Developing and Testing Automatic Models of Patient Communicative Health Literacy Using Linguistic Features

Scott A. Crossley[1], **Renu Baylan**[2], **Jennifer Liu**[3], **Dean Schillinger**[4], **Andrew J Karter**[3], **Danielle McNamara**[2]

[1]Georgia State University

[2]Arizona State University

[3]Kaiser Permanente Northern California

[4]University of California, San Francisco

Introduction

Health literacy (HL) is generally defined as a patient's ability to obtain, process, comprehend and communicate basic health information so as to make appropriate health decisions (Grossman, 2010; Institute of Medicine, 2004; Schillinger et al., 2017). HL has important practical implications for a healthy society because patients with limited HL have been shown to have worse health and to be subject to a higher risk of adverse health outcomes over time (Sudore, 2006). The negative consequence of limited HL have been most well-studied in the case of type 2 diabetes mellitus (DM2) (Bailey et al., 2014), where limited HL has been shown to be independently associated with poor patient knowledge of diabetes self-management, worse glycemic control (i.e., blood sugar control,), higher rates of severe hypoglycemia (dangerously low blood sugar levels, a treatment side effect) (Sarkar et al., 2010), poorer medication adherence (Karter et al., 2009), and higher rates of diseases complications, such as amputations and kidney failure (Schillinger et al., 2002).

With respect to assessing HL, nearly all established measures require patients to read health-related stimuli and respond to comprehension questions; a few require patients to read health-related stimuli and pronounce the words correctly (Haun, Valerio, McCormack, Sorensen, Paasche-Orlow, 2014). Others have suggested that HL is a complex construct that must be deconstructed into component skills of "functional HL", "communicative/interactive HL", and "critical HL" (Nutbeam, 2000). While HL is a complex construct and a topic of significant debate, studies have shown that existing measures of HL are strongly correlated with general linguistic and literacy skills (Nutbeam, 2009). To our knowledge, however, no studies have attempted to examine patients' written communications as a means to evaluate communicative HL, despite the face validity of such an approach. Furthermore, existing measures of HL are time-consuming, require in-person administration and challenging to scale. This study involves a large patient population experiencing DM2, a disease which is estimated to inflict 30.3 million people in the U.S. (~9% of adults) (Centers for Disease Control and Prevention 2014). Like many chronic conditions, DM2 self-management is complex and requires frequent communication with healthcare providers. It is argued that poor communication skills on the part of patients and physicians explains some of these

health disparities (Bailey et al., 2014). In addition, lower HL can impact patients' evaluation of online health information (Diviani et al., 2015) and negatively impact the effective use of online patient portals (Sarkar et al., 2010), which are digital tools that provides an avenue for e-mail exchanges ("secure messages", or SMs) between patients and their physicians. The overarching objective purpose of this study is to examine "communicative HL" in terms of patients' abilities to effectively convey their own medical information through on-line writing (i.e., communicate basic health information). Examination of patients' SMs can then be harnessed via natural language processing techniques to develop an automated measure of HL that could be applied to entire populations of patients.[1]

The specific goals of this study are two-fold. First, we aim to identify the linguistic features in patients' written data derived from SMs that can successfully categorize low vs. high HL patients in terms of their ability to effectively convey health information. Our purpose is to instantiate HL as a function of language skills that afford communication of basic health information. We do so by collecting expert judgments of written HL among SMs sent by patients to their physicians in an online patient portal. We then use these human judgments as gold standards of patient HL and employ linguistic features found in the patients' SMs and reported by natural language processing (NLP) tools to develop a machine learning (ML) model to categorize patients as being either low or high HL. Second, we assess links between the classifications made by the ML model and patient demographic information, health outcomes, and healthcare utilization rates to provide secondary, or predictive validation.

Such analyses can afford a better understanding of potential links between language production and communicative HL by developing models of communicative HL based on language production. In this study, we examine the extent to which these models of communicative HL predict patient demographic medical profiles, and explore whether the patterns that merge are similar to those observed in prior, smaller studies that measured HL directly on patients. Since lower HL is associated with preventable suffering, more rapid decline in physical function (Smith et al., 2015) and related healthcare costs (Bailey et al., 2014), linguistic models that can automatically predict patients with lower communicative HL could be used to systematically identify at risk patients. Early identification of these patients could allow health care providers to tailor and individualize provider communication and/or enable and health systems to deliver related interventions at a population level, providing greater support to those patients with the greatest communication needs. As such, developing automated models of HL has the potential to significantly impact patient care and public health.

Health Literacy

In the healthcare context, HL represents a set of skills that are developed through effective proactive, receptive and interactive communication behaviors (Sudore, 2009) between patients and clinicians across a number of inter-related domains, including literacy skills (reading and writing); verbal skills (speaking and listening); numeracy skills (understanding risks, probabilities, calculations) and, increasingly, digital communication skills (interacting with computer-based platforms) as they relate to health and healthcare topics. In fact, a

growing body of research has shown that poor communication exchange is an important mediator in the relationship between limited HL and health outcomes, and that improving communication exchange can mitigate HL-related health disparities (Schillinger et al., 2004, 2009).

In the US, limited HL is more common among socioeconomically disadvantaged groups (e.g., lower income, those with lower educational attainment), some minorities, older populations, and immigrants who speak English as a second language (Institute of Medicine, 2004; Kirsch, Jungeblut, Jenkins, & Kolstad, 2002; Smith et al. 2015). As stated above limited HL has been shown across a number of populations and medical conditions to be independently associated with higher prevalence of DM2, poor glycemic control (Schillinger et al., 2002), severe hypoglycemia (Sarkar et al., 2010), and poor medication adherence (Karter et al., 2009). Hence, limited HL represents a critical and costly clinical and public health problem (Bailey et al., 2014; Institute of Medicine, 2004;).

Suboptimal communication exchanges, and their resultant problems with treatment adherence, are a mediator between limited HL and DM2 outcomes. Communication exchanges center on patients' ability to both communicate health-related problems to their health care providers and understand health-related content. For instance, DM2 patients with limited HL are significantly more likely to report difficulties understanding medical jargon (Castro, Wilson, Wang, & Schillinger, 2007) and coming to a shared understanding with their physician about clinical problems and treatment options (Schillinger, Bindman, Wang, Stewart, & Piette, 2004). Synthesis studies on HL and DM2 demonstrate that patients with limited HL have poorer communications with providers because of a reduced ability to communicate health-related ideas (Bailey et al., 2014). In addition, limited DM2 HL patients, compared with high HL patients, more often report that they could better control their DM2 if communication between them and their providers were improved (Sarkar et al., 2008). Overall, patients with lower HL more often report problems with shared decision-making, more difficulties understanding their health problems and how to manage them, and being confused about their medical care (). Further, they report that few physicians appreciate the communication barriers they have when attempting to follow medical recommendations (Schillinger et al., 2004).

One solution to these problems is to increase the opportunity for communications between patients and physicians to occur electronically. Prior research has shown that communication technologies and digital platforms, when developed for and in collaboration with populations with limited HL, are not only accessible to patients with limited HL, but can differentially support those with the greatest communication needs and improve health outcomes (Schillinger, 2007; Schillinger et al., 2008). In nearly all health systems in the US, increased communication opportunities are generally provided through online patient portals, which are digital communication tools linked to a patient's electronic medical record. These portals are being heavily promoted by health systems, due in part to federal incentives. Patient uptake is growing and in some advanced systems uptake is high (Schillinger et al., 2017). Research demonstrates that patients who access them are more likely to have favorable healthcare utilization patterns, adhere to prescribed regimens and achieve better outcomes (Zhou, Kanter, Wang, & Garrido, 2010). Indeed, as healthcare

becomes increasingly dependent on asynchronous, between-visit electronic communications via SMs, patient portals are gaining primacy as a means of communication. This means that 21st century HL will require patients to have a certain degree of digital communicative HL skills to take advantage of online patient portals. However, patients with limited HL may have difficulty messaging their provider or understanding the provider's replies or instructions (Sarkar et al., 2010). Providers too must engage with patients in a manner that provides meaningful and actionable information and support in an easily comprehended style that promotes shared meaning (Brach, Dreyer, & Schillinger, 2014). Thus, the skillset inherent to patient HL in the 21st century includes the ability to both effectively write and process SMs. SMs are particularly relevant for patients with an illness as complex as DM2, given their need for relatively frequent communication as compared to other types of patients.

Examining the extent to which health systems and their clinicians accommodate to the communication needs of DM2 patients with limited HL is an important area for research and policy. HL limitations pose a barrier to patient-provider communication, undermine healthcare delivery, jeopardize health outcomes and increase healthcare costs. Hence, the ability to assess patients' HL has long been of interest to individual clinicians, healthcare delivery systems, and the public health community. Healthcare delivery systems increasingly recognize the importance of identifying the subset of patients who have limited HL. However, measuring HL to tailor interventions has proven painstaking and infeasible to scale (DeWalt et al., 2012). As a result, health systems are interested in incorporating predictive models and derived scores as a means of risk stratifying and targeting care. Using "big data" along with NLP approaches to estimate HL at the individual patient level could open up new avenues to enhance population management as well as individualize care. Failure to do so in population management interventions has previously been shown to amplify HL-related disparities (Karter et al., 2015).

NLP Approaches to Measuring Health Literacy

A recent review of all HL research measures found that at least 51 unique measures have been created and employed, with virtually all requiring paper and pencil responses, and individual measures requiring up to one hour to administer. Of the 51, 26 measured general HL, 15 measured disease or content-specific HL, and 10 measured specific sub-populations (Haun et al., 2014). No studies have attempted to measure communicative HL by assessing patients' own original written content, specifically written communications to their physicians. Of note, studies in the field of general literacy have shown that linguistic production (e.g. writing skill) is correlated with linguistic comprehension (e.g. reading skill), providing a basis for harnessing patients' SMs to assess communicative HL. As such, one approach to identifying patients with limited communicative HL is to employ NLP tools to assess the linguistic production found in patient writing. NLP encompasses all computerized approaches to analyzing language, specifically using or measuring some type of linguistic or language features to better understand natural language. The strength of an NLP approach is its efficiency in analyzing massive amounts of data by repeating analyses objectively and literally, something that is time consuming and difficult for humans to accomplish. A primary goal of NLP analyses is to gather information on how language is understood and

used. NLP tools are generally not used in isolation and are instead combined with statistical methods or machine learning algorithms to increase the reliability and validity of their output.

NLP tools have long been used in the medical domain to represent clinical narratives, assess medical text quality, or develop semantic lexicons for medical language processing (Johnson, 1999). A few studies have also examined HL in terms of medical text processing. Specifically, studies have investigated the readability of medical texts and how NLP features can be used to better predict medical text readability. Much of this research is based on early studies that used classic readability formulas such as the Flesch-Kincaid Grade Level (Kincaid, Fishburne, Rogers, & Chissom, 1975) to assess medical text readability. These studies generally report that health-related texts are not written at appropriate reading levels (Weiss, 2015). In response, readability researchers have investigated the potential for NLP features to better predict medical text readability with the goal of improving text processing on the part of patients. Studies across a variety of domains have shown that readability formulas based on NLP features more successfully distinguish difficult from easy texts compared to classic readability formulas (Kim et al., 2007; Wu et al., 2013; Zeng–Treitler, Kandula, Kim, & Hill, 2012; Zheng & Yu, 2018).

At least one recent study examined the potential for NLP tools to assess patient HL (Balyan et al., 2019). In this study, Authors applied NLP tools on a large corpus of SMs sent from 6,941 patients to their clinicians to develop a machine learning model of HL based on patients' self-reported HL obtained via survey. Results indicated that the developed models revealed patterns consistent with previous HL research in that patients identified as having limited HL by the model were more likely to be older, have limited educational attainment and be of a minority background. In addition, identified limited HL patients had poorer medication adherence and glycemic control and higher rates of hypoglycemia, comorbidities and healthcare utilization. One limitation of that study is that the linguistic patterns that informed the machine learning approaches were not clearly delineated, making it difficult to pinpoint the language features of limited HL patients. Additionally, the study used as its gold standard for HL patients' perceptions of their own communicative competence, a measure of HL-related self-efficacy. While the self-reported HL items have previously been validated (Sarkar et al., 2008), it is not known whether objective measures of communicative HL (such as measuring communicative HL via assessing patients' own SMs) would perform superiorly, or whether such a measure would tap into domains of HL that are different from the HL-related self-efficacy measured by self-report.

Current Study

Despite the use of NLP techniques in medical domain tasks, in predicting symptoms, and in assessing medical text readability, little attention has been paid to assessing how NLP techniques can be used to investigate patients' HL. Thus, we build on Balyan et al. (2019) by developing a novel NLP-based model of HL. Unlike Authors, we do not rely on self-reported survey results of patients' HL-related self-efficacy, but rather rely on expert ratings of how well patients can communicate medical information in SMs. This may help overcome inaccuracies in self-reported data. We also provide greater information about how

linguistic features pattern between limited and adequate HL patients providing greater connections between HL and linguistic features. In a similar fashion to Authors, we validate the resulting NLP model of HL by assessing associations between it and patient demographics and patient-related behaviors and events such as medication adherence, medical outcomes and hospitalizations. Thus, our study is guided by the following research questions:

1. Can linguistic features model expert ratings of communicative HL based on language features found in patient SMs?

2. Do the derived models demonstrate hypothesized differences in patient demographics, health outcomes, and hospitalizations?

An automated model of HL based on communicative skills may help overcome temporal, physical, and interpersonal obstacles reported in previous research which has attempted to measure HL through patient interviews or questionnaires. Such approaches make the process of measuring HL challenging, especially when scaling for larger patient populations. An automated categorization of HL based on NLP features would provide a more efficient means to automatically detect patients with limited HL based on productive language skills that, if found to be valid, could be used to identify at-risk patients so as to deliver appropriate interventions.

Method

Participants

Our study includes data obtained from Kaiser Permanente Northern California (KPNC), a nonprofit, fully integrated healthcare delivery system. KPNC provides services to 3.3 million plan members through 37 outpatient centers staffed by ~3,300 medical providers. The data extracted from the KPNC registry is for DM2 patient for whom we have measures of self-reported health literacy (Chew et al., 2008), patient reports of provider's communication quality, and a broad array of socio-behavioral and psychological measures (Ratanawongsa et al., 2013). This sample is known as the Diabetes Study of Northern California (DISTANCE) - a study designed to study social and ethnic disparities in diabetes outcomes - and comprises data for ~14,000 patients, with certain ethnic minorities oversampled to ensure ample numbers of minority patients. The variables in DISTANCE were collected from questionnaires completed via telephone, on-line, or paper and pencil (62% response rate).

Corpus

We extracted all of the secure messages (SMs) (N=1,050,577) exchanged between 12,286 patients and all their clinicians from KPNC' patient portal between 01/01/2006 and 12/31/2015. We identified those SMs that a patient sent to his or her primary care physician(s) and aggregated these SMs into a single file to represent patients' linguistic profiles. We removed all patients whose SM lacked sufficient words (<50 words) to provide linguistic coverage. We also removed all patients who did not have matching DISTANCE survey data. We then removed all SMs written in a language other than English. This left us with 9,530 patient SM threads.

We initially purposively sampled 500 SMs so that the SMs represented patients from a range of self-reported health literacy scores, ages, socio-economic status, and other demographic variables. Upon analysis of the data, it became apparent that many of the SMs included text written by patient proxies (i.e., SMs written for the patient by their caregivers) that did not represent the patients' linguistic profile. In some cases (~5% of patients) the SMs included more than 50% proxy messages. We next resampled the data to increase the initial sample size and removed all the patients whose SMs contained over 50% proxy data. After resampling and culling, we were left with 512 aggregate patient SMs that contained less than 50% proxy data. For these 512 messages, proxy data were manually removed.

We compared the demographic characteristics among the entire SM population (n=12286), the sample used in this study (n=9530) and the subsample (512) for which we collected human ratings of communicative HL. The three groups had similar sex distribution (men%: 52 vs 53 vs 51). The study sample and subsample had higher percentage of white patients (27 vs 29 vs 34). However, the subsample had a lower percent of patients with college education (62 vs 63 vs 56). The patients in were subsample were about 1 year younger comparing to those in the other two groups (57.5 vs 57.4 vs 55.71).

Human Ratings

By definition, all written content in patients' SMs was health-related. We developed a holistic rubric to obtain communicative HL ratings for each patient SM thread in our subset. The rubric was based on a 6-point rating scale adapted from that used by the SAT for essay scoring. The rubric was used to holistically assess the perceived communicative HL of the patients based on the language they produced in their SMs. High communicative HL (i.e., a score of 6) was defined as the following:

The patient demonstrates clear and consistent mastery of written English, although the writing may contain a <u>few minor errors</u>. The patient's writing is well organized and accurately focused providing clear access to the content of the message and the ideas that the patient wants to express. The writing demonstrates clear coherence and smooth progression of ideas, exhibits skillful use of language, using a varied, accurate, and apt vocabulary. In addition, the writing demonstrates meaningful variety in sentence structure and is free of most errors in grammar, usage, and mechanics.

Low HL (i.e., a score of 1) was defined as:

The patient demonstrates very little or no mastery of written English and the writing is <u>severely flawed</u> by ONE OR MORE of the following weaknesses: disorganized or unfocused writing that results in disjointed or incoherent writing that do not provide access to the content of the messages and the ideas that the patient wants to express. The writing also displays fundamental errors in vocabulary, demonstrates severe flaws in sentence structure, and/or contains pervasive errors in grammar, usage, or mechanics that persistently interfere with meaning.

Each SM in the corpus was scored by two HL experts using the holistic rubric. Both raters had advanced degrees in language-related fields and experience rating HL data. The raters

were trained on the rubric using a separate subset of 50 SMs until inter-rater reliability reached $r > .700$. Raters were informed that the distance between each number on the rating scale was to be considered equal. After calibrating on the initial 50 texts, the raters then independently scored the 512 SMs. After scoring, they were given the opportunity to adjudicate any disagreements greater than two. Final inter-rater reliability was Kappa = .678, indicating strong agreement (Landis & Koch, 1977). After adjudication, the raters' scores were averaged to provide an overall HL score for each patient. The SMs were split into two groups based on lower HL scores (< 4, n = 200) and higher HL scores (4, n = 312). These scores functioned as our dependent variable during analyses.

Linguistic Feature Selection

Prior research that has indicated that lexical features related to word choice, discourse features, and sentence structure are strong predictors of writing quality (Crossley, Kyle, & McNamara, 2015; Crossley & McNamara, 2016; McNamara, Crossley, Roscoe, Allen, & Dai, 2015). To capture these features, we used three natural language processing (NLP) tools that derive linguistic features related to lexical sophistication, text cohesion, and syntactic complexity. We discuss these tools and their output below.

Tool for the Automatic Analysis of Lexical Sophistication (TAALES)—TAALES (Kyle & Crossley, 2015; Kyle, Crossley, & Berger, 2018) is a computational tool that is freely available and easy to use, works on most operating systems, affords batch processing of text files, and incorporates over 100s of classic and newly developed indices of lexical sophistication. These indices measure word frequency, lexical range, n-gram frequency and proportion, academic words and phrases, word information, lexical and phrasal sophistication, and age of exposure. For many indices, TAALES calculates scores for all words (AW), content words (CW), and function words (FW). TAALES also reports on a number of word information and psycholinguistic scores derived from databases such the Edinburgh Associative Thesaurus (EAT; Kiss et al., 1973), which calculates number of word associations per word and the English Lexicon Project (ELP, Balota et al., 2007), which calculates many lexical features including the number of phonological neighbors a word has (i.e., how many words sound similar to the word in question) and lexical decision response times for words (i.e., how long does it take to decide a word is a word versus a non-word).

The Tool for the Automatic Analysis of Cohesion (TAACO)—TAACO (Crossley, Kyle, & McNamara, 2016; Crossley, Kyle, & Dascalu, in press) incorporates a number of classic and recently developed indices related to text cohesion. TAACO has features for content and function words and provides linguistic counts for both sentence and paragraph markers of cohesion. The tool incorporates WordNet synonym sets, latent semantic analysis, and Word2Vec features. Specifically, TAACO calculates sentence and paragraph overlap indices (i.e., local and global cohesion) and a variety of connective indices. For example, argument overlap is a count of arguments that are shared between sentences and paragraphs.

The Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC)—TAASSC (Kyle, 2015; Kyle & Crossley, 2018) measures large and fined grained clausal and phrasal indices of syntactic complexity and usage-based

frequency/contingency indices of syntactic sophistication. TAASSC includes a number of pre-developed fine-grained indices that measure clausal complexity and phrasal complexity using output reported by the Stanford CoreNLP (Manning et al., 2014). At the clausal level, TAASSC measure features such as the number of passive auxillary verbs and adjective complements per clause. At the phrasal level, TAASSC calculates features like determiners per nominal phrase and dependents per nominal subject. In addition, TAASCC reports on features related to verb argument constructions (VACs) including the frequency of VACs and the attested lemmas per VAC as found in reference corpora taken from sections (e.g., magazine or newspaper) of the Corpus of Contemporary American English (COCA; Davis, 2009).

Health Outcome Variables

Because of the known association between limited HL and sub-optimal patient-provider communication (Castro et al., 2007; Sarkar et al., 2008; Schillinger et al., 2003, 2004), we examined the relationship between communicative HL and patients' reports of physician communication using an adapted version of the most HL-relevant item from the 4-item Consumer Assessment of Provider and Systems (CAHPS) survey obtained from patients as part of the DISTANCE questionnaire: "In the last one year, how often have your physician and health care providers explained things in a way that you could understand?" We defined communication as "poor" if the patient reported that his or her doctor and health care team "never" or "sometimes" explained things in a way that he/she could understand (Ratanawongsa, 2013). Using data derived from the electronic health records from KPNC, we examined Hemoglobin A1c (HbA1c, an integrated measure of blood sugar control and the Charlson index (a measure of comorbidity and illness severity; Charlson, Szatrowski, Peterson, & Gold, 1994). We report HbA1c both as a mean value, as well as the proportion of poor glycemic control, defined as a HbA1c value >9%. Comorbid illness was measured with the Deyo version of the Charlson comorbidity index (Deyo, Cherkin, & Ciol, 1992), a validated measure of illness severity. Another set of analyses was conducted for health service utilization, using outpatient clinic visits, emergency room encounters and hospitalizations. All health-related outcome variables that represented single data points (HbA1c) represented the most recent value obtained just prior to the first SM sent by the patient. All health-related outcome variables that represented annual rates (comorbidity, utilization) reflected events accrued over the 12-month period prior to the first SM sent by the patient.

Statistical Analysis

To examine the predictive strength of the linguistic features to distinguish between SMs scored as low and high HL, we conducted a discriminant function analysis (DFA). Prior to all analyses, we reduced the number of features reported by the NLP tools in two ways. We first conducted a MANOVA with pairwise comparisons to examine if there were meaningful and significant differences between the secure messages judged to be written by patients with either high or low HL skills. To control for Type 1 errors, variables had to report a p value < .001. To ensure that variables entered into the model demonstrated a significant and meaningful relation with the dependent variable, we required a partial eta-squared (η_p^2) > .02. The remaining variables were then checked for multicollinearity with one another.

Multicollinearity was operationalized as any two variables demonstrating a strong correlation ($r > .700$). If two or more variables demonstrated multi-collinearity, the variable with the stronger η_p^2 was retained.

The remaining variables (n = 35) were entered as predictors in a stepwise DFA to predict low and high HL secure messages. The initial model was conducted on the full set of secure messages (N = 512). The model reported by this DFA was then used to predict group membership of the secure messages using leave-one-out-cross-validation (LOOCV). The LOOCV procedure allows testing of the accuracy of the model on an independent data set.

The DFA model was then extended to the entire patient SM dataset (n = 9,530) comprised of SMs that fit our initial requirements for inclusion (i.e., over 50 words, matching DISTANCE survey data, and SMs written in English). Each SM in the larger dataset was classified as being written by a low or high HL patient. We then examined bivariate associations between classifications made by the model with demographic (age, race, and education), health outcome and healthcare utilization variables using a two-sided p-value at the 0.05 level of significance. Categorical variables such as race were analyzed using chi-square analysis. Mean comparisons were conducted using t-tests for HbA1c, communication (CAHPS) score, Charlson (comorbidity) index, and healthcare utilization rates.

Results

Discriminant Function Analysis

The stepwise DFA retained nine linguistic indices as significant predictors of whether a SM was rated to be from low or high HL patient (see Table 1 for descriptive and MANOVA details for each selected variable).

The results demonstrate that the DFA using these nine indices correctly classified 411 of the 512 SMs as low or high HL, $\chi^2 (1)$ n=512) = 175.829, $p < .001$, C-statistic =0.79, for an accuracy of 80.3% (chance level for this analysis is 61%). The Kappa value for this analysis was .586, which suggests moderate agreement between the predicted classification of the SM ratings and their actual classification. The results from the LOOCV were similar to the initial DFA and classified 78.9% of the SMs accurately (see Table 2 for the confusion matrix for this analysis). The results indicate that the nine variables can strongly predict whether a SM was rated as low or high HL. The DFA coefficients indicate that SMs rated as low HL contained less sophisticated lexical items, were less cohesive, and reported less complex syntactic structures.

Demographic Analysis

The average age of our study population at the time of the DISTANCE study was 57.5 (±10); 52.4% were male and 27.3% were white. When applying the DFA model to the full dataset (n = 9,530), we found patterns that matched previously observed relationships between patient characteristics and HL. For example, ~75% of the predicted low HL patients were non-white, compared to ~68% of high HL patients. Against expectations, patients identified by the model to have high HL were about 1 year older than low HL patients (see

Table 3). Finally, ~40% of patients with predicted low HL had some college compared to ~56% of patients with high HL.

Health Outcomes

To evaluate whether the categorizations from the DFA model were associated with health outcomes, we assessed how well the modeled HL categories differed with respect to patient reports of communication (providers explaining things in a way that you could understand), glycemic control (mean HbA1c values and poor control based on the proportion with HbA1c >9%) and severity of illness (Charlson score) for the full dataset (n = 9,530). The results indicate that 15.5% of low HL and 11.3% of high HL patients reported poor provider communication and that 17.2% of low HL vs 12.5% of high HL patients had poor glycemic control (HbA1c value >9%). In addition, patients predicted to have low HL had a significantly higher prevalence of comorbid conditions (see Table 4). Analyses of healthcare service use rates demonstrated that patients with predicted low HL had significantly more outpatient clinic visits, emergency room visits, and impatient hospitalizations over the prior year, when compared to patients categorized as having high HL (see Table 5).

Discussion

Prior research has shown that patients with DM2 and limited HL have a higher risk of adverse health outcomes, including poor glycemic control, hypoglycemia, poor medication adherence and greater disease complications. These associations have been replicated across a number of chronic conditions, indicating that limited HL represents a significant clinical and public health problem (Institute of Medicine, 2004;). An important mediator between limited HL and poor health is suboptimal communication exchange between patients and physicians. For instance, patients with low HL report difficulties coming to a shared understanding with their physician about clinical problems and treatment options (Schillinger et al., 2004, 2017). These difficulties are generally attributed to a reduced ability to both communicate and understand health-related ideas (Bailey et al., 2014), combined with physicians' lack of skill in identifying those with limited HL and tailoring their communication to meet the needs of their patients. A recent innovation whose purpose is to enhance communication between patients and physicians is the use of online patient portals that allow asynchronous, between-visit electronic communications via SMs. Research indicates that patient use of these portals is associated with more favorable healthcare utilization patterns, medication adherence, and overall outcomes (Zhou et al., 2010). An assumption of online patient portals is that participants have adequate communicative HL to enable them to successfully extract information and interact effectively with health care providers within the system. Thus, it has become increasingly important to identify those patients with barriers posed by low HL so as to develop and implement interventions to improve healthcare delivery and outcomes and reduce HL-related disparities. However, measuring HL using traditional methods has proven difficult and infeasible to scale (DeWalt et al., 2012).

This study attempts to address this problem by using written corpus analyses, expert human ratings, and NLP approaches to estimate HL at the individual patient level, with the hopes of

not only better understanding HL from a linguistic perspective but also opening new avenues to individualize care and enhance population management. Specifically, we examined HL as a function of patients' demonstrated ability to communicate health-related information to their providers via SMs. In doing so, we build on and complement a recent study by Authors (2019) in our approach to building NLP-based HL models. Like Authors (published), we further validated our HL model by predicting patient-related experiences and events such as communication with their providers, medical outcomes and annual emergency room and hospitalization rates. However, unlike Authors (published), who developed their model based on patients' self-reported HL, we modeled communicative HL using expert ratings of the HL of patients' written health communications as our gold standard. We also provided new insights into the linguistic features that predict communicative HL.

Our results indicate that the developed NLP HL model based on nine linguistic features derived from the patients SMs, predicted human ratings of communicative HL (binary low or high) with ~80% accuracy. Validation testing on a larger sample of patients indicated that being non-white, older and having lower educational attainment were each associated with being more likely to be judged as having low communicative HL. In addition, patients with low HL reported worse communication experiences with their healthcare providers, experienced more negative health outcomes (i.e., worse glycemic control and more comorbid conditions) and had higher healthcare service utilization (i.e., outpatient clinic visits, emergency room visits, and impatient hospitalizations). Overall, the novel communicative HL model not only significantly predicted whether a patient was judged to be of low or high HL, but also was a consistent predictor of demographic, communication and health patterns that are known to be associated with HL.

The model also was informative in terms of the linguistic features that distinguish low and high HL patients. Overall, low HL patients produce words in their SMs that are less sophisticated and produce less complex syntactic structures within their SMs and less globally cohesive SMs. Unpacking the lexical features, we observed that low HL patients produced content words (e.g., verbs and nouns) that have quicker lexical response times, indicating that the words in low HL patients' SMs are more quickly identified as words (versus non-words) when compared to those of high HL patients. In terms of function words (e.g., prepositions, conjunctions, pronouns), low HL patients produced words that are seemingly more difficult in that they have fewer associated words and fewer words that are phonologically similar (i.e., words that sound the same). Thus, it appears that higher HL patients produced structural words (i.e., words that help provide structure to sentences) that are easier to process. This finding may be moderated by the syntactic complexity features in the model which indicated that low HL patients produced language that is less structurally complex than high HL patients. Simpler syntactic structures on the part of low HL patients were found at the phrasal level (e.g., noun phrases) and the clausal level (e.g., more passive structures and more adjective complements per clause). The production of more complex syntactic structures on the part of high HL patients may be a function of *less* sophisticated function words since many phrasal and clausal features depend on frequent prepositions. For example, passive clauses generally include the preposition *by* (e.g., "He was hit *by* the ball") while nominal complements are introduced by common preposition such as *to, in, by,* or *for* (e.g., "She slept *in* the bed").

In terms of discourse structure, one cohesion feature was a significant predictor of low vs. high HL patients: argument overlap at the paragraph level. This feature was lower in low HL patients, demonstrating that they had fewer overlap of arguments (i.e., a base nouns such as *play* which may have various forms such as *play, plays, playing, player, players*) across paragraphs. Less overlap of arguments across larger text segments such as paragraphs makes text less cohesive and more difficult to process because it lowers the global cohesion of a text.

The model developed here can be scaled to a larger population and across other health systems as long as language data are available for patients. As an example, our initial sample of SMs scored by the expert raters was ~500. The model developed for these ~500 patients was later extrapolated to a much larger dataset (~9,000 patients) for validity testing demonstrating that the model is feasible to scale. Our future work will extend this model to over 300,000 patients who are part of a larger sample. A scalable model such as the one reported here provides an efficient means to automatically detect patients with limited HL based on productive language skills. These results can then be used to identify at-risk patients, which can then be used to inform and deploy interventions at the patient-provider and system-populations levels.

We know of only one other approach at automatically assessing HL in patients using NLP techniques (Authors, in press). Similar to this study, the authors examined patients' written health communications to estimate communicative HL, using the linguistic features of patient SMs to predict HL. However, the authors relied on patient survey results (self-reported HL) to develop HL criteria as compared to objective expert ratings used in the current study. Additionally, they used a much larger set of predictor variables (185 linguistic features) and did not apply an ablation test vs. a limited set of linguistic features (nine) as we used in this study. Nevertheless, authors' results were similar to those reported here in that their HL models varied in their test characteristics, with C-statistics ranging from 0.61–0.74. Relations between their HL models and health outcomes revealed patterns consistent with previous HL research such that patients identified as limited HL were older and more likely of minority status, had poorer medication adherence and glycemic control, and exhibited higher rates of hypoglycemia, comorbidities and healthcare utilization. In future work, we intend to directly assess the comparative yield of these NLP models, as well as their performance relative to off-the-shelf, standard readability measures.

Conclusion and Limitations

We have demonstrated how NLP techniques can be used to develop a model of communicative HL derived from patients' written data produced within an online health portal. This HL model performed well at predicting communicative HL in a very large sample of ethnically diverse patients with DM2, and revealed associations with demographic characteristics, health outcomes, and healthcare utilization patterns that track with prior HL research. The results also help us to better understand communicative HL from a linguistic perspective and provide a foundation from which to automatically assess patients' communicative HL across large populations.

In the current study, we focused on patients' ability to communicate health information to their providers via SMs. Evidence from the general literacy field suggests that individuals' ability to write is strongly associated with other domains of literacy, linguistic competence, and problem-solving capacities (Allen, Snow, Crossley, Jackson, & McNamara, 2014; Allen, Dascalu, McNamara, Crossley, & Trausan-Matu, 2016; Crossley, Allen, Snow, & McNamara, 2016; Schoonen, 2019). However, future work in this vein should consider predicting other facets of HL. HL is a multifaceted construct that includes not only the ability of patients to communicate information but also the ability to process, comprehend, and act on health information that they receive. A more comprehensive measure of patients' HL would include not just communication ability, but also patients' ability to read and understand specific health topics, critically appraise and execute health instructions, including verbal instructions, and effectively problem-solve based on a foundation of health-related knowledge (Nutbeam, 2009). Additionally, while the model developed here is a strong indicator of patients' unidirectional communicative ability via online health portals (specifically using SMs), much health communication is not written, and written communication may occur outside of health portals. To capture this variance, future studies should also collect data from spoken exchanges between patients and physicians (Harrington & Valerio, 2014). Relatedly, while our objective was to measure patients' communicative HL, we acknowledge that assessing the linguistic content of only one actor in a communication exchange limited our ability to evaluate communication exchanges and seek evidence (or absence of) comprehension. Nevertheless, our findings that a model of communicative HL derived from expert ratings of patient SMs was predictive of patient reports of poor receptive communication as well as communication-sensitive health outcomes suggests that limited communicative HL may be a marker for less interactive and lower quality bi-directional communication. Further, while the healthcare system analyzed in this study is a large and integrated system with broad representation, it will be important to assess how well the developed model works in other systems and for other types of patients, beyond DM2 patients. Finally, it is possible that combining our NLP approach with other easily and electronically available information such as patient demographics and clinical characteristics could increase the accuracy of our model.

Better communication between physicians and patients has been shown to lessen confusion about medical care, build trust, forge therapeutic alliances, and help patients better manage their health problems (Ratanawongsa et al., 2013). Applying NLP-based strategies could contribute to the broader effort to address the significant clinical and public health problem that is limited HL. The ultimate goal of the current work is to optimize the extent of "shared meaning" resulting from the critical bidirectional communications that patients exchange with clinicians and health systems. Identifying patients likely to have limited HL could prove useful for alerting clinicians about potential difficulties among patients in comprehending written and/or verbal instructions. Additionally, patients identified as having limited HL could be supported better by receiving follow-up communications to ensure understanding of medication instructions and promote adherence (Sudore et al., 2010). To that end, the next stage in our work includes (a) developing automated measures of physicians' linguistic complexity so as to study how communication exchange is affected when there is a match or a mis-match in linguistic complexity between physicians and

patients; and (b) testing the effects of an automated feedback tool on the patient portal to lower the complexity of physician SMs so as to better meet the needs of patients with low communicative HL. Future implementation and dissemination research is needed. This research should include evaluating the transportability of our approach to deriving communicative HL from patients' SMs to diverse healthcare settings, developing provider workflow and/or novel population management approaches when patients with limited HL identified, comparing linguistic models of HL to models developed from demographic data or other sources, and examining the effects of interventions that harness this novel source of information on health-related outcomes.

Acknowledgments

References

Allen LK, Dascalu M, McNamara DS, Crossley S, & Trausan-Matu S (2016). Modeling Individual Differences among Writers Using ReaderBench. In Proceedings of the 8th annual International Conference on Education and New Learning Technologies (EduLearn) (pp. 5269–5279). Barcelona, Spain: IATED.

Allen LK, Snow EL, Jackson GT, Crossley SA, & McNamara DS (2014). Reading components and their relation to writing. L'Année psychologique/Topics in Cognitive Psychology. 114 (4), 663–691.

Bailey SC, Brega AG, Crutchfield TM, Elasy T, Herr H, Kaphingst K, … & Rothman R (2014). Update on health literacy and diabetes. The Diabetes Educator, 40(5), 581–604. [PubMed: 24947871]

Balota DA, Yap MJ, Hutchison KA, Cortese MJ, Kessler B, Loftis B, … & Treiman R (2007). The English lexicon project. Behavior Research Methods, 39, 445–459. [PubMed: 17958156]

Balyan R, Crossley SA, Brown W, Karter AJ, McNamara DS, Liu JY, et al. (2019) Using natural language processing and machine learning to classify health literacy from secure messages: The ECLIPPSE study. PLoS ONE 14(2): e0212488. 10.1371/journal.pone.0212488 [PubMed: 30794616]

Brach C, Dreyer BP, & Schillinger D (2014). Physicians' roles in creating health literate organizations: a call to action. Journal of general internal medicine, 29(2), 273–275. [PubMed: 24113805]

Castro CM, Wilson C, Wang F, & Schillinger D (2007). Babel babble: Physicians' use of unclarified medical jargon with patients. American Journal of Health Behavior, 31(1), S85–S95. [PubMed: 17931142]

Centers for Disease Control and Prevention. (2014). National diabetes statistics report: Estimates of diabetes and its burden in the United States, 2014. Atlanta, GA: US Department of Health and Human Services.

Charlson M, Szatrowski TP, Peterson J, & Gold J (1994). Validation of a combined comorbidity index. Journal of Clinical Epidemiology, 47(11), 1245–1251. [PubMed: 7722560]

Chew LD, Griffin JM, Partin MR, Noorbaloochi S, Grill JP, Snyder A… (2008). Validation of screening questions for limited health literacy in a large VA outpatient population. Journal of general internal medicine. 1 (5), 561–566.

Crossley SA, Allen L, Snow E, & McNamara DS (2016). Incorporating learning characteristics into automatic essay scoring models: What individual differences and linguistic features tell us about writing quality. Journal of Educational Data Mining, 8 (2), 1–19.

Crossley SA, Kyle K, & Dascalu M (in press). The Tool for the Automatic Analysis of Cohesion 2.0: Integrating Semantic Similarity and Text Overlap. Behavioral Research Methods.

Crossley SA, Kyle K & McNamara DS(2015). To aggregate or not? Linguistic features in automatic essay scoring and feedback systems. Journal of Writing Assessment, 8. Retrieved from: http://journalofwritingassessment.org/article.php?article=80.

Crossley SA, Kyle K, & McNamara DS (2016). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. Behavior Research Methods, 48, 1227–1237. [PubMed: 26416138]

Crossley SA, & McNamara DS (2016). Say more and be more coherent: How text elaboration and cohesion can increase writing quality. Journal of Writing Research, 7(3), 351–370.

Davies M (2009). The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. International Journal of Corpus Linguistics, 14(2), 159–190.

DeWalt DA, Schillinger D, Ruo B, Bibbins–Domingo K, Baker DW, Holmes GM, … & Grady KL (2012). Multisite randomized trial of a single–session versus multisession literacy-sensitive self-care intervention for patients with heart failure. Circulation, 125(23), 2854–2862. [PubMed: 22572916]

Deyo RA, Cherkin DC, & Ciol MA (1992). Adapting a clinical comorbidity index for use with ICD–9–CM administrative databases. Journal of Clinical Epidemiology, 45(6), 613–619. [PubMed: 1607900]

Diviani N, van den Putte B, Giani S, & van Weert JC (2015). Low health literacy and evaluation of online health information: A systematic review of the literature. Journal of Medical Internet Research, 17(5).

Grossman EG. (2010). Patient Protection and Affordable Care Act. Washington, D.C.: Department of Health & Human Services.

Harrington KF, & Valerio MA (2014). A conceptual model of verbal exchange health literacy. Patient Education and Counseling, 94(3), 403–410. [PubMed: 24291145]

Haun JN, Valerio MA, McCormack LA, Sørensen K, & Paasche-Orlow MK (2014). Health literacy measurement: an inventory and descriptive summary of 51 instruments. Journal of health communication, 19(sup2), 302–333. [PubMed: 25315600]

Hill–Briggs F, Schumann KP, & Dike O (2012). Five-step methodology for evaluation and adaptation of print patient health information to meet the< 5th grade readability criterion. Medical Care, 50(4), 294–301. [PubMed: 22354210]

Institute of Medicine (2004). Health literacy: A prescription to end confusion. Washington, D.C.: The National Academies Press.

Johnson SB (1999). A semantic lexicon for medical language processing. Journal of the American Medical Informatics Association, 6(3), 205–218. [PubMed: 10332654]

Karter AJ, Parker MM, Moffet HH, Ahmed AT, Schmittdiel JA, & Selby JV (2009). New prescription medication gaps: A comprehensive measure of adherence to new prescriptions. Health Services Research, 44(5p1), 1640–1661. [PubMed: 19500161]

Karter AJ, Parker MM, Duru OK, Schillinger D, Adler NE, Moffet HH, … & Schmittdiel JA (2015). Impact of a pharmacy benefit change on new use of mail order pharmacy among diabetes patients: The Diabetes Study of Northern California (DISTANCE). Health Services Research, 50(2), 537–559. [PubMed: 25131156]

Kim H, Goryachev S, Rosemblat G, Browne A, Keselman A, & Zeng–Treitler Q (2007). Beyond surface characteristics: a new health text-specific readability measurement. In AMIA Annual Symposium Proceedings (pp. 418–422). American Medical Informatics Association.

Kincaid JP, Fishburne RP Jr, Rogers RL, & Chissom BS (1975). Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. DTIC Document.

Kirsch IS, Jungeblut A, Jenkins L, & Kolstad A (2002). Adult literacy in America: A first look at the findings of the national adult literacy survey (NCES 1993–275). Washington, DC: U.S. Department of Education.

Kiss GR, Armstrong C, Milroy R, & Piper J (1973). An associative thesaurus of English and its computer analysis. Computer and Literary Studies, 153–165.

Kyle K (2015). Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage–based indices of syntactic sophistication. Doctoral dissertation. Georgia State University. Retrieved from http://scholarworks.gsu.edu/alesl_diss/35

Kyle K, & Crossley SA (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. TESOL Quarterly, 49(4), 757–786.

Kyle K, & Crossley SA (2018). Measuring Syntactic Complexity in L2 Writing Using Fine-Grained Clausal and Phrasal Indices. Modern Language Journal, 102 (2), 333–349.

Kyle K, Crossley S, & Berger C (2018). The tool for the automatic analysis of lexical sophistication (TAALES): version 2.0. Behavior Research Methods, 50(3), 1030–1046. [PubMed: 28699123]

Landis JR, & Koch GG (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics, 363–374. [PubMed: 884196]

Manning C, Surdeanu M, Bauer J, Finkel J, Bethard S, & McClosky D (2014). The Stanford CoreNLP natural language processing toolkit. In Proceedings of 52nd annual meeting of the association for computational linguistics: System demonstrations (pp. 55–60).

McNamara DS, Crossley SA, Roscoe RD, Allen LK, & Dai J (2015). A hierarchical classification approach to automated essay scoring. Assessing Writing, 23, 35–59.

Nutbeam D (2000). Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. Health promotion international, 15(3), 259–267.

Nutbeam D (2009). Defining and measuring health literacy: what can we learn from literacy studies? International Journal of Public Health, 54, 303. [PubMed: 19641847]

Ratanawongsa N, Karter AJ, Parker MM, Lyles CR, Heisler M, Moffet HH, … & Schillinger D (2013). Communication and medication refill adherence: The Diabetes Study of Northern California. JAMA Internal Medicine, 173(3), 210–218. [PubMed: 23277199]

Sarkar U, Karter AJ, Liu JY, Moffet HH, Adler NE, & Schillinger D (2010). Hypoglycemia is more common among type 2 diabetes patients with limited health literacy: The Diabetes Study of Northern California (DISTANCE). Journal of General Internal Medicine, 25(9), 962–968. [PubMed: 20480249]

Sarkar U, Piette JD, Gonzales R, Lessler D, Chew LD, Reilly B, … & Schillinger D (2008). Preferences for self–management support: Findings from a survey of diabetes patients in safety-net health systems. Patient Education and Counseling, 70(1), 102–110. [PubMed: 17997264]

Schillinger D (2007). Literacy and health communication: reversing the 'inverse care law'. The American Journal of Bioethics, 7(11), 15–18.

Schillinger D, Bindman A, Wang F, Stewart A, & Piette J (2004). Functional health literacy and the quality of physician-patient communication among diabetes patients. Patient Education and Counseling, 52(3), 315–323. [PubMed: 14998602]

Schillinger D, Grumbach K, Piette J, Wang F, Osmond D, Daher C, … & Bindman AB (2002). Association of health literacy with diabetes outcomes. JAMA, 288(4), 475–482. [PubMed: 12132978]

Schillinger D, Hammer H, Wang F, Palacios J, McLean I, Tang A, … & Handley M (2008). Seeing in 3-D: examining the reach of diabetes self-management support strategies in a public health care system. Health Education & Behavior, 35(5), 664–682. [PubMed: 17513690]

Schillinger D, Handley M, Wang F, & Hammer H (2009). Effects of self-management support on structure, process, and outcomes among vulnerable patients with diabetes: a three-arm practical clinical trial. Diabetes care, 32(4), 559–566. [PubMed: 19131469]

Schillinger D, McNamara D, Crossley S, Lyles C, Moffet HH, Sarkar U, … & Ratanawongsa N (2017). The next frontier in communication and the ECLIPPSE study: Bridging the linguistic divide in secure messaging. Journal of Diabetes Research.

Schoonen R (2019). Are reading and writing building on the same skills? The relationship between reading and writing in L1 and EFL. Reading and Writing, 32(3), 511–535.

Smith SG, O'conor R, Curtis LM, Waite K, Deary IJ, Paasche–Orlow M, & Wolf MS (2015). Low health literacy predicts decline in physical function among older adults: Findings from the LitCog cohort study. Journal of Epidemiol Community Health, 69(5), 474–480.

Sudore RL, Landefeld CS, Perez-Stable EJ, Bibbins-Domingo K, Williams BA, & Schillinger D (2009). Unraveling the relationship between literacy, language proficiency, and patient–physician communication. Patient education and counseling, 75(3), 398–402. [PubMed: 19442478]

Sudore RL, Yaffe K, Satterfield S, Harris TB, Mehta KM, Simonsick EM, … & Ayonayon HN (2006). Limited literacy and mortality in the elderly. Journal of general internal medicine, 21(8), 806–812. [PubMed: 16881938]

Weiss BD (2015). Health literacy research: Isn't there something better we could be doing?. Health Communication, 30(12), 1173–1175. [PubMed: 26372029]

Wu DT, Hanauer DA, Mei Q, Clark PM, An LC, Lei J, … & Zheng K (2013). Applying multiple methods to assess the readability of a large corpus of medical documents. Studies in Health Technology and Informatics, 192, 647–651. [PubMed: 23920636]

Zeng–Treitler Q, Kandula S, Kim H, & Hill B (2012). A method to estimate readability of health content. Association for Computing Machinery.

Zheng J, & Yu H (2018). Assessing the readability of medical documents: A ranking approach. JMIR medical informatics, 6(1).

Zhou YY, Kanter MH, Wang JJ, & Garrido T (2010). Improved quality at Kaiser Permanente through email between physicians and patients. Health Affairs, 29(7), 1370–1375. [PubMed: 20606190]

Table 1

Descriptive statistics and MANOVA results for linguistic features among patients with low and high communicative HL

Index	Low HL SMs mean (SD)	High HL SMs mean (SD)	F	η 2
Lexical decision response time (CW)	617.375 (5.768)	623.51 (6.133)	127.73	0.200
Attested lemmas (COCA magazine)	0.977 (0.021)	0.988 (0.01)	75.988	0.130
Determiner per nominal phrase	0.256 (0.07)	0.302 (0.053)	70.225	0.121
Passive auxiliary verbs per clause	0.024 (0.015)	0.035 (0.017)	51.706	0.092
Dependents per nominal subject	0.93 (0.279)	1.07 (0.22)	39.932	0.073
Number of associations (EAT, FW)	89.066 (2.088)	90.058 (1.694)	34.729	0.064
Argument overlap, paragraph	0.894 (0.142)	0.943 (0.077)	25.084	0.047
Adjective complements per clause	0.057 (0.031)	0.068 (0.022)	23.481	0.044
Phonological Neighborhood Frequency (FW)	8.879 (0.094)	8.907 (0.097)	10.937	0.021

CW = Content words, EAT =, FW = Function words

Table 2

Confusion matrix for DFA categorizing SMs as low and high communicative HL

		Low HL	High HL	
Whole set	Low HL	150	50	200
	High HL	51	261	312
		Low HL	High HL	
Cross-validated	Low HL	146	54	200
	High HL	54	258	312

Table 3

Descriptive statistics for demographic comparisons between patients with low and high communicative HL

Race – White %		Age at Survey – Mean (SD)		Education – Some college %		P-value
Low HL	High HL	Low HL	High HL	Low HL	High HL	
25.5	31.8	57.91 (9.84)	56.61 (10.5)	39.6	54.2	all < 0.001

Table 4

Descriptive statistics for health outcomes between lower and higher HL patients

A1c % (mean)		HbA1c > 9 (%),		Charlson Index		
Low HL	High HL	Low HL	High HL	Low HL	High HL	P-value
7.67 (1.62)	7.43 (1.48)	17.00	13.30	2.44 (1.80)	2.31 (1.69)	all < 0.001

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5

Descriptive statistics for annual healthcare service use between patients with low and high communicative HL

Outpatient clinic visit		Emergency room visits		Hospitalizations		P-value
Low HL	High HL	Low HL	High HL	Low HL	High HL	
10.06 (10.6)	9.53 (9.51)	0.49 (1.14)	0.41 (1.01)	0.22 (0.69)	0.19 (0.64)	all < 0.050