

# Crisis

## **Analyzing Language in Suicide Notes and Legacy Tokens: Investigating Clues to Harm of Self and Harm to Others in Writing**

Michael J. Egnoto and Darrin J. Griffin

Online First Publication, January 19, 2016. <http://dx.doi.org/10.1027/0227-5910/a000363>

### CITATION

Egnoto, M. J., & Griffin, D. J. (2016, January 19). Analyzing Language in Suicide Notes and Legacy Tokens: Investigating Clues to Harm of Self and Harm to Others in Writing. *Crisis*. Advance online publication. <http://dx.doi.org/10.1027/0227-5910/a000363>

# Analyzing Language in Suicide Notes and Legacy Tokens

## Investigating Clues to Harm of Self and Harm to Others in Writing

Michael J. Egnoto<sup>1</sup> and Darrin J. Griffin<sup>2</sup>

<sup>1</sup>FIRE Risk Communication and Resilience, National Consortium for the Study of Terrorism and Responses to Terrorism (START), University of Maryland, College Park, MD, USA

<sup>2</sup>Department of Communication Studies, College of Communication and Information Sciences, University of Alabama, Tuscaloosa, AL, USA

**Abstract.** *Background:* Identifying precursors that will aid in the discovery of individuals who may harm themselves or others has long been a focus of scholarly research. *Aim:* This work set out to determine if it is possible to use the legacy tokens of active shooters and notes left from individuals who completed suicide to uncover signals that foreshadow their behavior. *Method:* A total of 25 suicide notes and 21 legacy tokens were compared with a sample of over 20,000 student writings for a preliminary computer-assisted text analysis to determine what differences can be coded with existing computer software to better identify students who may commit self-harm or harm to others. *Results:* The results support that text analysis techniques with the Linguistic Inquiry and Word Count (LIWC) tool are effective for identifying suicidal or homicidal writings as distinct from each other and from a variety of student writings in an automated fashion. *Conclusion:* Findings indicate support for automated identification of writings that were associated with harm to self, harm to others, and various other student writing products. This work begins to uncover the viability or larger scale, low cost methods of automatic detection for individuals suffering from harmful ideation.

**Keywords:** prevention, legacy tokens, suicide writings, LIWC text analysis

Two tragic incidents that occur within colleges and universities are homicide and suicide. Both have been investigated with the goal of detecting individuals who may commit harm to self or harm to others (Fox & DeLateur, 2014; Mann et al., 2005). Often, crisis management teams are forced to rely on the past experience of team members to help make difficult judgment calls on the real level of threat presented in different scenarios (Albrecht, 2010). As such, scholars and administrators can benefit from the addition of a quantitative tool that can quickly process data to help with the identification of at-risk individuals. This work presents the findings of a cross-sectional computerized linguistic analysis of the writings (legacy tokens) of known spree killers as well as suicide notes in the hope of finding readily identifiable traits that may help prevent these tragedies.

### How Spree Killers Differ From Terrorists

Much debate exists over the types of murders that plague society. Generally, three categories are recognized: serial

murder, mass murder, and spree murder (Blackman, Leggett, Olson, & Jarvis, 1999; Fox & Levin, 1998). The distinction between mass murder and spree murder is difficult to make because of conflicting definitions between scholars (Edelstein, 2014). However, serial murder differs from both in that a serial killer focuses generally on one or a few victims at a time and operates over a significant duration (Edelstein, 2014). None of these categories adequately account for terrorism. Recent literature (Edelstein, 2014) has proposed a fourth category – the serial–spree murder (Edelstein, 2014). The serial–spree murderer differentiates from the mass or spree killer in that their intention is not just to kill multiple people without a cooling off period, but to achieve this goal multiple times, often with a political orientation; sometimes this category is called “terror murder” (Edelstein, 2014; Fox & Levin, 2007). Few works examine the writings of terrorists, but some studies have used WordNet (a text-analysis program; <http://www.wordnet.princeton.edu>) to detect terrorists through online writings by identifying word similarity, and shows promise for future applications in terrorism detection (Choi, Ko, Kim, & Kim, 2013).

## Spree Killers

Spree killings occur when one or more offenders commit multiple murders without a cooling-off period (Federal Bureau of Investigation [FBI], 2008). However, the value of treating spree killers as a separate class from serial killers has been debated (Morton & Hilts, 2005). Much of this study investigates school shooters within the class of spree killers. We feel this broad classification is further justified because the analysis of legacy tokens and the mental state of the spree killers is not impacted by how quickly their events are brought to an end.

Spree killers share many commonalities in personality and life experiences. These killers often hold strong sentiments of individualism, lack a family orientation, and possess a history of anger, depression, suicidal ideation, and feelings of ostracism (Verlinden, Henson, & Thomas, 2000). Paranoia, narcissism, bigotry, lack of empathy, and antisocial personality disorder are also reoccurring afflictions of shooters (Verlinden et al., 2000). Those who are intending or planning to harm others may manifest these desires to others through e-mails, social networking websites, or physical manifestos. Most recently, law enforcement groups like the FBI describe messages that shooters use to foreshadow their actions or leave a mark on the world as legacy tokens (Simons & Tunkel, 2013).

Legacy tokens are manifestos, wills, or other communications designed to claim credit for an attack and articulate the motivation behind violence (Meloy & Hoffman, 2013). Current evaluation of legacy tokens often relies on the consensus of threat assessment teams to evaluate the imminence and significance of impending violence (Meloy & Hoffman, 2013). Although legacy tokens are evaluated on a variety of categories by threat assessment teams, the time and resources required to create these teams and have them evaluate materials are substantial. Creation of parsimonious methodologies would benefit the evaluation of legacy tokens to better predict threats. Therefore, this work seeks to evaluate legacy tokens with a computer-assisted text analysis tool called Linguistic Inquiry and Word Count (LIWC; Pennebaker, Booth, & Francis, 2007).

Spree killers have strong negative afflictions toward school and peer orientation. Spree killers also express high levels of narcissism, anger, depression, and individualism, and share common social environmental factors such as socioeconomic status and familial educational attainment as well as feelings of ostracism (Langman, 2009a; McGee & DeBernado, 1999; O'Toole 2000; Verlinden et al., 2000). Measuring all facets of these phenomena from text analyses would prove laborious and full of error. Instead, focusing on the primary components of peer and school engagement is the more effective course of action. Recall that Meloy and Hoffman (2013) and O'Toole (2010) focused on the psychological aspects of emotionality and violence, which likely manifest in the form of anger, negative emotion, and death. LIWC has been used before to predict forms of aggression (Pennebaker, 2011) and to identify psychopathy (Hancock, Woodworth, & Porter, 2013).

Tausczik and Pennebaker (2010) in a comprehensive meta-analysis of LIWC evaluated the categories from the

LIWC database that had been published on, and reported the psychological correlates they are generally associated with. Despite negative emotion being among the most prolific categories used by LIWC in publication, a canon of psychological correlates were not reported for it, nor were they reported for anger or death category usage, despite a number of publications also utilizing them. However, an analysis of the documents used for the meta-analysis showed that negative emotion was commonly used in studies of trauma, bereavement, failed relationships, mental health, conflict, sexual abuse, depression, lying, and violence (Tausczik & Pennebaker, 2010). Consequently, negative emotion is a strong category for inclusion in this analysis.

Anger was associated with emotional health, sexual abuse, and depression, with the death category occurring most frequently in studies of trauma, bereavement, and psychological well-being (Tausczik & Pennebaker, 2010). Therefore, it is reasonable to assume that the written vocabulary of spree killers as quantified by the LIWC language categories would differ from that of a student population. Our inquiry was guided by the following research hypotheses:

- H1:* The legacy tokens of spree killers will contain more negative emotion references than the writing of students.
- H2:* The legacy tokens of spree killers will possess more anger references than the writing of students.
- H3:* The legacy tokens of spree killers will have more sexual references than text taken from students.

## Past and Present Examinations of Suicidal Behavior

Shneidman and Farberow (1956) noted that distinguishing suicidal persons based on case history alone is extremely difficult. Further, the vast majority of people who complete suicide have a history of suicidal threats or attempts, with subjects suffering from paranoia or schizophrenia being the most likely to complete suicide (Shneidman & Farberow, 1956). Later research found that simulated and authentic suicide notes could be differentiated (Osgood & Walker, 1959; Shneidman & Farberow, 1957) but that further refinement is necessary to improve accuracy in segregating the notes of the suicidal and nonsuicidal (Black, 1993). Ogilvie, Stone, and Shneidman (1966) are credited with the first computerized attempt to examine actual suicide notes and simulated ones. They relied on an early computer system for content analysis named the General Inquirer System (see Stone & Hunt, 1963). Computer technology has advanced greatly allowing complex algorithms and machine learning to increase detection of authentic suicide notes (Pestian, Nasrallah, Matykievicz, Bennett, & Leenaars, 2010). Recently, a review of screening tools and procedures for identification of risk among youth collected a great number of tools to evaluate individuals for suicidal ideation (National Action Alliance for Suicide Prevention, 2013). Quite effectively, many of these tools could be administered quickly and with minimal training; regrettably, the tools all required active

engagement of the potentially suicidal person interacting with a suicide prevention stakeholder.

## Suicidal Ideation

Unlike for legacy tokens, a robust body of research exists that has employed LIWC and linguistic analysis tools to investigate the relationship between suicidal ideation and writing. Work on suicidal poets found increased reference to the self as compared with nonsuicidal poets (Stirman & Pennebaker, 2001). Additionally, future tense usage, positive emotion, and social references increased in notes from completed suicides as opposed to attempted suicides (Handelman & Lester, 2007). Echoing the use of positive emotion in suicide attempts, Lester (2007) also found an increase in positive emotion words immediately prior to completed suicide. Finally, additional work revealed that references to the self as well as positive emotion words increased as the time to completed suicide counted down to the end of the individual's life (Lester, 2009).

*H4:* Suicidal writings will have more self-references than texts from students.

*H5:* Suicidal writings will have more future tense usage than texts from students.

*H6:* Suicidal writings will have more positive emotion than texts from students.

## Text Analysis and LIWC

Text analysis is a reliable tool for predicting interactions and cohesiveness in group situations (Gonzales, Hancock, & Pennebaker, 2010) as well as for establishing behavioral intentions and emotionality at the individual level (Tausczik & Pennebaker, 2010). The LIWC (Pennebaker et al., 2007) is a program that analyzes text on a word-by-word basis. LIWC compares the words with a predefined set of dictionaries (version 1.11 2/13/2011; Pennebaker, 2007; dictionary, English). These dictionaries are further divided into 74 psychological dimensions correlating to psychological effects and cognitive operations (Gonzales et al., 2010; Pennebaker & Stone, 2003). We proposed that LIWC can be used to examine the psychological dimensions present in the legacy tokens of spree killers in addition to the work that it has already been utilized in for suicide.

LIWC also provides comparisons with several different types of text that have been aggregated over the years (Pennebaker, Chung, Ireland, Gonzalez, & Booth, 2007; Pennebaker & Francis, 1996). Categorically, they represent emotional writing, control writing, blog writing, academic writing, novel writing, and talking. Much of these writings were completed by students during the validation of LIWC, and therefore represent a solid data source for comparison. Emotional writing, control writing, blogs, novels, and talking were aggregated to create a robust comparison group to compare against.

*Table 1.* Summary of spree killers and data, sorted by year

Name	Year	Wounded	Killed	Total	Type of data	Location
Charles Whitman	1966	32	16	48	Letters/suicide note	University of Texas, Austin, USA
Marc Lépine	1989	14	14	28	Suicide note	Ecole Polytechnique, Canada
Timothy McVeigh	1995	600	168	768	Manifesto	Oklahoma City, USA
Luke Woodham	1997	7	3	10	School paper/journal/essay	Pearl High School, USA
Kipland Kinkel	1998	24	4	28	Notes/journal	Thurston High School, USA
Dylan Klebold	1999	12	7	19	School paper	Columbine High School, USA
Eric Harris	1999	12	6	18	Journal	Columbine High School, USA
Mark O. Barton	1999	13	12	25	Letter/suicide note	Stockbridge, USA
Jeffrey Weise	2005	5	9	14	Social media (web forum)	Red Lake High School, USA
Erik Hainstock	2006	1	0	1	Letters from prison	Weston High School, USA
Seung-Hui Cho	2007	17	32	49	School papers/manifesto	Virginia Tech University, USA
Pekka-Eric Auvinen	2007	12	9	21	Manifesto	Jokela HS, Finland
Robert Hawkins	2007	6	9	15	Letters/suicide note	Westroads Mall, USA
Jiverly A. Wong	2009	4	14	18	Suicide note	Binghamton, USA
Michael McLendon	2009	6	11	17	Suicide note/letter	Geneva, USA
Wellington Oliveira	2011	20	12	32	Suicide note	Esc. Municipal Tasso Da Silv., Brazil
Jared Lee Loughner	2011	13	6	19	Social media (YouTube)	Tucson, USA
T. J. Lane	2012	3	3	6	Social media (Facebook)	Chardon High School, USA
Christopher Dorner	2013	3	5	8	Manifesto	Orange / San Bernardino, USA
Elliot Rodger	2014	13	7	20	Manifesto	Isla Vista, USA
Justin Bourque	2014	2	3	5	Social media (Facebook)	Moncton, Canada

In application, LIWC has been utilized to reasonably identify lying (Almela, Valencia-Garcia, & Cantos, 2012; Newman, Pennebaker, Berry, & Richards, 2003), predict forms of aggression (Pennebaker, 2011), evaluate suicide notes in a variety of contexts (Lester, 2007, 2009, 2014; Lightman, McCarthy, Dufty, & McNamara, 2007), identify psychopathy (Hancock et al., 2013), and most recently to examine pronoun usage and narcissism (Carey et al., 2015). As such, it was considered a robust tool appropriate for analyzing data that address legacy tokens of spree killers as well as for making comparisons with suicide notes and student writing samples. A discussion of textual analysis methods and computerized programs in addition to LIWC can be found elsewhere (Tausczik & Pennebaker, 2010).

## Method

### Spree Killers Data

The rarity of spree killings presents sampling challenges. Although active shooter events (a subset of spree killers) are on the rise (Blair & Martaindale, 2013), the overall incidence and impact of spree killers has changed little over the years (Huff-Corzine et al., 2014). Therefore, sampling a sufficient amount of data to test our research hypotheses proved difficult. Original documents were collected from news organizations, law enforcement, and a variety of Internet sources (provided that at least two independent Internet sources could verify the data).

The data comprise journal entries, school papers, suicide notes, manifestos, essays, transcriptions of YouTube videos, and social media postings. A total of 21 legacy tokens were collected for analysis (see Table 1). All documents were re-typed, spell checked, and visually inspected for accuracy by proofreading from authors to ensure that there was no degradation in the text during the preparation of the documents before entry into LIWC. A total of seven legacy tokens were taken from schoolshooters.info (<http://www.schoolshooters.info>), which is a dedicated website for the investigation of school shooters from a psychiatric perspective run by Dr. P. Langman. Langman is known for his work on school shooters through his book (Langman, 2009b) and recent publications (Langman, 2009a, 2010, 2013). Each killer's individual writing served as the unit of analysis. All tokens were verified as often as possible by confirmation of at least two sources.

The 21 tokens ranged between 187 and 108,190 words, with a mean of 7,169, median 984 words ( $SD = 23,337$ ). All but one of the writings was done before the spree killing occurred. Removal of Elliot Roger's manifesto brings the mean to 2,118 ( $SD = 3,054$ ), with a maximum of 11,483. The majority of the tokens were between 400 and 1,850 words long (see Table 2).

### Suicide Notes Data

An open-source search of suicide notes was conducted of published suicide notes through public sources such as

newspapers, magazines, etc., through LexisNexis from 1990 to 2014 with the phrase *suicide note*. The search yielded 1,025 results, many of which were discarded for redundancy. The remaining search results (located only in newspapers and magazines at this point) were inspected for unabridged suicide notes that had been replicated or provided in image format. Any notes that were published as an image were transcribed to allow for further analysis. Many of the notes were anonymous in the publications found by the authors. For consistency, any identifiers for the portion of the sample that could be identified were removed before analysis. In total, 25 suicide notes were collected for analysis (range = 17–4,094 words,  $M = 569.88$ ,  $SD = 897.48$ ,  $Mdn = 180$ ). Removal of the longest suicide note (male, early 20s, college education) brought the average to 422 words ( $SD = 516$ ,  $Mdn = 174$ ).

Suicide notes are a private, often restricted phenomena. Although many suicide notes are publicly available, they are often from individuals who are somewhat famous, talented writers, or in some cases both (e.g., George Eastman, Kurt Cobain). Still other notes seemed dated, or were from people who committed such horrible atrocities that their psychological well-being was considered too far afield for comparison with a college sample (i.e., Virginia Woolf, 1941, and Adolf Hitler, 1945). Finally, determining the authenticity of some notes proved difficult (e.g., Ernest Hemingway's alleged suicide note). In light of this, writers, actors, politicians, and persons of historical significance were omitted from analysis, narrowing the field of data considerably.

The remaining suicide notes selected for analysis had to be from 1990 or after, and were transcribed where possible from archival photos of the actual documents published in newspapers or other media outlets. The year 1990 was selected as a cut-off because it was hoped that people who were alive until that point had at least some common experiences with many adults still alive today. Where no original source could be found (a photographic copy of the note), validation was sought through the repetition of the same note in multiple sources (i.e., multiple publications reporting the same content in the note). In total, 25 notes were found and transcribed for analysis with three having to be confirmed through multiple sources. Gender and age were maintained where possible so that future investigations could determine how these variables impact writing style. Age could not be determined precisely in one case, where the suicide was listed as a teen. As an exact age could not be computed, this case was not included in age-oriented analyses, but was included in all other analyses where age was not a factor.

The suicide notes ranged from 17 to 4,094 words in length ( $M = 569.88$ ,  $Mdn = 180$ ,  $SD = 897.48$ ). As best as could be determined, all notes were written by individuals who had successfully completed suicide actions with the attempt associated with the note used for analysis. Ages ranged from 13 to 74 years for males ( $n = 14$ ) and 16 to 61 years for females ( $n = 11$ ). The notes of males averaged 796.35 words ( $SD = 1,148.27$ ) and those of females averaged 281.64 words ( $SD = 236.32$ ).



Table 2. Summary of data by type, age, sex, and length of writing

Type	Name (if known)	Age	Sex	Number of words
Spree killer	Charles Whitman	25	M	817
Spree killer	Marc Lépine	25	M	518
Spree killer	Timothy McVeigh	33	M	960
Spree killer	Luke Woodham	16	M	1,402
Spree killer	Kipland Kinkel	16	M	1,526
Spree killer	Dylan Klebold	17	M	1,267
Spree killer	Eric Harris	18	M	6,711
Spree killer	Mark O Barton	44	M	440
Spree killer	Jeffrey Weise	16	M	8,343
Spree killer	Erik Hainstock	15	M	1,665
Spree killer	Seung-Hui Cho	23	M	2,090
Spree killer	Pekka-Eric Auvinen	18	M	1,855
Spree killer	Robert Hawkins	19	M	330
Spree killer	Jiverly A. Wong	41	M	555
Spree killer	Michael McLendon	28	M	187
Spree killer	Wellington Oliveira	23	M	376
Spree killer	Jared Lee Loughner	23	M	984
Spree killer	T.J. Lane	17	M	385
Spree killer	Christopher Dorner	33	M	11,483
Spree killer	Elliot Rodger	22	M	108,190
Spree killer	Justin Bourque	24	M	468
Suicide	Anonymous	13	M	17
Suicide	Anonymous	16	M	2,039
Suicide	Anonymous	21	M	1440
Suicide	Anonymous	21	M	365
Suicide	Anonymous	28	M	4,094
Suicide	Anonymous	30	M	1,501
Suicide	Anonymous	40	M	62
Suicide	Anonymous	45	M	272
Suicide	Anonymous	48	M	365
Suicide	Anonymous	51	M	46
Suicide	Anonymous	52	M	168
Suicide	Anonymous	67	M	55
Suicide	Anonymous	74	M	136
Suicide	Anonymous	<sup>a</sup>	M	623
Suicide	Anonymous	16	F	180
Suicide	Anonymous	21	F	473
Suicide	Anonymous	21	F	83
Suicide	Anonymous	27	F	651
Suicide	Anonymous	34	F	155
Suicide	Anonymous	37	F	147
Suicide	Anonymous	50	F	64
Suicide	Anonymous	52	F	340
Suicide	Anonymous	56	F	168
Suicide	Anonymous	59	F	735
Suicide	Anonymous	61	F	102

Note. <sup>a</sup>Exact age not available, but was described as a teen.  
M = male. F = female.

## Statistical Analysis

To address Hypotheses 1–6, a series of one-way ANOVAs were conducted comparing the linguistic categories of each hypothesis between the student groups, suicide notes, and spree killer writings. Student groups are reported from LIWC (<http://www.liwc.net/>) comprising the collection of over 20,000 student writings. Additional post hoc testing with Tukey's honest significant difference (HSD) was used to determine the differences between groups. Even though we did not hypothesize differences between all groups, determining the extent of the differences of all categories offers obvious benefit, and was therefore included in these expanded analyses. For all significant differences, the  $p$  value of Tukey's HSD is reported. Where nonsignificant results are reported, the ANOVA  $p$  value is used.

H1 evaluated the negative emotion of spree killers ( $M = 3.73$ ,  $SD = 1.50$ ) and compared them with students ( $M = 1.78$ ,  $SD = .73$ ) and suicide cases ( $M = 1.80$ ,  $SD = 1.39$ ); results indicated significant differences between spree killers and students,  $F(4, 51) = 6.47$ ,  $p = .05$ , and suicides,  $F(4, 51) = 6.47$ ,  $p < .01$ . Therefore, H1 is fully supported. Similarly, H2 evaluated the anger of spree killers ( $M = 2.01$ ,  $SD = 1.31$ ) and compared them with students ( $M = 0.54$ ,  $SD = 0.24$ ) and suicide cases ( $M = 0.49$ ,  $SD = 0.56$ ); results indicated significant differences between spree killers and students,  $F(4, 51) = 8.41$ ,  $p = .02$ , and suicides  $F(4, 51) = 8.41$ ,  $p < .01$ . Therefore, H2 is also fully supported. H3 evaluated sexual references of spree killers; however, results indicated no significant differences between spree killers and students and suicides,  $F(4, 51) = 2.02$ ,  $p = .1$ . Therefore, H3 is not supported. H4 evaluated personal pronoun usage in the suicide notes ( $M = 17.27$ ,  $SD = 5.64$ ) and compared them with the writings of students ( $M = 6.83$ ,  $SD = 2.93$ ) and spree killers ( $M = 6.24$ ,  $SD = 4.19$ ); results indicated significant differences between suicide notes and the writings of students,  $F(4, 51) = 19.66$ ,  $p < .01$ , and spree killers,  $F(4, 51) = 19.66$ ,  $p < .01$ . Therefore, H4 is fully supported. H5 evaluated the use of future tense in suicide notes ( $M = 6.38$ ,  $SD = 3.04$ ) and compared it with the use of future tense by students ( $M = 1.17$ ,  $SD = .21$ ) and spree killers ( $M = 1.20$ ,  $SD = .54$ ); results indicated significant differences between suicide notes and the writing of students,  $F(4, 51) = 21.40$ ,  $p < .01$ , and spree killers,  $F(4, 51) = 21.40$ ,  $p < .01$ . Therefore, H5 is also fully supported. Finally, H6 evaluated the use of positive emotion in suicide notes versus the writing of students and spree killers; however, results indicated no significant differences between suicide notes and student writings,  $F(4, 51) = 0.58$ ,  $p = .68$ . Therefore, H6 is not supported (see Table 3).

## Results

This manuscript explored a variety of semantic characteristics associated with spree killers and individuals who had completed suicide to determine whether these lexical characteristics are generalizable and indicative of different types of harm. Determining vocabulary usage related to words

Table 3. Variable characteristics for ANOVAs

H	Variable	<i>M</i>	<i>SD</i>	Group differences
1	Negative Emotion			
	Student	1.78	.73	$F(4, 51) = 6.47, p = .05$
	Suicide	1.80	1.39	$F(4, 51) = 6.47, p < .01$
	Spree killer	3.73	1.50	
2	Anger			
	Student	.54	.24	$F(4, 51) = 8.41, p = .02$
	Suicide	.49	.56	$F(4, 51) = 8.41, p < .01$
	Spree killer	2.01	1.31	
3	Sexual References			
	Student	1.27	.39	N/A
	Suicide	1.44	1.32	N/A
	Spree killer	1.30	.97	
4	Personal Pronoun			
	Student	6.83	2.93	$F(4, 51) = 19.66, p < .01$
	Suicide	17.27	5.64	
	Spree killer	6.24	4.19	$F(4, 51) = 19.66, p < .01$
5	Future Tense			
	Student	1.17	.21	$F(4, 51) = 21.40, p < .01$
	Suicide	6.38	3.04	
	Spree killer	1.20	.54	$F(4, 51) = 21.40, p = \leq .01$
6	Positive Emotion			
	Student	2.74	.23	N/A
	Suicide	2.76	1.78	
	Spree killer	3.03	1.30	N/A

Note. Bold denotes primary group for comparisons in each variable. H = hypothesis. N/A = not applicable.

associated with negative emotions, anger, personal pronoun usage, and future tense use provides a foundation for investigating student writings in academic settings, thereby offering crisis responders another evaluative tool.

Specifically, the use of negative emotion (H1) and anger (H2) is particularly relevant for differentiating those who have harm ideation versus other ideation from either suicidal or student populations. Each of these characteristics is exhibited 2 (H1) to 4 (H2) times as often in spree killer writings than in either suicidal or general student texts. These variables manifest substantially more frequently in violent populations than in nonviolent populations, and further development in spree killer identification will likely reduce the standard deviation of these variables, further pronouncing the differences. Personal pronoun usage (H4) and future tense usage (H5) in suicide notes manifested even stronger than negative emotion (H1) and anger (H2) did in spree killer writings. Personal pronouns were nearly 3 times as likely in suicidal writings than in other writings, and future tense usage was nearly 4 times as likely in suicidal writings as compared with other types of writing.

For H1 and H2 of shooter writings, and H4 and H5 of suicidal writings, these variables only distinguish that par-

ticular behavior from the others – there is very little bleed-through. In other words, spree killers focus on negative emotion and anger, whereas suicidal and student writings exhibit similar levels of this behavior. Conversely, suicidal individuals strongly emphasize personal pronoun usage and future tense orientation, whereas killers and students are about equal in this regard. It could be the case that these variables tap into a deeper cognitive construct that is unique to internal or external harm orientation.

## Limitations and Future Research

This work is limited by its sample sizes and its preliminary nature. Currently legacy tokens that are appropriate for analysis are elusive, expensive, and time consuming to locate. Similarly, suicide notes that can be validated are also rare. The comparison groups from LIWC have aggregated standard deviations, which makes it difficult to identify interesting characteristics of any subpopulations within those groups. Future research should seek to collect unique data from appropriate student samples for comparison.

Specific differences in suicide notes exist between the sexes. Women express more references to positive feelings, references to cognitive processes, and personal pronoun usage than men do (Lester, Haines, & Williams, 2010). Understanding sex differences in suicidal ideation becomes increasingly important when previous work has noted that females experience higher levels of depression, suicidal ideation, and suicide attempts than men (Lester, 2014). However, the sample was insufficient in size to investigate these differences, and the preliminary nature of this investigation cannot be ignored. Future research should evaluate what impacts the sex of the writer has in these contexts.

## Conclusion

This project is a first step toward examining whether or not quantitative analytic techniques can be used to analyze the writing of those who have capacity for harm of the self or others. At least four specific instances can be envisioned where this methodology may prove to be a benefit for violence prevention.

First, as a longitudinal, within-subject measure. Individuals who seem to be at risk could be identified over time so as to track progress either toward or away from violent ideation by comparing their writings not only with suicidal or spree killer writings, but also with their past writings over time. In this way, inclinations toward (or away from) violent behavior could be assessed to determine when interventions should be used, and how successful those interventions are on a writing-specific metric.

Second, as a between-subject measure at any point in time. College is a stressful experience, particularly with changing workloads throughout the semester that are further compounded by other life changes such as sleep pattern disruption or impending life changes such as work

(Ross, Niebling, & Heckert, 1999). During the academic year, this methodology could be employed to actively identify individuals who may be struggling with negative ideations and who have not self-selected treatment. In this way, administrators are provided with a proactive tool to help conduct earlier interventions with people who may be at risk to self-harm or to harm others.

Third, this tool could be used as a group assessment measure. After a campus tragedy, natural disaster, etc., affected persons could be grouped together and have their progress tracked overtime to help assess their health as individuals and as a collective. For instance, an institution may have an event that occurs within a subpopulation such as an athlete being killed in a traffic accident. The institution would begin to provide as many resources as possible to help the whole community deal with the event, but may want to put additional emphasis on teammates, friends, roommates, and significant others who had an especially close relationship with the deceased. The methodology discussed here would allow for assessment as a group to determine which individuals are members of the affected community, how that subcommunity was affected, and how well they are recovering as a group as well as individuals.

Fourth, this measure could be used as a national assessment tool to try to discern patterns that lead to suicidal or homicidal ideations. These events are rare, and many institutions may have difficulty sharing information on a substantial enough scale to further refine effective intervention, prevention, and recovery measures that would benefit the broader academic community. Having a standardized tool to assess these ideations creates a comparative metric that allows for meaningful investigations of other related factors that may lead to negative behaviors. Instead of burdening individual crisis teams to assess cases and share findings with each other, a large database could easily be constructed that allows for greatly improved accuracy and information sharing between stakeholders to prevent these events.

## References

- Albrecht, S. (2010). Threat assessment teams: Workplace and school violence prevention. *FBI Law Enforcement Bulletin*, 79(2), 15.
- Almela, Á., Valencia-García, R., & Cantos, P. (2012). Seeing through deception: A computational approach to deceit detection in written communication. In *Proceedings of the Workshop on Computational Approaches to Deception Detection* (pp. 15–22). Stroudsburg, PA: Association for Computational Linguistics.
- Black, S. T. (1993). Comparing genuine and simulated suicide notes: A new perspective. *Journal of Consulting and Clinical Psychology*, 61, 699–702.
- Blackman, P., Leggett, V. L., Olson, B. L., & Jarvis, J. P. (Eds.). (1999). The varieties of homicide and its research. *Proceedings of the 1999 Meeting of the Homicide Research Working Group*. Quantico, VA: FBI Academy, FBI, US Department of Justice.
- Blair, J. P., & Martaindale, M. H. (2013). *United States active shooter events from 2000 to 2010: Training and equipment implications*. San Marcos, TX: Texas State University.
- Carey, A. L., Brucks, M., Küfner, A. C. P., Holtzman, N. S., große Deters, F., Back, M. D., ... Mehl, M. R. (2015). Narcissism and the use of personal pronouns revisited. *Journal of Personality and Social Psychology*. Advance online publication. <http://dx.doi.org/10.1037/pspp0000029>
- Choi, D., Ko, B., Kim, H., & Kim, P. (2013). Text analysis for detecting terrorism-related articles on the web. *Journal of Network and Computer Applications*, 38, 16–21.
- Edelstein, A. (2014). Re-thinking typologies of multiple murders: The missing category of serial-mass murder and its theoretical and practical implications. *International Journal of Emergency Mental Health and Human Resilience*, 16, 122–126.
- Federal Bureau of Investigation. (2008). *Serial murder, multi-disciplinary perspectives for investigators. Behavioral analysis unit for violent crime*. Retrieved from <http://www.fbi.gov/stats-services/publications/serial-murder>
- Fox, J. A., & DeLateur, M. J. (2014). Mass shootings in America. Moving beyond Newtown. *Homicide Studies*, 18(1), 125–145.
- Fox, J. A., & Levin, J. (1998). Multiple homicide: Patterns of serial and mass murder. *Crime and Justice*, 23, 407–455.
- Fox, J. A., & Levin, J. (2007). *Extreme killing: Understanding serial and mass murder*. Thousand Oaks, CA: SAGE Publications.
- Gonzales, A. L., Hancock, J. T., & Pennebaker, J. W. (2010). Language style matching as a predictor of social dynamics in small groups. *Communication Research*, 37, 3–19.
- Handelman, L. D., & Lester, D. (2007). The content of suicide notes from attempters and completers. *Crisis*, 28, 102–104.
- Hancock, J. T., Woodworth, M. T., & Porter, S. (2013). Hungry like the wolf: A word-pattern analysis of the language of psychopaths. *Legal and Criminological Psychology*, 18(1), 102–114.
- Huff-Corzine, L., McCutcheon, J. C., Corzine, J., Jarvis, J. P., Tetzlaff-Bemiller, M. J., Weller, M., & Landon, M. (2014). Shooting for accuracy: Comparing data sources on mass murder. *Homicide Studies*, 18(1), 105–124.
- Langman, P. F. (2009a). Rampage school shooters: A typology. *Aggression and Violent Behavior*, 14, 79–86.
- Langman, P. F. (2009b). *Why kids kill: Inside the minds of school shooters*. London, UK: Macmillan.
- Langman, P. F. (2010). *Expanding the sample: Five school shooters*. Retrieved from <http://www.schoolshooters.info>
- Langman, P. F. (2013). *Psychiatric medications and school shooters*. Retrieved from <http://www.schoolshooters.info>
- Lester, D. (2007). Suicide and culture. In D. Lester & J. R. Rogers, *Suicide: A global issue* (pp. 209–232). Santa Barbara, CA: ABC-CLIO.
- Lester, D. (2009). Learning about suicide from the diary of Cesare Pavese. *Crisis*, 30, 222–224.
- Lester, D. (2014). College student stressors, depression, and suicidal ideation. *Psychological Reports*, 114, 293–296.
- Lester, D., Haines, J., & Williams, C. L. (2010). Content differences in suicide notes by sex, age, and method: A study of Australian suicide notes. *Psychological Reports*, 106, 475–476.
- Lightman, E. J., McCarthy, P. M., Dufty, D. F., & McNamara, D. S. (2007). Using computational text analysis tools to compare the lyrics of suicidal and non-suicidal songwriters. In D. S. McNamara & G. Trafton (Eds.). *Proceedings of the 29th Annual Cognitive Science Society*. Hillsdale, NJ: Erlbaum.
- Mann, J. J., Apter, A., Bertolote, J., Beautrais, A., Currier, D., Haas, A., ... Hendin, H. (2005). Suicide prevention strategies: A systematic review. *JAMA*, 294, 2064–2074.
- McGee, J., & DeBernardo, C. (1999). The classroom avenger: A behavioral profile of school based shootings. *The Forensic Examiner*, 8, 16–18.
- Meloy, J. R., & Hoffmann, J. (Eds.). (2013). *International handbook of threat assessment*. Oxford, UK: Oxford University Press.



- Morton, R. J., & Hilts, M. A. (2005). *Serial murder: Multi-disciplinary perspectives for investigators*. Washington, DC: U.S. Department of Justice.
- National Action Alliance for Suicide Prevention: Youth in Contact with the Juvenile Justice System Task Force. (2013). *Screening and assessment for suicide prevention: Tools and procedures for risk identification among juvenile justice youth*. Washington, DC: National Action Alliance for Suicide Prevention.
- Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, 29(5), 665–675.
- Ogilvie, D., Stone, P., & Shneidman, E. (1966). Some characteristics of genuine versus simulated suicide notes. In P. Stone, D. Dunphy, M. Smith, & D. Ogilvie (Eds.). *The General Inquirer: A computer approach to content analysis* (pp. 527–535). Cambridge, MA: MIT Press.
- Osgood, C. E., & Walker, E. G. (1959). Motivation and language behavior: A content analysis of suicide notes. *The Journal of Abnormal and Social Psychology*, 59, 58–67.
- O'Toole, M. (2000). The school shooter: A threat assessment perspective. Quantico, VA: Critical Incident Response Group, National Center for the Analysis of Violent Crime, FBI Academy.
- O'Toole, M. E. (2010). *The school shooter: A threat assessment perspective*. Quantico, VA: Federal Bureau of Investigations.
- Pennebaker, J. W. (2007). Current issues and new directions in Psychology and Health: Listening to what people say – the value of narrative and computational linguistics in health psychology. *Psychology and Health*, 22, 631–635.
- Pennebaker, J. W. (2011). Using computer analyses to identify language style and aggressive intent: The secret life of function words. *Dynamics of Asymmetric Conflict*, 4(2), 92–102.
- Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). Linguistic Inquiry and Word Count (LIWC2001) [Software]. Austin, TX: LIWC.net.
- Pennebaker, J. W., & Francis, M. E. (1996). Cognitive, emotional, and language processes in disclosure. *Cognition & Emotion*, 10, 601–626.
- Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). *The development and psychometric properties of LIWC2007*. Austin, TX: LIWC.net.
- Pennebaker, J. W., & Stone, L. D. (2003). Words of wisdom: Language use over the life span. *Journal of Personality and Social Psychology*, 85(2), 291.
- Pestian, J., Nasrallah, H., Matykiewicz, P., Bennett, A., & Leenaars, A. (2010). Suicide note classification using natural language processing: A content analysis. *Biomedical Informatics Insights*, 3, 19–28.
- Ross, S., Niebling, B. C., & Heckert, T. M. (1999). Sources of stress among college students. *College Student Journal*, 33, 312–318.
- Simons, A., & Tunkel, R. (2013). The assessment of anonymous threatening communications. *International Handbook of Threat Assessment*. New York, NY: Oxford University Press.
- Stirman, S. W., & Pennebaker, J. W. (2001). Word use in the poetry of suicidal and nonsuicidal poets. *Psychosomatic Medicine*, 63, 517–522.
- Stone, P. J., & Hunt, E. B. (1963). A computer approach to content analysis: Studies using the General Inquirer System. In *Proceedings of the May 21-23, 1963, Spring Joint Computer Conference* (pp. 241–256). New York, NY: ACM.
- Shneidman, E. S., & Farberow, N. L. (1956). Clues to suicide. *Public Health Reports*, 71, 109–114.
- Shneidman, E. S., & Farberow, N. L. (1957). *Clues to suicide*. New York, NY: McGraw-Hill.
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29, 24–54.
- Verlinden, S., Hersen, M., & Thomas, J. (2000). Risk factors in school shootings. *Clinical Psychology Review*, 29, 3–56.

Received January 21, 2015

Revision received August 3, 2015

Accepted August 8, 2015

Published online January 19, 2016

#### About the authors

Michael Egnoto is an assistant clinical professor of FIRE and a researcher at START in the Risk Communication and Resilience division at the University of Maryland, USA. His work focuses on quantitative assessments of risk and crisis communication during natural, man-made intentional, and man-made unintentional disasters.

Darrin J. Griffin is an assistant professor in the College of Communication & Information Sciences at the University of Alabama, USA. His work focuses on deceptive communication, nonverbal behavior, spree-killer research, and tactics relevant to law enforcement.

#### Michael J. Egnoto

FIRE Risk Communication and Resilience  
National Consortium for the Study of Terrorism and Responses to Terrorism (START)  
University of Maryland  
College Park, MD 20740  
USA  
Tel. +1 (301) 314-1873  
Fax +1 (301) 314-1980  
E-mail megnoto@umd.edu