically different as the stressful bullying stories were found to contain more negative emotional words and fewer positive emotional words than the non-stressful stories. Additionally, stressful stories were found to contain more words, larger words, fewer spatial terms, fewer sensory and perceptual terms, more self references and more tentative terms. True and false stressful reports were only found to significantly differ in terms of self references and tentative words, suggesting that children can report false stressful stories that linguistically mimic their true accounts.

3.2 Connexor FDG Parser Analyses

Connexor FDG software program analyzed each sentence and produced both a measure of complexity and words per sentence. Sentences were parsed into noun and verb phrases and the total number of noun and verb phrases per sentence served as our measure of complexity. Each sentence was then further parsed into tokens (i.e., words), which served as our measure of words per sentence. Mean words per sentence and complexity scores were created for each story type and will be assessed in turn below.

To assess whether there was a significant difference in the words per sentence of children’s statements by story type a 2 (Story Content) x 2 (Veracity) x 4 (Age Groups) repeated measures ANOVA on children’s mean words per sentence scores was performed. No significant
main effects or interactions were found indicating that there was no significant difference in the words per sentence of children’s reports by story type or age.

Next, to examine whether there was a significant difference in the complexity of children’s statements by story type, a 2 (Story Content) x 2 (Veracity) x 4 (Age Groups) repeated measures ANOVA on mean complexity scores was performed. Again, results did not reveal any significant main effects or interactions, suggesting that there are no significant differences in the complexity of children’s reports by story type or age.

To further investigate the relation between age and the complexity of children’s statements, a series of linear regressions were completed for each of the four story types (true non-stressful, false non-stressful, true stressful, and false stressful). The mean complexity scores for each of the four story types was used as the predicted variable, with words per sentence entered on the first step followed by age in years on the second step. Because words per sentence and complexity were found to be correlated with one another for each of the four story types (true non-stressful: $r(40) = .82, p < .01$; false non-stressful: $r(40) = .78, p < .01$; true stressful: $r(40) = .85, p < .01$, false stressful: $r(40) = .94, p < .01$), words per sentence scores were entered first to control for the amount of information reported by children to clearly assess the relation between age and the complexity of children’s reports.

As expected, the first model with words per sentence was significant for all four story types (true non-stressful: $F(1, 40) = 82.71, R^2 = .68, p < .05$; false non-stressful: $F(1, 39) = 58.66, R^2 = .61, p < .05$; true stressful: $F(1, 39) = 96.44, R^2 = .71, p < .05$; false stressful: $F(1, 40) = 283.18, R^2 = .87, p < .05$), indicating that as words per sentence increased, children were significantly more likely to use more complex sentences ($\beta = .82$; $\beta = .78$; $\beta = .85$; $\beta = .94$, $ps < .05$ for true and false non-stressful and true and false stressful stories respectively).
When examining the effect of age in the second model, age only significantly contributed above and beyond mean words per sentence in the false sport stories, $F$ change $(1, 37) = 7.18$, $R^2$ change $= .06$, $p < .05$. Specifically, as age increased children were significantly more likely to use more complex sentences in their false sport stories, $\beta = .23$, $t(37) = 2.68$, $p < .05$. No other significant results were found for any other story type. This finding suggests that younger children may have had more difficulty fabricating stories about non-stressful events and as a result, used more simplistic language. It is also important to note that there was no significant relation between age and the complexity of stressful stories, suggesting that younger children were capable of creating stressful stories using complex language similar to older children.

### 3.3 Discriminant Analysis

Given that there were significant linguistic differences between true and false stressful and non-stressful stories, two stepwise discriminant analyses using the Wilks’ Lambda method were conducted to determine whether the linguistic trends identified by LIWC and Connexor FDG could predict the veracity of children’s statements. The veracity of the statement was used as the classifying variable. Eleven dependent variables were entered into the analysis, which were age group, syntactic complexity, the length of the statement (word count), six-letter words, spatial terms, motion terms, positive emotional words, negative emotional words, sensory and perceptual process words, self references, and tentative words. Discriminant analysis were completed for both stressful and non-stressful stories in order to determine if the emotional content of stories impacted how accurately the stories could be classified.

#### 3.3.1 Non-stressful Stories

A stepwise discriminant analysis was conducted with the 11 variables as predictors for the true and false non-stressful sport stories. The overall Wilks’ lambda was significant, Wilks’ $\Lambda = .33$, $\chi^2 (6, N=84) = 91.56$, $p < .05$, indicating that true and
false non-stressful stories could be successfully differentiated based on the 11 linguistic predictors. Specifically, these factors accurately predicted true and false stories 92.9% of the time. Self references were identified as the most significant predictor of story veracity, with more self references significantly predicting that the story is false, \(F(1, 82) = 170.19, p < .05\).

3.32 Stressful Stories. When considering the true and false stressful stories, the stepwise discriminant analysis yielded an overall Wilks’ lambda that was significant, Wilks’s \(\Lambda = .95, \chi^2(1, N=84) = 4.46, p < .05\), with 59.5% of the stories being classified correctly. This suggests that the eleven variables significantly predicted which bullying stories were true or false. Tentativeness was shown to be a unique significant predictor of veracity, with more tentative terms indicating truth, \(F(1, 82) = 4.62, p < .05\). No other variables were found to be significant predictors of veracity with regards to stressful bullying scenarios.

The discriminant analysis revealed that the linguistic cues can be used to discriminate between true and false reports of both stressful and non-stressful events. Furthermore, it appears as though the linguistic cures more accurately differentiate between true and false non-stressful stories (92.9%) than true and false stressful stories (59.5%). The difference in the veracity classification accuracy of the stressful and non-stressful stories suggests that there may be more distinct linguistic differences between true and false non-stressful stories than true and false stressful stories.

Results revealed that there are subtle yet definable differences in the way that children report true and false stressful and non-stressful events. Further, discriminant analysis could classify a significant portion of both stressful and non-stressful stories. The way that these subtle linguistic differences impact the veracity classification of reports by adults must be considered.
Experiment 2

As Experiment 1 revealed that children’s true and fabricated reports vary on a variety of subtle linguistic cues, whether or not adults can identify and differentiate true and fabricated stressful stories must also be investigated. Such findings will assist in understanding the potential use of technology in lie detection and may clarify ways in which humans can improve their lie detection accuracy. Past literature reveals that generally adults are slightly above chance levels when differentiating between children’s truths and lies concerning neutral events (Chahal & Cassidy, 1995; Leach, et al., 2004; Vrij et al., 2006). However, there is currently limited research considering adult detection of children’s lies concerning stressful events. It is possible that the impact of the emotional content of a statement on the veracity classifications of children’s reports by adult raters remains to be clarified.

4.1 Experience and Lie Detection

The influence of experience with children has also been investigated in research attempting to classify and differentiate truthful from fabricated accounts. Past research has discovered that only those with experience with lie detection, such as Secret Service agents, CIA agents, judges and clinical psychologists, are able to perform much above chance levels on detecting children’s lies (Bala et al., 2005; Crossman & Lewis, 2006; Leach et al., 2004; Talwar & Lee, 2002; Vrij, et al., 2006). However, it is interesting to consider if experience with a population and with specific events, though not necessarily with lie detection, can impact and improve deception detection.

4.2 The Current Study

The present investigation considered whether adults would accurately identify children’s true and false stories of a stressful event. Both teacher and non-teacher populations were
included in the present study in order to investigate if experience would assist in identifying stressful stories that commonly occur in school situations. Adults were provided with transcripts of the stories and were not informed of the age, gender or any differentiating features of the child. This was done in accordance with findings that adults may be slightly (though not significantly) more accurate when provided with a transcript of the child’s statement as opposed to a video (Tye et al., 1999). It was expected that based solely on the information from the transcript, the stressful nature of the stories would make the false stories easier to detect. Further, due to teacher’s experience with children generally and with bullying incidents’ particularly, we considered that teacher experience may improve veracity classification of stressful bullying stories. However, due to past research suggesting that general experience does not increase lie detection accuracy, we hypothesized that teachers would not be more accurate than non-teachers.

Methods

5.1 Participants

A total of 89 adults between the ages of 18-60 years ($M = 29.20, SD = 9.71, 74$ females) participated by filling out an online survey, with 69 participants (75%) completing the entire survey. A total of 35 participants (39.3% of sample, $M$ age = 29.69, $SD = 8.66, 32$ females) were teachers, with 25 (71.4%) completing the survey.

5.2 Materials and Procedure

Transcripts of children’s true and fabricated reports of stressful bullying events were read by adults and their judgments concerning statement veracity were given. As we were interested in identifying if certain linguistic cues would assist in the classification of children’s reports, stories were selected based on linguistic components identified in Experiment 1. Specifically, participants whose true stressful statements contained more tentative words and fewer self
references compared to fabricated stressful statements were noted with a differential analysis. After subtracting the false stressful statements from the true stressful statements, sixteen participants with the largest positive difference for their tentative word use and the largest negative difference in their self reference use were identified. Given that each of the participants had two stories (one true and one false), a total of 32 stories were included in the study with half being true and half being false.

Stories were posted on ‘Survey Monkey’ (www.surveymonkey.com), an online survey system. This system allows surveys to be disseminated via email and to be posted online, allowing for a wide variety of individuals to participate. Four different versions of the questionnaire were created, with each having a different random order of the 32 stories. Participants were asked to identify their age, occupation and the amount of experience they had with children. Further, it was also specifically sent to both teachers and non-teachers in order to assess differences in performances due to experience with children.

After completing the demographic information, participants were asked to read each of the 32 statements individually. After reading each story, participants identified whether they believed the story to be true or false. Due to the emotional content of the stories, participants were also given the option to skip any questions they were not comfortable answering. The frequency of skipped responses was calculated for each story, with percentages ranging from 0-10% and comprising only 3.46% of total responses. For further analysis, these responses were not included and coded as missing data, which is similar to previous detection studies (Talwar et al., 2006).

After determining the veracity of each statement, participants were asked to provide a confidence rating of their judgement on a five-point Likert scale ranging from ‘Not at all
Confident’ to ‘Very Confident’. This was done to determine how many participants were simply guessing and how many were confident in their judgements.

The final stage of the survey asked participants about different linguistic components of statements and what linguistic cues they would expect in true and false stories. For example, participants were asked questions such as “Compared to false stories, true stories would contain more self references”. Participants were asked about eight other categories including statement length, if statement length increases by age, word length, statement complexity, negative emotional words, tentative words, inclusive words and exclusive words. To make their judgement, participants used a five-point Likert scale that ranged from “Strongly Agree” to “Strongly Disagree”.

Results and Discussion

Preliminary analysis revealed that there were no significant differences between male and female participants. As such, subsequent analysis combined both genders.

6.1 Overall Accuracy Rates

Accuracy rates were calculated for each participant by averaging their accuracy across both true and false stories. Overall accuracy rates ranged from 25% to 59% ($M = .48, SD = .07$). A one-sample t-tests comparing accuracy rates to chance (.50) indicated that the overall accuracy rate was significantly below chance, $t(51) = 2.50, p < .05$ (see Figure 2).

Further one-sample t-tests were run to compare the accuracy rates of truthful and fabricated stories separately to chance levels. Adult accuracy rates for true stories was found to be at chance levels ($M = .54, SD = .15)$, $t(61) = 1.85, p > .05$, while false stories were significantly below chance levels, ($M = .44, SD = .14), t(52) = 3.30, p < .05$. These findings indicate that false stories were mistakenly judged to be true a significant portion of the time. Two possibilities could
account for this significant finding. One was that participants were simply biased to respond “true” when reading the false story, or they were able to discriminate between the true and false stories but misclassified the true stories as “false” and the false stories as “true”.

6.11 Signal Detection Analysis. To address these two possibilities, Signal Detection Analysis was performed. This analysis was used to ascertain adults’ sensitivity in discriminating between true and false reports. Discrimination abilities (d’), or the ability to differentiate between true and false stories, ranged from -.95 to .32, with a score of zero indicating that there is no discrimination occurring between the two stories. Overall discrimination ability was found to significantly differ from zero, \( M = -.08, SD = .23 \), \( t(51) = -2.50, p < .05 \). The negative difference suggests that while adults are discriminating between true and false stories, they are significantly misclassifying true stories as false and false stories as true.

Whether adults displayed a truth or lie bias was investigated by considering criterion c, which is a measure of whether participants had a bias of responding either “true” or “false”. Criterion c measures were found to range between -.0008 to .0003, with a score of zero indicating that no bias exists. Overall criterion c \( (M = .00, SD = .00) \) was not significantly different from zero, \( t(51) = -1.00 \). However, when biases were further investigated for true and false stories separately, one-sample t-tests revealed that criterion c was significantly different from zero for the true stories \( (M = -.003, SD = .006) \), \( t(61) = -3.72, p < .05 \) as well as for false stories, \( (M = -.01, SD = .01) \), \( t(53) = -4.47, p < .05 \). The negative direction of the criterion c value indicates a minor but significant truth bias for both true and false stories, suggesting that adults are generally slightly more likely to believe a bullying story than to judge it false. These results are consistent with previous studies which suggest adults have a truth bias when detecting children’s true and false statements (e.g., Talwar et al, 2006).
6.2 Experience and Classification Accuracy

Experience with children was also examined to establish if it is related to accuracy rates. A One-Way ANOVA compared the overall accuracy rates of teachers and non-teachers and revealed no significant differences between the accuracy rate of teachers ($M = .49, SD = .05$) and non-teachers ($M = .47, SD = .07$), $F(1, 50) = 1.09, p > .05$. To investigate if teacher and non-teacher accuracy rates were significantly different from chance, one-sample t-tests were conducted and revealed that non-teachers were significantly below chance, $t(34) = 2.48, p < .05$, though teachers were not, $t(16) = .69, p > .05$. Subsequent ANOVAs investigating differences between teachers and non-teachers for accuracy rates of truthful and fabricated stories separately were also completed. There were no differences in the classification accuracy rates of true and false statements for teachers, ($M = .52, SD = .15; M = .47, SD = .14$) and non-teachers ($M = .54, SD = .16; M = .42, SD = .14$), $F(1, 60) = .21, p > .05$ and $F(1, 52) = 1.13, p > .05$, respectively, though non-teachers false story classifications were found to be significantly below chance levels, $t(35) = 3.28, p < .05$. This suggests that while there are no significant differences between the groups, teachers are performing at chance levels while non-teachers are slightly below.

Next, a linear regression was run to determine if the amount of teaching experience significantly predicted the accuracy of teacher ratings. The regression was run with the teachers who completed the survey, with the number of years of experience entered as the predictor and the total classification accuracy as the dependent variable. The model was not significant, $R^2 = .01, F(1, 16) = 2.27, p > .05$, indicating that the number of years of teaching experience was not significantly related to accuracy rates. Further, these findings indicate that similar to past research (e.g. Leach et al., 2004; Vrij et al., 2006), experience with children does not improve
one’s truth and lie detection accuracy, as both teachers and non-teachers performed slightly below chance levels.

6.21 Signal Detection Analysis. Signal detection analysis was also complete to determine if the discrimination abilities and biases of teachers and non-teachers differed. A paired-samples t-tests did not reveal any significant differences in $d'$ between teachers ($M = -.03$, $SD = .19$) and non-teachers ($M = -.11$, $SD = .25$). One-sample t-tests also revealed that teacher discrimination abilities were marginally significantly below zero, $t(16) = -.69$, $p = .05$. Non-teachers discrimination ability was found to be significantly below zero, $t(34) = -2.46$, $p < .05$. This suggests that both teachers and non-teachers discriminated between the truths and lies, but did so inaccurately (i.e., classifying truths as lies and lies as truths). An independent samples t-test did not reveal any differences were found in criterion c between teachers ($M = .00$, $SD = .00$) and non-teachers ($M = .00$, $SD = .00$), $t(50) = .69$, $p > .05$.

Teacher and non-teacher biases were also investigated and criterion c for teachers was found to be significantly different from zero, ($M = .00$, $SD = .00$), $t(16) = 7.00$, $p < .05$, which was mainly due to the lack variability among the teachers who showed very little response bias. This result should be interpreted with caution, as non-teacher’s criterion c was not found to differ significantly from zero, ($M = .00$, $SD = .002$), $t(34) = -1.00$, $p > .05$.

6.3 Judgement Confidence

Confidence scores for each participant were calculated by averaging the confidence evaluations made for each story. To determine whether adults’ confidence significantly impacted accuracy, a linear regression was run with overall confidence ratings as the predictor and overall accuracy rates as the dependent variable. Preliminary findings revealed no significant differences between teachers and non-teachers confidence rating, thus confidence scores were collapsed
across occupation. Overall confidence was not significantly related to accuracy rates, $F(1,43) = .69, p > .05$. This is in line with past research revealing that confidence did not necessarily predict accuracy (Ekman & O’Sullivan, 1991; Leach et al., 2006).

6.4 Linguistic Judgements

For the linguistic judgement portion of the online survey, participants identified the linguistic cues they would expect for true and false stories. Each linguistic judgement was made on a 5-point scale from ‘Strongly Agree’ to ‘Strongly Disagree’, with a middle neutral value of 3 (“Neither Agree nor Disagree”). A one-sample t-test was used to evaluate if judgements were significantly different from the neutral judgement of 3. It was discovered that there were significant differences for four of the nine linguistic variables. Results revealed that participants thought that true stressful stories would have fewer large words ($t(68) = 5.18, p < .05$), be less complex ($t(68) = 3.01, p < .05$), contain more self references ($t(68) = 3.10, p < .05$) and more negative emotional words ($t (68) = 2.87, p < .05$) compared to false stressful stories. This suggests that participants generally had some ideas about the contents of true and false stories. However, with regards to judgements about large words, complexity and negative emotional words, judgements were not supported by findings from Experiment 1. Therefore, children were not found to fulfill expectations in terms of the linguistic patterns in their true and false stories. Further, judgements concerning self references contradict actual findings in Experiment 1; participants expected children’s true stories to contain more self references, which is in line with previous adult research but are opposite to findings from Experiment 1.

As adults tend to have significant misconceptions about linguistic differences between children’s true and false statements, it does not seem surprising that overall, adults were significantly below chance levels in veracity classifications. They inaccurately classified a
significant portion of the false stories as true. Few differences were found between teachers and non-teachers, as accuracy classifications did not significantly differ but non-teachers were found to be significantly below chance levels. The present findings appear consistent with past literature investigating truth and lie classification accuracy in that teacher experience with children generally and bullying specifically did not significantly increase their veracity detection rates (Chahal & Cassidy, 1995; Leach, et al., 2004; Vrij et al., 2006).

General Discussion

The present studies examined the classification of children’s true and false reports of stressful and non-stressful events. To date, few studies have made comparisons between children’s true and false reports and to our knowledge, children’s stressful (and potentially criminal) and non-stressful fabricated reports have not been directly compared. Results of the present study revealed significant linguistic differences between children’s true and false statements as well as between stressful and non-stressful reports of events.

7.1 Developmental Linguistic Differences

Contrary to our original hypothesis, when evaluating developmental differences in children between the ages of 7 to 14 years, few significant linguistic differences emerged. Specifically, we discovered that sentence complexity significantly increased with age for the false non-stressful stories. This developmental difference implies that false non-stressful stories may be more difficult for younger children to fabricate and as such, these stories were told in a more simplistic manner. It is possible that the reason for the developmental complexity finding is due to the content itself; the sport stories were generally less specific in their content and contained less structure than the false bullying stories. Conversely, bullying stories were typically more formulaic and contained some background information, an incident, and a resolution, making
them easier to construct and resulting in no developmental difference. This developmental finding for the false non-stressful stories is important as it suggests that age differences discovered when considering stories regarding neutral content may be a misrepresentation of the child’s actual ability to fabricate a story of a stressful or criminal event.

7.2 True versus fabricated stories

An important finding of the current study was that children’s fabricated stories closely resembled their true accounts in terms of statement length and complexity. Further, of the 19 LIWC categories considered, only three significantly differed in terms of Veracity: statement length, motion terms, and self references. However, even with the limited findings, the discriminant analysis was able to use linguistic patterns and differentiate between true and false reports for both stressful and non-stressful stories.

Further, linguistic cues to deceptive reports were discovered. Consistent with our hypothesis, children’s true reports were found to be significantly longer than false reports. Additionally, we found that children used more motion terms in their false stressful stories compared to their true stressful stories, which is consistent with past adult literature. However, it should be noted that these results were specifically more pronounced with 9- and 10-year-olds. Finally, findings regarding self references were contrary to the adult literature and actually occurred more in false than in true stories (Bond & Lee, 2005; Newman et al., 2003).

Results regarding differences between true and false stressful reports also varied slightly from the non-stressful reports. Children’s false stressful reports were found to contain more self references, more motion terms, and fewer tentative terms. Children’s non-stressful stories differed by self references, but not by motion terms or tentative terms. Findings reveal that children may be using more simple terms in false stressful reports, possibly in an effort to
decrease cognitive demands when attempting to deceive about a stressful event. Further, children appear to be more hesitant when disclosing true stressful accounts, which may also be related to the stressful nature of the story.

Methods attempting to use linguistic patterns to differentiate between adults’ true and fabricated reports have been shown to have some successes. Recent studies have attempted to combine methods such as Reality Monitoring with LIWC automated analysis (Bond & Lee, 2005). The LIWC categories of spatial terms, sensory processes, temporal terms, and cognitive mechanisms can be used to represent Reality Monitoring categories and were found to correctly classify the statements in terms of veracity 71.1% of the time (Bond & Lee, 2005). The current study found that results concerning spatial terms and sensory processes were significant only in terms of content, with the frequency of each category varying depending on the stressful nature of the story. However, no significant results for temporal terms of cognitive mechanisms were found. Taken together, this suggests that Reality Monitoring procedures using the LIWC analysis may not be as useful with children as it is with adults because it does not consider children’s cognitive capacity or sensitivity to story content.

When examining adults’ expectations of linguistic differences we found that adults tended to hold expectations that were not supported by the linguistic analyses performed in Experiment 1. For example, consistent with the literature on adults’ lies (Bond & Lee, 2005; Newman et al., 2003), adults in the present investigation expected children to use more self references when telling the truth. In contrast, Experiment 1 revealed that children use fewer self references in true stories. The differences between adults’ expectations and children’s actual behaviours may have resulted in adults doubting children’s true stressful reports.
7.3 Stressful vs. Non-stressful Stories

The current study provides unique insight into children’s event reports of both stressful and non-stressful stories. Results revealed that the linguistic content of stressful and non-stressful stories differed significantly, as was noted by differences in the positive and negative emotional content of the stories. It also appears as though emotional content may impact the type of description given as more sensory processing and spatial words were spoken in stressful bullying stories compared to non-stressful sport stories. Further, tentative word use was markedly different for stressful stories compared to non-stressful stories, suggesting that emotional content, as opposed to veracity, can significantly impact the tentativeness or confidence in children’s language.

Additionally, the discriminant analysis appears to be more successful in differentiating between true and false stories for non-stressful events (92.9%) as opposed to stressful events (59.5%). This suggests that children may have been better able to fabricate stressful stories to resemble true stressful stories than with non-stressful content. One possible explanation for this finding is that after experiencing the emotional event of being bullied, children ruminate and can replicate their emotions and feelings more comprehensively when fabricating an account. Indeed, familiarity of the event has been shown to improve one’s ability to create a more believable story (Blandon-Gitlan, et al., 2005; Pezdek et al, 2004). Future studies should include participants who have not been bullied in any way or experienced the stressful event to determine if the ability to closely replicate true stories is a result of experience. It is clear that the content and the manner in which these stories are told significantly differ.

Future studies should also have adults view videotapes of children’s reports of stressful events such as bullying and give their veracity judgments. While Tye et al. (1999) found adults
to be slightly more accurate with transcripts as opposed to videotapes, viewing children reporting such stressful incidents would be more ecologically relevant and potentially more applicable to a courtroom situation.

While this research has contributed to a greater understanding of children’s linguistic patterns when reporting true and fabricated events, the results and applications of linguistic software programs must be considered. The advantage of using such software is that it objectively detects subtle nuances in language and gives an idea of the sophistication of children’s statements. In addition, the two computer software programs combined were able to differentiate between true and false bullying stories 59.5% of the time, whereas adult readers were only able to correctly differentiate between the stories 48% of the time, which was below chance. The success of the automated linguistic program’s discrimination compared to adult raters suggests that these tools present an objective and seemingly more accurate method of lie detection. Further, the linguistic software programs can be aided with the help of automated transcribing programs. However, this process is not yet ‘fine-tuned’ enough to be applied to a wide setting or to be used with efficiency and the accuracy rates still need improvement. However, such linguistic analyses succeed in demonstrating the verbal differences in children’s accounts and highlight the significance of ecologically relevant research.

The ability to effectively detect children’s deceptive statements about stressful events has potential implications for the justice system and the professionals who work in it, such as police, lawyers, child protection workers, and judges. Children and adolescents are often witnesses in courtroom situations and are required to testify about their own abuse, neglect, or victimization, and the outcomes of such cases have serious implications. Further, those working with children in non-forensic situations, such as teachers, doctors, and social workers, require methods for
detecting true or fabricated reports of events to proceed appropriately. The current study provides a starting point for further research that may eventually be used to support professionals in obtaining the truth.
References


<table>
<thead>
<tr>
<th>Dimension</th>
<th>Example</th>
<th>True Stressful</th>
<th></th>
<th></th>
<th>False Stressful</th>
<th></th>
<th></th>
<th>True Non-Stressful</th>
<th></th>
<th></th>
<th>False Non-Stressful</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td><strong>Linguistic Dimensions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word Count</td>
<td></td>
<td>315.67</td>
<td>240.48</td>
<td>258.26</td>
<td>181.00</td>
<td>220.33</td>
<td>136.81</td>
<td>212.10</td>
<td>157.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Words per Sentence</td>
<td></td>
<td>17.15</td>
<td>6.73</td>
<td>17.50</td>
<td>10.06</td>
<td>15.94</td>
<td>8.14</td>
<td>15.02</td>
<td>5.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unique Words</td>
<td></td>
<td>54.88</td>
<td>11.56</td>
<td>55.57</td>
<td>9.10</td>
<td>56.73</td>
<td>9.38</td>
<td>57.50</td>
<td>10.38</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Words longer than six letters</td>
<td></td>
<td>9.71</td>
<td>2.36</td>
<td>9.29</td>
<td>2.67</td>
<td>7.65</td>
<td>2.41</td>
<td>9.00</td>
<td>3.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First-person Singular</td>
<td>I, my</td>
<td>8.89</td>
<td>2.42</td>
<td>9.97</td>
<td>2.70</td>
<td>9.18</td>
<td>3.8</td>
<td>8.25</td>
<td>3.38</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First-person Plural</td>
<td>we, us</td>
<td>0.85</td>
<td>1.02</td>
<td>0.76</td>
<td>0.87</td>
<td>2.82</td>
<td>2.74</td>
<td>2.88</td>
<td>3.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Self References</td>
<td>I, we</td>
<td>9.75</td>
<td>2.09</td>
<td>10.74</td>
<td>2.47</td>
<td>2.82</td>
<td>2.74</td>
<td>11.13</td>
<td>3.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Second Person Pronouns</td>
<td>you, your</td>
<td>0.62</td>
<td>0.81</td>
<td>0.61</td>
<td>0.82</td>
<td>0.39</td>
<td>0.83</td>
<td>0.25</td>
<td>0.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Third person Pronouns</td>
<td>she, their</td>
<td>6.21</td>
<td>2.02</td>
<td>5.97</td>
<td>2.55</td>
<td>2.19</td>
<td>1.53</td>
<td>2.87</td>
<td>2.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Psychological Processes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive Emotions</td>
<td>happy, pretty</td>
<td>1.64</td>
<td>1.16</td>
<td>1.75</td>
<td>1.32</td>
<td>3.44</td>
<td>2.06</td>
<td>3.62</td>
<td>2.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive Feelings</td>
<td>joy, love</td>
<td>0.43</td>
<td>0.87</td>
<td>0.26</td>
<td>0.32</td>
<td>0.42</td>
<td>0.52</td>
<td>0.31</td>
<td>0.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative emotions</td>
<td>hate, enemy</td>
<td>2.10</td>
<td>1.17</td>
<td>1.91</td>
<td>1.15</td>
<td>1.07</td>
<td>0.88</td>
<td>1.53</td>
<td>1.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive Processes</td>
<td>know, ought</td>
<td>5.60</td>
<td>2.89</td>
<td>5.45</td>
<td>2.56</td>
<td>4.58</td>
<td>2.47</td>
<td>5.08</td>
<td>2.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrepancy Terms</td>
<td>should, would</td>
<td>1.85</td>
<td>1.37</td>
<td>1.79</td>
<td>1.45</td>
<td>1.23</td>
<td>1.36</td>
<td>1.67</td>
<td>1.39</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tentative Terms</td>
<td>maybe, guess</td>
<td>2.75</td>
<td>1.68</td>
<td>2.06</td>
<td>1.21</td>
<td>1.80</td>
<td>2.02</td>
<td>1.64</td>
<td>1.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Certainty</td>
<td>always, never</td>
<td>0.66</td>
<td>0.71</td>
<td>0.76</td>
<td>0.68</td>
<td>0.59</td>
<td>0.79</td>
<td>0.75</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensory and Perceptual</td>
<td>see, touch</td>
<td>3.46</td>
<td>1.33</td>
<td>3.56</td>
<td>1.48</td>
<td>1.44</td>
<td>1.28</td>
<td>1.57</td>
<td>1.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Relativity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>hour, day</td>
<td>6.05</td>
<td>3.06</td>
<td>5.76</td>
<td>2.55</td>
<td>6.31</td>
<td>3.03</td>
<td>6.04</td>
<td>2.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space</td>
<td>around, over</td>
<td>2.61</td>
<td>1.49</td>
<td>3.02</td>
<td>1.52</td>
<td>3.28</td>
<td>1.62</td>
<td>2.81</td>
<td>1.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motion verbs</td>
<td>walk, move</td>
<td>1.55</td>
<td>1.14</td>
<td>1.95</td>
<td>1.34</td>
<td>2.16</td>
<td>1.74</td>
<td>1.87</td>
<td>1.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. The Frequency of the Linguistic Inquiry Word Count Categories by Story Type
Figure 1. Frequency of LIWC Categories
Figure 2. Adult Truth and Lie Classification Accuracy Rates
Appendix A

Child Bullying Questionnaire

Please complete this questionnaire using a scale ranging from 1 to 5:

1 = never
2 = has happened once or twice
3 = two or three times a month
4 = about once a week
5 = several times a week

Are you, or have you ever been…

| Question                                                                 | Response
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbally harassed (called mean names, made fun of or teased)?</td>
<td></td>
</tr>
<tr>
<td>Excluded or ignored by other people?</td>
<td></td>
</tr>
<tr>
<td>Physically assaulted (hitting, kicking, shoving, biting)?</td>
<td></td>
</tr>
<tr>
<td>Gossiped about (others told lies or spread false rumours)?</td>
<td></td>
</tr>
<tr>
<td>Robbed (violence used to take money or property)?</td>
<td></td>
</tr>
<tr>
<td>Physically threatened?</td>
<td></td>
</tr>
<tr>
<td>Threatened so that someone can get something from you (extortion)?</td>
<td></td>
</tr>
<tr>
<td>Made fun of because of your race, ethnicity, religion, etc?</td>
<td></td>
</tr>
<tr>
<td>Sexually harassed (comments, names or gestures)?</td>
<td></td>
</tr>
<tr>
<td>Other? (Please identify)</td>
<td></td>
</tr>
</tbody>
</table>

How often have you been bullied in the last three months?