

Social Sciences

Read Online
24/7

WILEY

[Back to old version](#)

Journal of Computer-Mediated Communication

Volume 10, Issue 2

January 2005

Page 00

Gender, Identity, and Language Use in Teenage Blogs

David A. Huffaker , Sandra L. Calvert

First published:January 2005 [Full publication history](#)**DOI:**

10.1111/j.1083-6101.2005.tb00238.x

[Citing literature](#)

Address: Northwestern University, Frances Searle Building Room 2-147, 2240 Campus Drive, Evanston, IL 60208 USA

Address: Department of Psychology, Georgetown University, 309-C White Gravenor Building, 37th & O Streets, Washington, DC 20057 USA

Abstract

This study examines issues of online identity and language use among male and female teenagers who created and maintained weblogs, personal journals made publicly accessible on the World Wide Web. Online identity and language use were examined in terms of the disclosure of personal information, sexual identity, emotive features, and semantic themes. Male and female teenagers presented themselves similarly in their blogs, often revealing personal information such as their real names, ages, and locations. Males more so than females used emoticons, employed an active and resolute style of language, and were more likely to present themselves as gay. The results suggest that teenagers stay closer to reality in

their online expressions of self than has previously been suggested, and that these explorations involve issues, such as learning about their sexuality, that commonly occur during the adolescent years.

Enhanced Article Feedback

Introduction

Identity is a central task that begins in infancy and ends with the culmination of one's life. Its role in adolescent development has been particularly important as youth come to know and define themselves in ways that were not possible during their childhood (Calvert, 2002). More specifically, the ability to reflect on one's own thoughts, and hence on one's self, adds a new dimension to self-discovery, particularly of one's sexual identity.

As adolescents seek to define who they are near the beginning of the 21st century, their forums for self-discovery have expanded. One place that adolescents now spend a considerable amount of time is in online settings, and these online venues, such as multi-user domains (MUDs), have been linked to identity exploration (Turkle, 1995). One of the newest venues for exploration is the weblog, a reversed-chronological online journal, which is used in a variety of ways, but often as a personal journal or ongoing commentary about oneself (Herring, Scheidt, Bonus, & Wright, 2004a; Huffaker, 2004a). Our purpose here is to examine how adolescents use weblogs to explore their identity. In particular, we examine the language and emotional codes that adolescents use to express themselves in weblogs.

Background: Identity, Adolescent Development, and Online Discourse

Identity is often characterized in terms of one's interpersonal characteristics, such as self-definition or personality traits, the roles and relationships one takes on in various interactions, and one's personal values or moral beliefs (Calvert, 2002). Identity also involves a sense of continuity of self images over time (Grotevant, 1998), a continuity that may be disrupted when puberty creates radical alterations in one's physical appearance. With sexual maturation comes changes in the roles that one is expected to assume with members of the opposite sex, for the adolescent is increasingly expected to assume a sexual identity, one of the markers and anchors of a mature identity (Erikson, 1993; Grotevant, 1998).

For Erikson (1993), a unitary sense of identity is constructed after a successful search for *who one is*. However, other perspectives of adolescent development view the construction of self as one that involves multiple “public” selves which are presented according to the demands and constraints of particular situations (Harter, 1998). Within the social interactionist perspective, adolescents take on the roles of others through playful stances where they assume different perspectives, thereby allowing them to try on different facets of who they will become. For the social interactionist, language is a key means through which those roles are

explored and constructed (Harter, 1998).

While physical constraints such as the body, biological sex, race, or age can have a profound effect on self-definition and self-presentation (Collins & Kuczaj, 1991), many of these attributes become flexible in online environments. In a virtual world, one even gets to construct one's body. The anonymity afforded to youth within virtual worlds allows adolescents more flexibility in exploring their identity through their language, their role play, and the personae they assume (Calvert, 2002).

The most pronounced search for identity in the adolescent years generally begins with puberty and can end at various social markers, such as the acquisition of certain legal markers (e.g., the right to vote, or formal adult status, that are acquired at age 18 in the United States of America). For our purposes, we will focus on an intersection of those adolescent years and the teenage years, generally defined as ages 13-19. Specifically, our sample will begin at age 13 and extend through age 17, before legal status is acquired, since many youth still live at home during those years, thereby providing a more homogenous sample than if some of our sample was in junior high school while others were in college. We will refer to this 13-17 year-old age group as adolescents or teens.

In contemporary U.S. society, many adolescents spend considerable amounts of time in online interactions (Subramanyam, Greenfield, Kraut, & Gross, 2002). The language used on the Internet demonstrates an evolution of discourse (Crystal, 2001), and adolescents are in the midst of that language evolution (Greenfield & Subrahmanyam, 2003).

Sometimes referred to as *netspeak*, the language of the Internet entails both traditional linguistic forms and adapted ones that include slang and non-standard forms that are sometimes used in offline life. Netspeak is an emergent discourse that is shaped entirely by the creativity of its community (Crystal, 2001). The introduction of acronyms (e.g., “lol = laugh out loud,” “brb = be right back”), plays or variations on words (e.g., “cya = see you”, “latah = later”), graphical icons that represent emotions, called *emoticons* (e.g., :) or :-), or graphical icons that represent a real person in a virtual context, called *avatars*, are all examples of language produced by online communicators. This language continues to evolve and remains an important area of study when considering the ways in which Internet users interact and express who they are.

Scholars have actively explored how identity and language are manifested in online interactions. Current research in computer-mediated communication (CMC) environments such as chat rooms, newsgroups, and MUDs has revealed interesting trends in the way individual identity is presented, language is used, and interactions have transpired (Calvert, 2002; Crystal, 2001; Greenfield & Subrahmanyam, 2003; Herring, 2000; Turkle, 1995). Yet, as new Internet applications are created and embraced, CMC studies must continue to strive toward the understanding of online identity, language, and interaction.

Such is the case for weblogs or “blogs.” Blogs are personal journals or reversed-chronological commentaries written by individuals and made publicly accessible on the web, and they have distinctive technological features that set them apart from other forms of CMC (Herring et al., 2004a; Huffaker, 2004b). These features include: 1) ease-of-use, as users do not need to know HTML or other web programming languages to publish onto the Internet; 2) ways to archive information and knowledge; 3) opportunities for others to comment or provide feedback for each blog post; and 4) links to other “bloggers” to form online communities.

These features are especially important for constructing online identity. First, the lack of technical expertise needed to create or maintain blogs makes the application more accessible regardless of gender and age. Next, the ability to archive blog posts creates a way to scaffold on previous impressions and expressions; thus, constructing identity can be a continual process for adolescents, and one to which they can refer. Finally, when blog software offers ways to provide feedback or link to other bloggers, this can foster a sense of peer group relationships, another important aspect for the developing adolescent. In short, weblogs represent a new medium for computer-mediated communication and may offer insight into the ways in which adolescents present themselves online, especially in terms of self-expression and peer group relationships, both of which impact the construction of identity.

Age and Gender Characteristics of Bloggers

Most surveys suggest that a significant portion of the total blog population is inhabited by teenagers, and the split between genders is relatively close. *Perseus Development Corporation*, for instance, finds blogs are dominated by youth, with 52% of all blogs being developed and maintained by teens that are ages 13-19. They also find 56% of the *total* bloggers are female and 44% are male (Henning, 2003). A similar study of 203 randomly-selected weblogs finds that 54% have male authors and 46% have female authors, as well as that 40% of blog authors are under age 20 (Herring et al., 2004a).

A later study of 358 randomly selected blogs found 52% male and 48% female total bloggers, with 39% of bloggers being less than 20 years old. However, they also found more females than males in the 'teen' category (Herring, Kouper, Scheidt, & Wright, 2004b). *BlogCensus* randomly sampled 490,000 blogs to find 40% male and 36% female, with the rest of the blogs unidentifiable in terms of gender.² Finally, *Jupiter Research* found that blogging is split evenly between the genders (Greenspan, 2003). Thus, not only is a significant portion of bloggers under 20 years old, the distribution between males and females demonstrates the egalitarian acceptance and usage of weblogs (Herring et al., 2004b).

Given the heavy use of weblogs by both male and female adolescents, it is timely to link their online weblog creations to their emerging self-constructions. Because many aspects of identity, both online and offline, involve language use, we examine online identity as the disclosure of personally identifying information, emotive features, sexual identity and semantic themes through discourse expressed in online weblogs.

Disclosure of Personal Information in Blogs

Some CMC applications, such as Instant Messenger (IM) or MUDs, require names (user IDs) to utilize the application, but blogs do not. They can be constructed in ways that leave the author completely unidentified, and users can post anonymously. This is an interesting feature when names, age, and even location can be strongly associated with both online and offline identity. How often do adolescent bloggers explore identity in terms of names or other personal information? How flexible are these explorations of identity? Herring, Scheidt, Bonus, and Wright (2004a) find that 92% of blog authors provide a name, whether it is a full name (31%), a first name (36%), or a pseudonym (29%). Also, more than half (54%) of blog authors provide explicit demographic information such as age, occupation, or geographic location on the home page of their blogs (Herring et al., 2004a).

It is intriguing that blog authors often reveal their real names, along with other personal information such as age or location. In a virtual world, where identity is flexible, why would authors choose to present themselves as they do in nonvirtual worlds? Perhaps the idea of the personal journal encourages authors to reveal exactly who they are. Perhaps there is a certain sense of empowerment in revealing thoughts and feelings without hiding behind a public mask. On the other hand, if the content of a blog is personal, candid, or intimate, would not a certain sense of anonymity make authors feel more comfortable in being explicit? In either case, it would seem that names, age, and other forms of personal information play an important role in creating and maintaining blogs since these kinds of descriptors reflect the self, and hence how blog authors want to present themselves to others.

Gender and Emoticons

Emoticons, also referred to as *smileys*, derive from the hybrid of “emotion” and “icons”, and are either composed of punctuation characters or of graphical symbols. Because online interactions lack the facial expressions and body gestures vital to expressing opinions and attitudes, emoticons were introduced to fill a void in online communication (Crystal, 2001).

How do emoticons affect the interpretation of a message? While one study suggests verbal content still outshines emoticons in the interpretation of a message (Walther & D'Addario, 2001), the impact of emoticons and smileys is not relegated to the message itself. They also help form impressions of the author's disposition or attitude. For instance, in a study of chat room moderators, the use of emoticons caused the moderator to be perceived as more “dynamic,” “friendlier,” “valuable,” and “talkative” than moderators who did not use emoticons (Constantin, Kalyanaraman, Stavrositu, & Wagoner, 2002a).

Are there gender variations in the use of emoticons in CMC? In a study of 3,000 online messages, Witmer and Katzman (1997) found that females used more graphical accents, including emoticons, to express emotion in their discourse than males (Witmer & Katzman, 1997), a pattern that is also reflected in adolescent males' tendency to deny their feelings, including sadness (Ruble & Martin, 1998). In an analysis using instant messaging dialogues, males rarely use emoticons in conversations with other males, but do use them with females, while females use an equal amount of emoticons in both male and female conversations (Lee, 2003). Similarly, an examination of online newsgroups finds that women are more apt to express emotions, but males will adapt in mixed-gendered newsgroups and express more (Wolf, 2000).

Sexual Identity in Teenagers

One component of gender roles that is a key developmental milestone during adolescence involves sexual orientation, or sexual identity (Huston, 1983). During adolescence, sexual orientations, whether heterosexual, homosexual, bisexual, or transgender, often emerge (Grotevant, 1998). While the challenges of assuming a mature sexual identity occur for all youth, these challenges may be particularly difficult for those who are gay, lesbian, bisexual, or transgendered. Historically, those with non-heterosexual orientations encountered discrimination, disparagement, and even punishment (Foucault, 1990). Even now, gay, lesbian, bisexual, and transgendered youth face erotic feelings and fantasies that they may not be able to discuss with their families and peers (Grotevant, 1998).

In a virtual world, where flexibility and anonymity are possible, adolescents may feel more comfortable expressing their sexual orientation and exploring their sexual identity beyond social prescriptions. In online forums, including weblogs, language is a key means through which sexual identity can be expressed and explored.

Gender and Language Use

Traditional gender roles embody the male role as agentive, where action, self-expansion, and individuality are the rule. By contrast, traditional gender roles define the female role as communal, embodying emotional expressiveness and a focus on the needs of others (Bakan, 1966). The work of Deborah Tannen suggests that the communication patterns of males and females often differ, with males using a direct and forceful style while females use a more indirect and intimate style of interaction (Tannen, 1995). Such linguistic styles parallel the masculine principle of agency and the feminine principle of communion.

Gendered language use in CMC contexts is not very different from that in face-to-face interactions and includes similar features of “verbosity, assertiveness, use of profanity, politeness (and rudeness), typed representations of smiling and laughter, and degree of interactive engagement” (Herring, 2000). There are, however, differences in the modes of CMC, which may have consequences for language use or social interactions. For instance, online chatting differs from the turn-taking patterns of face-to-face or telephone conversation because online posts appear out of sequence, thereby impacting language coherence (Herring, 1999; Greenfield & Subrahmanyam, 2003). Instant Messaging (IM) may foster intimacy among users, including self-disclosure and sentimental feelings, because it cultivates social connectedness (Hu, Smith, Westbrook, & Wood, 2003).

Robin Lakoff's (1975) theories on women's language suggest that females use a language style that reflects diffidence, shyness, and lower self-confidence, indicating a lack of commitment or strong opinion (Eckert & McConnell-Ginet, 2003). One device is euphemism, where a person uses words such as “fudge” or “heck” instead of profanity. Another device is the use of tag questions and hedges, such as “This weather is terrible, isn't it?” or “I kinda got angry.” Another device is indirection when there is a reluctance to commit to something, for instance, “Oh sorry, I've got a doctor's appointment around that time.” For Lakoff, women's language represents an overall conventional politeness (Eckert & McConnell-Ginet, 2003).

How do Lakoff's theories relate to gender within a CMC context? A study of 2,692 messages of Internet discussion groups finds that groups dominated by females tend to ‘self-disclose’ and avoid or attempt to reduce tension (Savicki, 1996). Similarly, Herring (2000) finds that women are “more likely to thank, appreciate and apologize, and to be upset by violations of politeness” (Herring, 2000).

In contrast, discussion groups dominated by males tend to use impersonal, fact-oriented language (Savicki, 1996), and males seem less concerned with politeness and sometimes violate expected online conduct (Herring, 2000). In an analysis of personal web pages, females are found to be “friendly” and “smiling”, while males present themselves as “confident” (Arnold & Miller, 1999).

Michelle Rodino (1997) disagrees with conceptualizing male and female language use in terms of a binary opposition. In her study of Internet-Relay Chat (IRC), she finds that IRC participants construct and express

gender in a variety of ways, and language use is flexible and dynamic (Rodino, 1997). Therefore, traditional gender and language research may oversimplify online language patterns because being “virtual” allows more freedom and flexibility (Rodino, 1997).

In either case, studying CMC can provide a variety of insights into the ways males and females present themselves and interact with others in online settings. For this study, gender similarities and differences in language use represent an important means by which adolescents form an online identity.

The Present Study

The purpose of this study is to examine gender similarities and differences among weblogs created by teenagers. By using content analyses of their weblogs, we examined how adolescents present their identities online, as well as how they use language to express their experiences and feelings. Of particular interest to us is: 1) the extent to which personal information, such as name, age, or location, is disclosed; 2) how emotive features are conveyed; 3) how sexual identity is intimated; and 4) how language is used to express ideas and feelings.

Hypotheses

Our hypotheses are as follows:

- H1: Males more so than females will provide personally identifying information, such as their name, age, location, and contact information.
- H2: In keeping with traditional gender roles, females will use emoticons to express their feelings more often than males.
- H3: In keeping with traditional gender roles, females will discuss intimate topics like their sexual identity more openly than males.
- H4: In keeping with traditional gender roles, males will use language that is more aggressive, resolute, and active than females.
- H5: In keeping with traditional gender roles, females will use language that is more passive, cooperative, and accommodating than males.

Methodology

To test these hypotheses, we conducted a content analysis of randomly-selected blogs that were created and maintained by teenagers. We examined the discourse, as expressed in language and emoticons, to conduct our analyses of weblog content.

Participants

Participants were randomly selected weblogs created by teenage males and females. These blogs resided on sites that provide hosting and administration services and include *LiveJournal* (n=30), *Xanga* (n=6), *Blogspot* (n=27), *Blog-City* (n=2), *t-Blog* (n=1), and *Journalspace* (n=1), except for three cases, where the

blog is a personally hosted web site with a vanity URL such as “www.myname.com.”

The blogs were retrieved using two weblog search engines, as well as from *Blogspot* and *LiveJournal*,³ two of the oldest and most popular blog hosting sites. The search engines were *Blogsearchengine*,⁴ which categorizes blogs into primary content areas such as “entertainment blogs,” “writing blogs,” or “teen blogs,” and *blo.gs*,⁵ which provides a directory of recently updated weblogs and allows users to track favorite blogs.

Our sampling methods may be biased in two ways: 1) the majority of sites are from LiveJournal and Blogspot, the two most popular free blog hosting sites, which may include a high percentage of novice users; and 2) Blogsearchengine and blo.gs may only link to or index blogs that conform to their technical specifications.

Blogs were retrieved using search strings such as “teens,” “teen blogs,” and “teenager,” yielding 347 teen blogs. After reviewing each link, and removing inactive URLs or links to sites created by adults (including 18- and 19-year-olds), 184 blogs remained (males = 63, females = 121), which represents the final sample of active teenage blogs. Because of the abundance of female blog results, the sample was separated by gender and a stratified random sampling method was employed to even the sample. The final sample consisted of a total of 70 weblogs, equally distributed across male and female authors.

Ages of the blog authors ranged from 13 through 17 (mean age = 15.47, $SD = 1.23$) when age was revealed on the site (67% of participants revealed their age). The mean age for the 26 males was 15.31 ($SD = 1.32$). The mean age for the 21 females was 15.67 ($SD = 1.11$). It must be noted that there was no way to validate the physical identities of blog authors. Because actual age or gender could be falsified in the virtual environment, this study could only explore the online personae that were displayed in the blogs.

Procedure

The front page of each weblog was analyzed and scored for personally identifiable information, emotive features, sexual identity, and gendered language. DICTION 5.0, a content analysis software package which evaluates documents in terms of word counts as well as content types and language tone, was used to create language scores for tone and semantic features.

There were two scorers. Inter-observer reliability was calculated for 29% of the blog sample for all dependent variables. The formula used was two times the number of agreements divided by the total number of scores for Scorer 1 and Scorer 2 for all dependent variables. Specific reliability scores are reported within their respective categories below.

Dependent Measures

Disclosure of Personal Information











Each blog was examined for the amount of personal information revealed in the text. This included: first name; full name; age; birth date, and location of the blog author; and contact information (email address, instant messaging user name, or a link to another personal web page or home page). If the blogger revealed an online handle or pseudonym, such as “MonkeyDork” or “CoolGrrl,” this was not scored as a real name. Inter-observer reliability for personal information was as follows: first name = 90%; full name = 90%; age =

91%; birth date = 85%; location = 85%; and contact information = 100%, which entails an email address, where inter-observer reliability = 100%, an instant messenger name = 95%, and a URL to a personal home page = 80%.

Emotive Features

The blogs were examined for any use of emoticons or smileys. The total number of emoticons was counted and divided into five categories: 1) happy; 2) sad; 3) angry; 4) flirty; and 5) tired. These categories were developed based on web sites⁶ dedicated to emoticon definitions and previous research on emoticon use in computer interactions (Danet, 1998, 2001; Danet, Ruedenberg, & Rosenbaum-Tamari, 1998). Table 1 provides examples of both text-based and graphical emoticons. Coders produced a total emoticon count for each type (happy, sad, etc.) as well as a count of graphical and text-based emoticons.

Table 1. Examples of text-based and graphical emoticons

TYPE OF EMOTICON	TEXT-BASED	GRAPHICAL
Happy	:) or :-)	 or 
Sad	:(or :-( or 
Angry	>:(or :O	 or 
Flirty	;) or :P	 or 
Tired	n/a	 or 

Inter-observer reliability was 85% for the use of emoticons and 95% for the number of instances of emoticons. Inter-observer reliability for type of emoticons was as follows: happy = 97%; sad = 90%; angry = 94%; flirty = 83%; and tired = 100%. Inter-observer reliability for the use of graphical emoticons was 96%, and 99% for text-based emoticons.

Sexual Identity

The coders also scored for references to romantic relationships with other boys or girls (e.g., a boyfriend/girlfriend, or a “crush”), for which inter-observer reliability was 95%, and if the blog author intimated in their language a heterosexual or homosexual preference, for which inter-observer reliability was 100%.

Overview of DICTION

In order to evaluate the language used on blogs, we used DICTION, a content analysis software program that takes into account language context as well as word frequencies. DICTION 5.0 analyzes texts and creates numerical frequencies based on word ratios. To do so, DICTION counts the words of a text and compares them with 33 built-in dictionaries to create a standard set of scores based on those lists of words. For instance, any reference to “I,” “I’m,” “me,” or “my” in the text, an obvious indicator of identity, falls under the “self-reference” dictionary and scores are generated for the number of instances.

Thirty-five (35) individual language scores also make up formulas for four master variables, which measure the major tone of the text, and five calculated variables, which are based on word ratios, including adjective to verb ratios, and variety and complexity of the total words used. The master variables used in this study were *activity*, *certainty*, and *commonality*. The language scores were determined using DICTION's built-in dictionaries, which were word lists generated from an analysis of thousands of texts, including political, academic, journalistic, business, religious, and creative texts (Hart, 2000). DICTION also uses “Normative Profiles,” in this case “Online Discussion Groups,” to focus its analysis on particular discourses such as online communication.

We used DICTION to examine: 1) the overall tone and theme of blogs using DICTION's master variables; and 2) individual language scores, i.e., subscores, which are used to calculate the master variables. Each HTML file was converted to a text document (.TXT) for analysis by DICTION. Conversion to a text document did not remove any of the words, only the user interface and graphics in each blog. DICTION did not count text-based emotions in its analysis of semantic themes.

Resolute and Active Language

Resolute and active language scores were derived from DICTION's master variables, and were formulated using individual language scores. The master variables used to analyze resolute and active language were *certainty* and *activity*. The master variable of certainty is based on language that indicates “resoluteness, inflexibility and completeness, and a tendency to speak *ex cathedra*”⁷ (Hart, 2000). The master variable of activity is based on language that features “movement, change, the implementation of ideas and the avoidance of inertia” (Hart, 2000). The subcomponents of each master variable and the formula used to create the master variable scores are presented in the Appendix.

Communal Language Patterns

Cooperative and accommodating language scores are derived from DICTION's master variables, which are also formulated using individual language scores. The master variable used to analyze cooperative and accommodating language is *commonality*. This master variable is based on language that highlights “agreed upon values of a group,” rejecting “idiosyncratic modes of engagement” (Hart, 2000). The subcomponents of this master variable, and the formula used to create the master variable, are presented in the Appendix.

Results

The results cover aspects of online identity and language use for the total sample, as well as for males and females separately. Specifically, the results entail the disclosure of personally identifiable information, emotive features, sexual identity, common blog themes, and masculine and feminine language. For each dependent measure, descriptive information about weblogs is initially presented. Then comparisons of male and female blogs are made. Statistical analyses were conducted using chi-square analysis and independent t-tests.

Disclosure of Personal Information

As depicted in [Figure 1](#), teen bloggers reveal a considerable amount of personal information about themselves. The most disclosed personal information includes first name (70%), age (67%), and contact information (61%) in the form of email, an instant messenger user name, or a link to a personal web page. Less disclosed information includes a birth date (39%) or full name (20%). Accounting for the different types of contact information, an email address (44%) or instant messenger user name (44%) is provided more often than a link to a home page (30%).

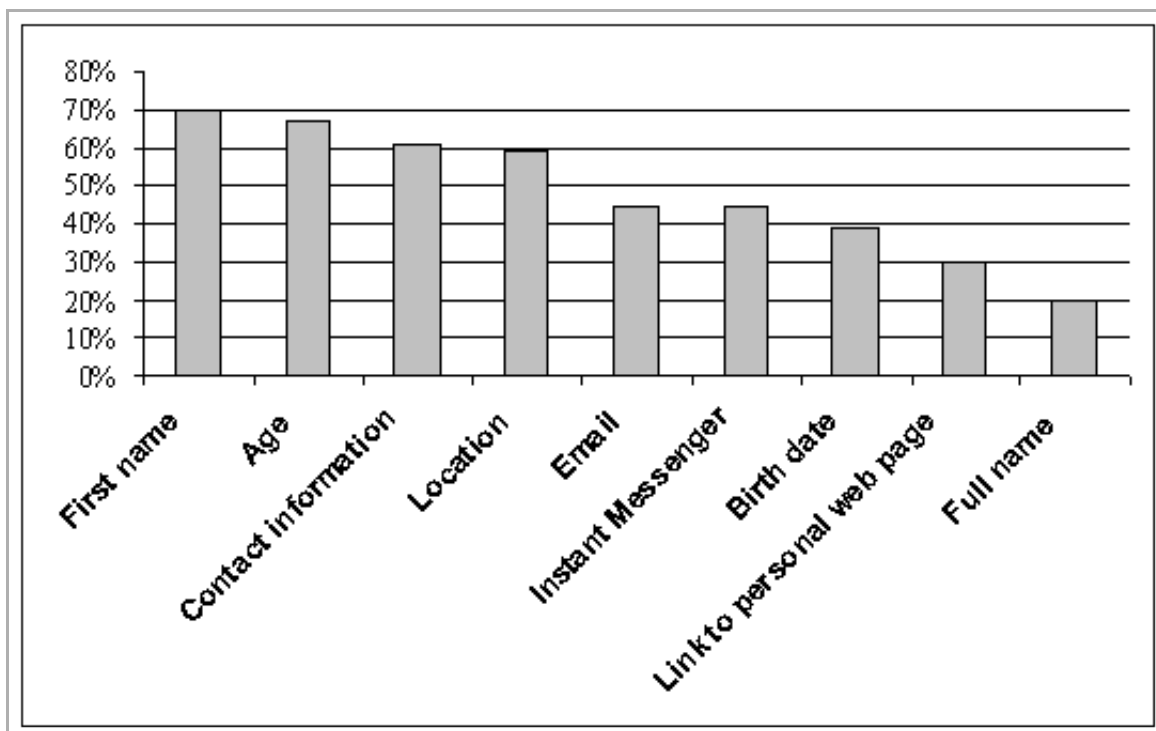


Figure 1.

[Open in figure viewer](#)

Distribution of disclosed personal information

The first hypothesis was that males more so than females would provide personally identifiable information, such as their name, age, location, and contact information. Contrary to prediction, there were no gender differences for the majority of categories: first name; full name; age; birth date; contact information; an email

address; or an instant messenger user name. There was a trend for males to provide their location more often than females, $c2(1) = 2.89$, $p < .07$. however, females were more likely than males to provide a link to a personal web page, $c2(1) = 5.51$, $p < .02$. see [Table 2](#).

Table 2. Contingency table for disclosure of personal information among teen bloggers

DOES AUTHOR REVEAL...?	TOTAL (n=70)	MALE (n=35)	FEMALE (n=3)
First name	70% (49)	36% (25)	34% (24)
Full name	20% (14)	13% (9)	7% (5)
Age	67% (47)	37% (26)	30% (21)
Birth date	39% (27)	21% (15)	18% (12)
Location	59% (41)	34% (24)	25% (17)
Contact information	61% (43)	27% (19)	34% (24)
Email	44% (31)	19% (13)	25% (18)
Instant Messenger name	44% (31)	24% (16)	20% (14)
Home page URL	30% (21)	9% (6)	21% (15)

Emotive Features

The descriptive analysis revealed that more than half (63%) of the total population of bloggers use emoticons in their blogs, whether in the form of a graphic, such as J, or a text-based smiley, such as;-). Emoticons used in the blogs were overwhelmingly Happy (53%). Sad (30%) emoticons were also very popular. Angry (4%), Flirty (5%), or Tired (5%) emoticons were used infrequently. [Figure 2](#) portrays the percentage of emoticon types within the total sample.

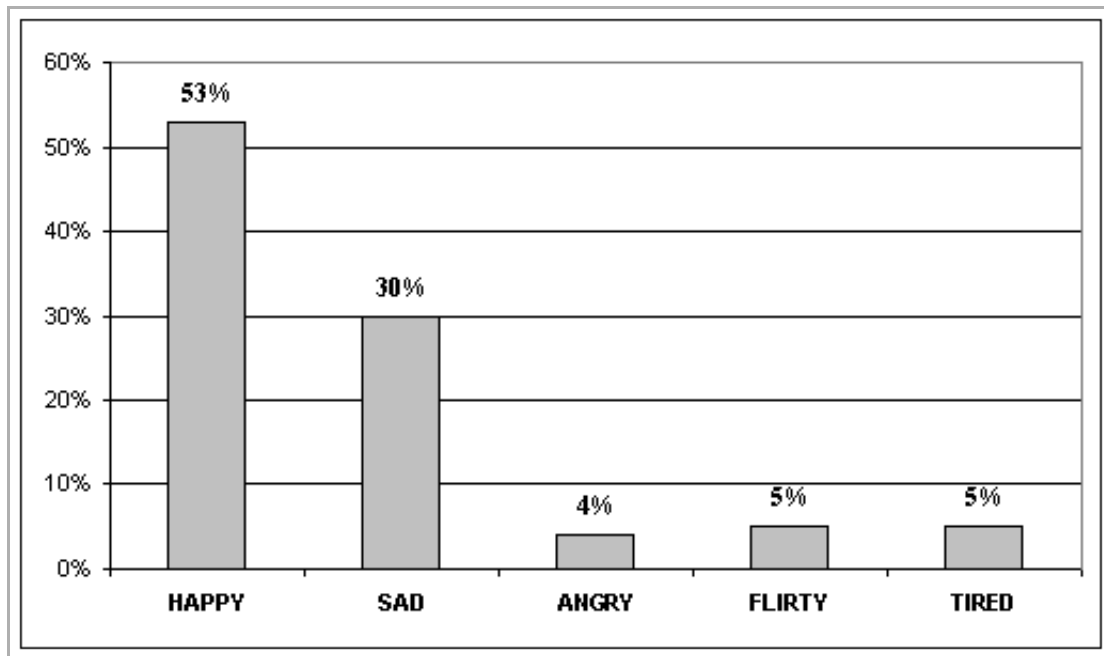


Figure 2.

[Open in figure viewer](#)

Overall emoticon use

The emoticons used in the blogs represent both graphical icons and text-based smileys. The division between these two types is very close: 49% of the total emoticon count is graphical; 51% is text-based. [Figure 3](#) demonstrates how graphics and text were used within each emoticon choice: Happy; Sad; Angry; Flirty; or Tired. Happy emoticons tipped toward text-based over graphical. Sad emoticons were almost evenly split between graphical and text-based forms. Angry and Tired emoticons were overwhelmingly graphical, mainly because replicating these emotions are difficult in text form. By contrast, Flirty emoticons (often in the form of a;) or a :P which represent a “wink” or a “tongue sticks out”) were mostly text-based.

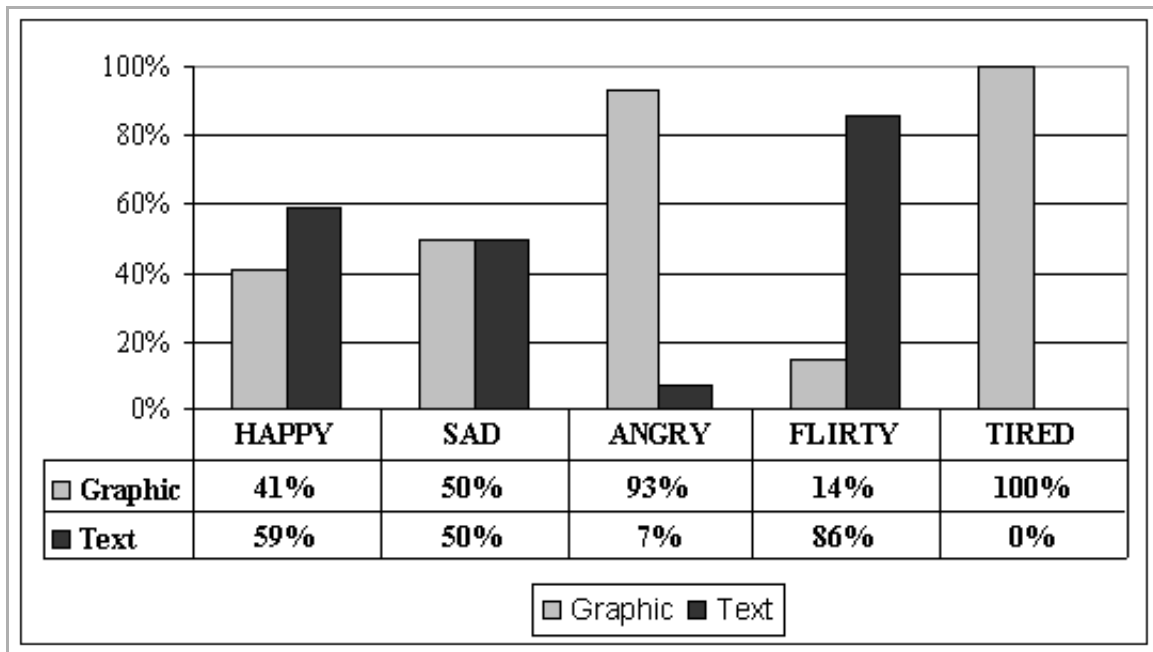


Figure 3.

[Open in figure viewer](#)

Percentage of graphic vs. text emoticons

Because females are traditionally more emotionally expressive than are males, we expected females to use emoticons in their blogs more often than males. Contrary to prediction, there were no overall gender differences for how often emoticons were used. Surprisingly, of those who used emoticons in their blogs, there was a trend for males to use more emoticons than females ($= 8.6$ vs. $= 3.9$, respectively), $t(68) = 1.68$, $p < .10$. Males used more flirty emoticons in their blogs than females ($= 0.5$ vs. $= 0.1$, respectively), where $t(68) = 2.13$, $p < .04$. There was also a trend for males to use more sad emoticons than females ($= 2.4$ vs. $= 0.9$, respectively), $t(68) = 1.74$, $p < .09$.

Relationships and Sexual Identity

Mentioning relationships with another boy or girl occurred in almost half (49%) of the blogs. Contrary to prediction, rather than favoring females, discussions regarding relationships were split evenly between male and female bloggers. Most of this discussion revolved around boyfriends, girlfriends, or other people they “like” or have a “crush on.” While 17% of authors in this sample discuss being homosexual, these authors were overwhelmingly male (83%), $\chi^2(1) = 6.44$, $p < .01$, with only two instances of females discussing being a lesbian or a transgendered person (see [Table 3](#)).

Table 3. Contingency table of sexual identity among male and female bloggers

	TOTAL (n=70)	MALE (n=35)	FEMALE (n=35)
Relationships	49% (34)	25% (17)	25% (17)
Homosexuality	17% (12)	14% (10)	3% (2)

Resolute and Active Language Patterns

In keeping with the male gender role, we expected males more so than females to use language that reflects resoluteness, aggression, activity, and a sureness of the self, while we expected females to use more passive language patterns. As seen in [Table 4](#), males demonstrated more sureness, i.e., resoluteness, of self than did females as reflected in the master variable, certainty, $t(66) = 2.16, p < .03$.

Table 4. Contingency table of sexual identity among male and female bloggers

	Mean Score	Mean Score
	MALE	FEMALE
Certainty	42.44	40.33
Tenacity	32.31	31.65
Leveling	8.56	7.89
Collectivity	4.46	3.18
Insistence	43.90	37.57
Numerical Terms	32.99	33.09
Ambivalence	17.49	19.62
Self-reference	35.43	38.84
Variety	0.53	.55

As [Table 5](#) highlights, there was also trend for males to score higher than females (= 51.04 vs. = 50.06, respectively) for the master variable, activity, $t(66) = 1.78, p < .08$. within the activity dimension, males tended to display more motion in their language than did females (= 4.94 vs. = 3.67, respectively), $t(66) = 1.94, p < .06$. contrary to prediction, however, there were no gender differences for the sub-score of aggression favoring males, nor were there differences favoring females for the sub-score of passivity.

Table 5. Mean language scores for activity, aggression, and passivity for male and female bloggers

	Mean Score	Mean Score
	MALE	FEMALE
Activity	51.04	50.06
Aggression	3.03	2.14
Accomplishment	5.14	4.60
Communication	8.83	9.38
Motion	4.94	3.67
Cognitive Terms	7.82	8.68
Passivity	3.96	3.83
Embellishment	43.90	37.57

Communal Language Patterns

In keeping with traditional gender patterns, we expected that females would use language that is more cooperative and accommodating than males. Contrary to prediction, there were no gender differences for the master variable, commonality (males $M = 49.06$ vs. females $M = 49.38$), or the sub score of cooperation.

Discussion

The purpose of this study was to examine online identity construction, focusing on how teenagers present and express themselves using weblogs. Online identity was explored in several ways: 1) the disclosure of real-world personal information; 2) emotive features used to express thoughts and feelings; 3) the language that adolescents use to express their sexual identity; and 4) the tone and semantics of the blogs, focusing on resolute, active, and communal language styles and the subcomponents of passive, aggressive, and cooperative language features. By examining language and emoticon usage, we were able to consider both the roles that adolescents assumed for themselves and in relation to one another, which may be more or less stable over time, as well as aspects of identity that may be more stable over time, such as sexual orientation.

Disclosure of Personal information

Anonymity and flexibility are inherent in the Internet arena. Freedom from physical constraints, as well as the ability to design one's persona, creates an expectation that users would experiment with online identity (Turkle, 1995). Exploring identity can take place on a variety of levels: in the name or character an online user assumes; in the types of information that are revealed online; and in the relationship of this information between the virtual and the nonvirtual world (Calvert, 2002; Döring, 2002; Jacobson, 1999; Turkle, 1995). In short, people can present themselves in a realistic way or a fictitious way. In our study, youth chose to present themselves realistically.

In some CMC contexts, such as multi-user domains, anonymity and flexibility are common experiences (Calvert, 2002; Turkle, 1995). This is not the case, however, with teenage blogs. Blogs provide a space for self-expression, usually in the form of long, personal, and thoughtful entries (Herring et al., 2004a). The online presentations of teenagers demonstrate that blogs are an extension of the real world, rather than a place where people like to pretend. For instance, teenagers reveal a considerable amount of personal information in their blogs. This includes first and sometimes last names, age, and location. The majority of teenage bloggers discuss things that influence or impact their real world, such as relationships, and attach their real-world identity to these discussions. In the case of location, males provide a little more personal information than females; but in all other cases, both males and females are revealing the same kinds of information.

Teenagers also provide diverse ways for others to contact them online. This includes an email address, an instant messenger user name, or even a link to a personal home page (which may contain other forms of contact information). While males and females provide the same types of contact information, females make a link to personal web site available more often than do males. This is interesting in light of previous studies that suggest females may be hesitant or fearful of male dominance, sexual harassment, or cyberstalking, which includes stalking, threatening, or harassing another person in an online environment (Earley, 1999; Gilbert, 1995; Herring, 1993).

That teenagers are providing so much personally identifying information about themselves is a cause for concern. Sexual predators can pose a serious threat to minors who are online (Thornburgh & Lin, 2002). "Don't talk to strangers" is a parental idiom for teaching awareness of the dangers of pedophilia and sexual predation to their children. However, on the Internet, where virtuality provides a sense of freedom from physical harm, people may feel less afraid of the online stranger. At the same time, the Internet's anonymity makes it an attractive medium for sexual predators and cyberstalkers, as sexual predators can deceive online adolescents by pretending to be younger than they are (Earley, 1999).

Because teenage bloggers are revealing a considerable amount of personal information, as well as multiple ways to contact them online, the danger of cyberstalking and communicating with strangers online is a serious issue. An awareness of the dangers of revealing personal information online should be cultivated in young bloggers, for cyberstalkers can arrange to meet their victims in offline settings.

Emotive Features

Emoticons and smileys were prominent in the blogs in our sample. More than half of the total bloggers use emoticons, whether graphical or text-based. While the majority of emoticons are Happy or Sad, bloggers

sometimes use Angry, Flirty, or Tired emoticons. Graphical and text-based emoticons are used with the same frequency, except in the case of Flirty, which is dominated by text-based smileys, and the cases of Tired and Angry, which are almost completely graphical. The reason for this may be two-fold: first, it is difficult to represent Angry or Tired in a text-based form, or Flirty in a graphical form; second, while :) and :(are standard, other types of smileys may be less well-known.

As David Crystal (2001) suggests, emoticons are used to fill a void in online communication. Because online interactions cannot rely on facial and body gestures to express thoughts or feelings, emoticons are used for several reasons. First, they help to accentuate or emphasize a tone or meaning during message creation and interpretation (Crystal, 2001). Second, emoticons help establish a current mood or impression of the author (Constantin, Kalyanaraman, Stavrositu, & Wagoner, 2002b). For instance, if the author is complaining about school, and including emoticons of anger or sadness, the reader may derive more depth or feeling from the message. Third, emoticons are a creative and visually-salient way to add expression to an otherwise completely textual form (Crystal, 2001).

Some studies suggest that females use emoticons more often than males in instant messaging applications and newsgroups (Lee, 2003; Witmer & Katzman, 1997; Wolf, 2000). However, this was not the case with our blogs. Among authors who use emoticons in their blogs, males actually posted more emoticons than did females. Similarly, males used more Sad and Flirty emoticons than did females. The use of Flirty emoticons by males fits well within the literature where males more so than females actively search for online sexual interactions, sometimes referred to as “tiny sex” (Turkle, 1995). The reason for the higher levels of Sad emoticons by males is less clear, particularly given that adolescent females are more likely to report feeling sadness than are adolescent males (Ruble & Martin, 1998). The overall heavy use of emoticons in weblogs suggests that they are prevalent in online interactions, partly because emoticons are now often built into CMC applications such as instant messaging, chat rooms, message boards, and now blogs.

Sexual Identity

Half of all bloggers discuss sexual identity and love relationships, boyfriends/girlfriends, or “crushes.” Discourse of this type dovetails nicely with the kinds of developmental issues, e.g., establishing a mature sexual identity (Erikson, 1993), that adolescents are experiencing in real life.

The gender of those 17% of bloggers who discuss a homosexual identity is overwhelmingly male. This suggests that male bloggers may be using blogs as a safe and comfortable environment to be honest, or even candid, about their sexual identity and feelings. Unlike a private diary, discussing “coming out” online can be empowering, as the blog author is aware that his posts can be read in the public sphere. While being “out” has been historically uncharacteristic of many homosexual males in the nonvirtual world, given that past discourse about homosexuality was marked with negative connotations and even punishment (Foucault, 1990), increased acceptance of homosexuality by certain subgroups, particularly youth, may make the disclosure of sexual orientation less of a risk than it has been in the past. Disclosing one's sexual identity online may also provide a way for gay teens to find others who share their sexual identity. Taken together, our findings suggest that adolescents seek a continuity of representations of who they are, as well as a confirmation of those representations by their peers.

Active, Resolute, and Passive Language Use

Lakoff's (1975) theories on gender and language suggest that females use language that is passive, indicating shyness and lower self-confidence (Eckert & McConnell-Ginet, 2003). In other CMC contexts, studies have found that females use language that is polite, appreciative, and cooperative (Herring, 2000, 2001; Savicki, 1996). By contrast, males use language that is more confident (Arnold & Miller, 1999), more aggressive, and less concerned with politeness (Herring, 2000, 2001).

In our study of blogs, males did use language that was more active, inflexible, and resolute, which resonates with some of Susan Herring's findings for other modes of CMC (Herring, 1993, 2000, 2001). Females, however, did not use more passive, cooperative, or accommodating language as Lakoff's work would suggest. One possible implication is that the language and the social interactions on the Internet are changing, perhaps because the participants are changing. That is, the latest wave of teenage females, at least female *bloggers*, may have different gender roles from those of earlier generations that Lakoff observed. Alternately, females who choose to create blogs may be less traditional in their gender roles than the general population.

In sum, the language of blogs found in our study echoes Michelle Rodino's (1997) claim that comparing language on the Internet to traditional gender and language studies is often oversimplified because online interactions offer more freedom and flexibility. While males may use more active and resolute language in their blogs than females do, females do not necessarily use passive or accommodating language. The latter findings support the concept of androgyny where masculinity and femininity are two separate dimensions rather than being polar opposites on a single continuum (Huston, 1983).

Conclusion

Identity is an essential part of the human experience. For adolescents, especially, identity marks a developmental milestone (Calvert, 2002; Erikson, 1993). In previous work, identity has been approached in terms of the relationships between the internal experience, such as personality and self-definition, and the external world, such as social relationships and shared values (Erikson, 1993; Freud, 1989; Jung, 1976; Lacan, 1986). The Internet has provided a new context for identity exploration, as the virtual world provides a venue to explore a complex set of relationships that is flexible and potentially anonymous. Language on the Internet represents a new type of discourse that is shaped by the creativity and innovation of its communities of users (Crystal, 2001). This emerging discourse can then be used to express the identities of its adolescent users.

Weblogs represent a CMC environment where both identity and language play important roles. Not only are teenagers using weblogs to present an online identity, but also to express their ideas, experiences, and feelings using an adapted language. In some cases, these blogs interlink to form online communities, similar to the peer relationships observed in the offline world.

While this study supports the social interactionist perspective, in as much as language is being used to construct and reflect adolescent identity (Harter, 1998), multiple “public” faces were not the norm. Instead,

our data suggest a tendency for adolescents to use language to create an anchor and a consistent public face as they engage in the very serious business of constructing a stable cohesive set of representations of who they are (Erikson, 1993; Grotevant, 1998). This interpretation of the data is supported by adolescent bloggers' choices to reveal personally identifying information about themselves, including their names and their sexual orientations. For example, gay males used blogs to discuss their sexual identity or to come out, and most adolescents discussed boyfriends or girlfriends, concerns that are central to their developing sexual identity.

Interestingly, the blogs created by young males and females are more alike than different. Perhaps the technical ease of use of blogs levels the differences between males and females, or perhaps this generation of Internet users is becoming more androgynous in its online communication and interaction. For instance, the trend that males are averaging more emoticons than females contradicts early literature on emoticon use in instant messaging applications and newsgroups (Lee, 2003; Witmer & Katzman, 1997; Wolf, 2000), but supports recent findings favoring preadolescent males' use of emoticons in a MUD (Calvert, Mahler, Zehnder, Jenkins, & Lee, 2003). Similarly, females are not using language that is more passive, accommodating, or cooperative as depicted in earlier studies (Eckert & McConnell-Ginet, 2003; Herring, 2000; Savicki, 1996).

Overall, blogs deserve further exploration by scholars. They are important for technologists, teachers, parents, and researchers who are interested in computer-mediated communication, online communities, children and technology, and adolescent development.

Notes

- 1 We thank the National Science Foundation for a grant to the Children's Digital Media Center (Grant # 0126014) that made this research possible. We also thank Gabrielle Strouse and other members of the Children's Digital Media Center for their assistance with this research project.
- 2 <http://www.blogcensus.net/weblog/>
- 3 <http://www.livejournal.com>
- 4 <http://www.blogsearchengine.com>
- 5 <http://blo.gs>
- 6 For example, <http://www.angelfire.com/hi/hahakiam/emoticon.html>.
- 7 From Dictionary.com: "Ex Cathedra—With the authority derived from one's office or position: the pope speaking ex cathedra; ex cathedra determinations."

Appendix

Certainty: Subcomponents and Calculation

Certainty is calculated as: [Tenacity + Leveling Terms + Collectivity + Insistence] - [Numerical Terms + Ambivalence + Self-Reference + Variety]. Each of these language scores was defined as follows:

- 1 *Tenacity*: Words that express confidence and totality.
- 2 *Leveling*: Words that express completeness and assurance.
- 3 *Collectivity*: Words that express social groupings such as a crowd or a world.
- 4 *Insistence*: This score assumes that repetition of words suggests a “preference for a limited, ordered world” (Hart, 2000). The calculation uses this formula:

Heavily Used Words x Total Occurrences

10

- 5 *Numerical Terms*: Any date or number that deals with quantitative or numerical operations.
- 6 *Ambivalence*: Words that express hesitation or uncertainty. This includes hedge phrases, vagueness, or confusion.
- 7 *Self-Reference*: Words that express first-person references, such as I, I'd, I'll, me, my, mine.
- 8 *Variety*: This score divides the number of different words in a text by the total words of a passage, assuming that a high score “indicates a speaker's avoidance of overstatement and a preference for precise statements” (Hart, 2000). The formula is:

Number of Different Words

Number of Total Words

Activity: Subcomponents and Calculation

Activity is calculated as: [Aggression + Accomplishment + Communication + Motion] - [Cognitive Terms + Passivity + Embellishment]. Each of these language scores is defined as follows:

- 1 *Aggression*: Words that express competition or forceful action, including terms that imply physical energy or domination.
- 2 *Accomplishment*: Words that express the completion of a task, or methodical human behavior.
- 3 *Communication*: Words that express social interaction, including face-to-face or mediated modes, such as a film or telephone.
- 4 *Motion*: Words that express movement, speed, journeys, transit, or physical processes.
- 5 *Cognitive Terms*: Words that express “cerebral processes”, including discovery, psychology, logic,

mental challenges, or learning practices.

6 *Passivity*: Words that express inactivity, compliance, or docility.

7 *Embellishment*: This score calculates the ratio of adjectives to verbs, and assumes that “heavy modification slows down a verbal passage, de-emphasizing human and material action” (Hart, 2000). The calculation uses this formula:

$$\text{Praise} + \text{Blame} + 1$$

$$\text{Present Concern} + \text{Past Concern} + 1$$

Commonality: Subcomponents and Calculation

Commonality is calculated as: [Centrality + Cooperation + Rapport] - [Diversity + Exclusion + Liberation].

Each of these language scores is defined as follows:

- 1 *Centrality*: Words that express regularity, congruence, predictability, universality, or an agreement on central values.
- 2 *Cooperation*: Words that express formal, informal, and intimate associations and interactions.
- 3 *Rapport*: Words that express an affinity toward similarities among a group of people
- 4 *Diversity*: Words that express non-conformity or heterogeneity.
- 5 *Exclusion*: Words that express social isolation.
- 6 *Liberation*: Words that express a rejection of social standards.

References

Wiley Online Library

[Browse Publications](#) [Browse by Subject](#) [Resources](#) [Help](#)

[About Us](#) | [Advertisers](#) | [Agents](#) | [Contact Us](#) | [Cookies](#)

[Media](#) | [Privacy](#) | [Site Map](#) | [Terms & Conditions](#)

WILEY

[About Wiley](#)

[Wiley.com](#)

[Wiley Job Network](#)

Copyright © 1999-2014 John Wiley & Sons, Inc. All Rights Reserved.