RESEARCH ARTICLE

Employing computational linguistics techniques to identify limited patient health literacy: Findings from the ECLIPPSE study

Dean Schillinger MD[1,2,3] | Renu Balyan PhD[4] | Scott A. Crossley PhD[5] | Danielle S. McNamara PhD[6] | Jennifer Y. Liu MPH[2] | Andrew J. Karter PhD[1,2]

[1]UCSF Division of General Internal Medicine, Zuckerberg San Francisco General Hospital and Trauma Center, San Francisco, California, USA

[2]Division of Research, Kaiser Permanente Northern California, Oakland, California, USA

[3]UCSF Health Communications Research Program, Center for Vulnerable Populations, Zuckerberg San Francisco General Hospital and Trauma Center, San Francisco, California, USA

[4]Ira A. Fulton School of Engineering, Arizona State University, Mesa, Arizona, USA

[5]Department of Applied Linguistics/ESL, College of Arts and Sciences, Georgia State University, Atlanta, Georgia, USA

[6]Department of Psychology, Arizona State University, Tempe, Arizona, USA

Correspondence
Dean Schillinger, MD, UCSF Division of General Internal Medicine, Zuckerberg San Francisco General Hospital and Trauma Center, San Francisco, CA 94110, USA.
Email: dean.schillinger@ucsf.edu

Abstract

Objective: To develop novel, scalable, and valid literacy profiles for identifying limited health literacy patients by harnessing natural language processing.

Data Source: With respect to the linguistic content, we analyzed 283 216 secure messages sent by 6941 diabetes patients to physicians within an integrated system's electronic portal. Sociodemographic, clinical, and utilization data were obtained via questionnaire and electronic health records.

Study Design: Retrospective study used natural language processing and machine learning to generate five unique "Literacy Profiles" by employing various sets of linguistic indices: Flesch-Kincaid (LP_FK); basic indices of writing complexity, including lexical diversity (LP_LD) and writing quality (LP_WQ); and advanced indices related to syntactic complexity, lexical sophistication, and diversity, modeled from self-reported (LP_SR), and expert-rated (LP_Exp) health literacy. We first determined the performance of each literacy profile relative to self-reported and expert-rated health literacy to discriminate between high and low health literacy and then assessed Literacy Profiles' relationships with known correlates of health literacy, such as patient sociodemographics and a range of health-related outcomes, including ratings of physician communication, medication adherence, diabetes control, comorbidities, and utilization.

Principal Findings: LP_SR and LP_Exp performed best in discriminating between high and low self-reported (*C*-statistics: 0.86 and 0.58, respectively) and expert-rated health literacy (*C*-statistics: 0.71 and 0.87, respectively) and were significantly associated with educational attainment, race/ethnicity, Consumer Assessment of Provider and Systems (CAHPS) scores, adherence, glycemia, comorbidities, and emergency department visits.

Conclusions: Since health literacy is a potentially remediable explanatory factor in health care disparities, the development of automated health literacy indicators

Dean Schillinger and Renu Balyan are co-first authors.

represents a significant accomplishment with broad clinical and population health applications. Health systems could apply literacy profiles to efficiently determine whether quality of care and outcomes vary by patient health literacy; identify at-risk populations for targeting tailored health communications and self-management support interventions; and inform clinicians to promote improvements in individual-level care.

1 | INTRODUCTION

Patient-physician communication is a fundamental pillar of care that influences patient satisfaction and health outcomes,[1] particularly in diabetes mellitus.[2] More than 30 million US adults are living with diabetes,[3] and one quarter to one third of them has limited health literacy skills. Limited health literacy is associated with untoward and costly diabetes outcomes that contribute to health disparities.[4-6] Limited health literacy impedes physician-patient communication, as well as imparts a barrier to patients' learning and understanding across numerous communication domains.[7-10]

Being able to assess patients' health literacy is of interest to clinicians, delivery systems, and the public health community.[11] Clinicians often are unaware of the health literacy status of their patients and have been found to both be receptive to receiving this information as well as responsive.[12] Ignoring differences in health literacy in population management has been shown to amplify health literacy-related disparities.[6] To date, identifying limited health literacy patients has proven painstaking and infeasible to scale.[13] Because "big data"—in this case data derived from patients' written secure messages sent via patient portals—are increasingly available, we sought to determine whether natural language processing tools and machine learning approaches can be utilized to identify patients with limited health literacy. An automated process, if it could generate health literacy estimates with sufficient accuracy, would provide an efficient means to identify patients with limited health literacy, with a number of implications for improving health services delivery. The few formulas used in prior health literacy studies of written text (eg, Flesch-Kincaid, SMOG) depend on surface-level lexical and sentential features, have not examined secure messages, and have not used natural language processing and machine learning. We are aware of only two studies that attempted to identify patient health literacy using secure message content.[14,15] Both studies developed predictive models of health literacy using natural language processing based on linguistic features extracted from secure messages. These "Literacy Profiles" were generated from patients' self-reported health literacy[14] and expert ratings of health literacy based on secure message quality[15] and showed promising results.

WHAT IS KNOWN ON THIS TOPIC

- Limited health literacy is associated with untoward and costly health outcomes that contribute to health disparities, and poor communication exchange is an important mediator in the relationship between limited health literacy and health outcomes.
- Given the time and personnel demands intrinsic to current health literacy instruments, combined with the sensitive nature of screening, measuring health literacy is both challenging and controversial.
- Electronic patient portals are an increasingly popular channel for patients and providers to communicate via secure messaging, and secure messages contain linguistic content that could be anlayzed to measure patient health literacy.

WHAT THIS STUDY ADDS

- Two valid literacy profiles from patients' secure message were generated by applying computational linguistics approaches to "big linguistic data", creating a novel, feasible, automated, and scalable strategy to identify patients and subpopulations with limited health literacy.
- Literacy profiles can provide a health IT tool to enable tailored communication support and other targeted interventions with potential to reduce health literacy-related disparities.

To advance methods for identifying patients' health literacy automatically, we built on this prior work, and developed and compared five literacy profiles based on distinct theoretical models and associated natural language processing (NLP) tools and machine learning techniques. The primary goal of the current study was to compare the relative performance of these literacy profiles with respect to their ability to discriminate between limited vs. adequate health literacy in
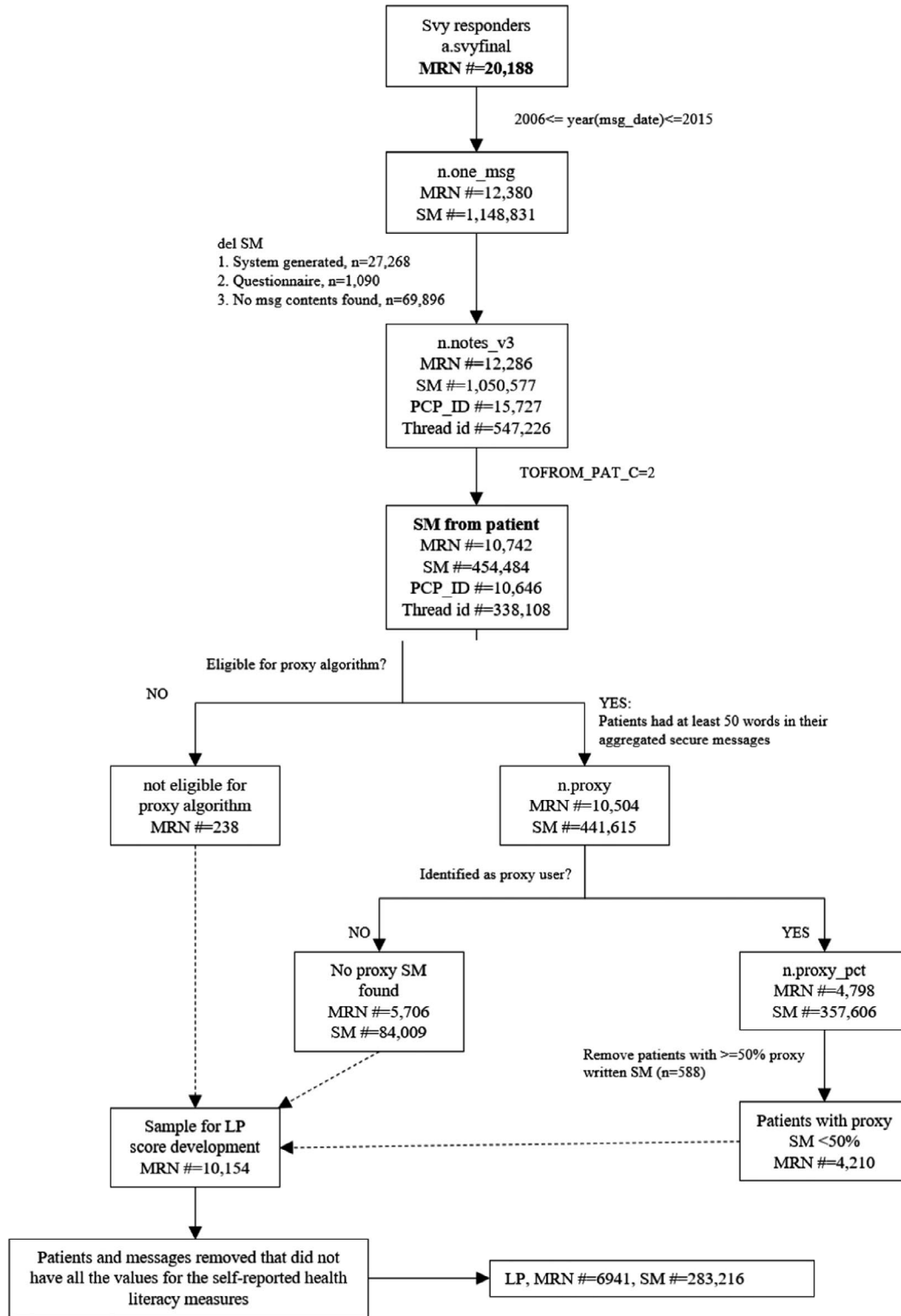
FIGURE 1 Patient and secure messages inclusion/exclusion flowchart*. *MRN#: Patient ID; msg_date: Date of message sent; Svy: survey; SM#: number of secure messages; LP: literacy profile; PCP_ID: primary care provider ID; proxy_pct: % of proxy messages; TOFROM_PAT_C: SM sent by the patient

a large sample of diabetes patients based on a large written corpus of patients' secure messages. The secondary goal was to assess the extent to which these different literacy profiles are associated with

patterns that mirror previous research in terms of their relationships with patient sociodemographics, ratings of physician communication, and a range of diabetes-related health outcomes.

2 | METHODS

2.1 | Data sources and participants

Our sampling frame included over one million secure messages generated by >150 000 ethnically diverse diabetes patients in the Kaiser Permanente Northern California (KPNC) Diabetes Registry, and >9000 primary care physicians. KPNC is a fully integrated health system that provides care to ~4.4 million patients and supports a well-developed and mature patient portal (kp.org).

The current study includes the subset of the KPNC registry patients who completed a 2005-2006 survey as part of the Diabetes Study of Northern California (DISTANCE) and responded to the self-reported health literacy items on the associated questionnaire (N = 14 357).[2,16,17] DISTANCE surveyed diabetes patients, oversampling minority subgroups to assess the role of sociodemographic factors on quality and outcomes of care. The average age of the study population at the time was 56.8 (±10); 54.3 percent were male; and 18.4 percent Latino, 16.9 percent African American, 22.8 percent Caucasian, 11.9 percent Filipino, 11.4 percent Asian (Chinese or Japanese), 7.5 percent South Asian/Pacific Islander/ Native American/Eskimo, and 11.0 percent multi-racial. Variables were collected from questionnaires completed via telephone, online or paper and pencil (62 percent response rate). Details of the DISTANCE Study have been reported previously.[17]

We first extracted all secure messages (N = 1 050 577) exchanged from 01/01/2006 through 12/31/2015 between diabetes patients and all clinicians from KPNC's patient portal. For the current analyses, only those secure messages that a patient sent to his or her primary care physician were included. We excluded all secure messages: from patients who did not have matching DISTANCE survey data; written in a language other than English; and written by proxy caregivers (determined by the KP.org proxy check-box or by a validated NLP algorithm[18]). The study flowchart (Figure 1) shows details about inclusion/ exclusion of patients and associated secure messages. The final dataset consisted of 283 216 secure messages sent by 6941 patients to their primary care physicians.

The number of individual SMs sent by a patient to their physician(s) ranged between 2 and 205, and the mean number of SMs sent was 39.88. For each patient, all secure messages were then collated into a single file. The length of patients' aggregated SMs ranged from 1 word and 16 469 words, with a mean length of 2058.95 words. To provide appropriate linguistic coverage to develop literacy profiles, we excluded patients whose aggregated secure messages lacked sufficient words (<50 words, see Figure 1), a threshold based on previous NLP text research in learning analytics domains.[19,20]

This study was approved by the KPNC and UCSF Institutional Review Boards (IRBs). All analyses involved secondary data and all data were housed on a password-protected secure KPNC server that could only be accessed by authorized researchers.

2.2 | Health literacy "Gold Standards"

DISTANCE survey included three validated health literacy items that measure self-efficacy in specific health literacy competencies using a 5-point Likert scale in which a response of 1 referred to "Always" and a response of 5 to "Never".[21] Questions include self-reported confidence in filling out medical forms, problems understanding written medical information, and frequency of needing help in reading and understanding health materials. We combined these items to create a self-reported health literacy variable to compare performance of the linguistic models,[14] by averaging scores across the health literacy items. Average scores were dichotomized to create binary data, with scores <4.5 indicating limited health literacy and ≥4.5 indicating adequate health literacy.[14] The threshold was determined based on the distribution of these average scores to maintain the appropriate balance between the health literacy categories and is consistent with prior studies that have employed these measures.[10,16,21]

We also generated health literacy scores based on expert ratings of the quality of patients' secure messages. These ratings used a subset of the DISTANCE sample, comprised of aggregated secure messages written by 512 patients purposively sampled to represent a balance of self-reported health literacy, as well as a range of age, race/ethnicity, and socio-economic status.[15] A health literacy scoring rubric was used to holistically assess the perceived health literacy of the patients based on their secure messages, adapting an established rubric used to score the writing abilities of high school students entering college.[22] An ordinal scale ranging from 1 to 6 [15] assessed the extent to which patients' secure messages demonstrated mastery of written English, organization and focus, and a varied, accurate, and appropriate health vocabulary to enable clear access to the health-related content and ideas the patient intended to express to their physician. Because of limited relevance to the construct of health literacy, we removed aspects of the rubric related to length, developing point of views, and discourse-related elements important in argumentative writing including the use of examples, reason, and evidence. Two raters experienced in linguistics and health literacy research were trained twice on 25 separate aggregated secure messages not included in the 512 messages used in the final analysis. After reaching a satisfactory inter-rater reliability (IRR, $r > .70$), raters independently scored the 512 messages. Secure messages were categorized into two groups: limited health literacy (scores < 4, n = 200) and adequate health literacy (scores ≥ 4, n = 312).

We examined for existence of any associations between the self-reported and expert-rated health literacy measures before these two measures were employed to train literacy profiles. The Cramer's V and chi-squared tests were used to measure the strength of association and significance of that association (effect size). The two variables were significantly different ($P = .01$) and only weakly correlated ($r = 0.118$, $P = .001$).

Literacy profile	Linguistic indices	Description
LP_FK	Readability	The length of words (ie, number of letters or syllables) and length of sentences (ie, number of words)
LP_LD	Lexical Diversity	The variety of words used in a text based on D
LP_WQ	Word Frequency	Frequency of word in a reference corpus
	Syntactic Complexity	Number of words before the main verb in a sentence
	Lexical Diversity	The variety of words used in a text based on MTLD
LP_SR	Concreteness	The degree to which a word is concrete
	Lexical diversity	The variety of words used in a text based on two measures of lexical diversity: MTLD, and D
	Present tense	Incidence of present tense
	Determiners	Incidence of determiners (eg, *a*, *the*)
	Adjectives	Incidence of adjectives
	Function words	Incidence of function words such as prepositions, pronouns etc
LP_Exp	Age of Exposure	The estimated age at which a word first appears in a child's vocabulary
	Lexical decision response time	The time it takes for a human to judge a string of characters as a word
	Attested lemmas	Number of attested lemmas used per verb argument construction
	Determiner per nominal phrase	Number of determiners in each noun phrase
	Dependents per nominal subject	Number of structural dependents for each subject in a noun phrase
	Number of associations	Number of words strongly associated with a single word

TABLE 1 Linguistic indices used in five literacy profiles[a]

Abbreviations: LP_Exp, Literacy Profile Expert-Rated Health Literacy; LP_FK, Literacy Profile Flesch-Kincaid; LP_LD, Literacy Profile Lexical Diversity; LP_SR, Literacy Profile Self-Reported Health Literacy; LP_WQ, Literacy Profile Writing Quality.

[a]We present examples of linguistic indices for LP_SR (n = 185) and LP_Exp (n = 8).

2.3 | Natural language processing (NLP) tools

The linguistic features we examined were derived from the patients' secure messages using several NLP tools that measure different language aspects, such as text level information (eg, number of words in the text, type-token ratio), lexical sophistication (eg, word frequency, concreteness), syntactic complexity (embedded clause and phrasal complexity), and text cohesion (eg, connectives, word overlap). These tools were selected because they measure linguistic features that are important aspects of literacy, including text complexity, readability, and cohesion. The tools included the Tool for the Automatic Assessment of Lexical Sophistication,[23,24] the Tool for the Automatic Analysis of Cohesion,[25] the Tool for the Automatic Assessment of Syntactic Sophistication and Complexity,[26,27] the SEntiment ANalysis and Cognition Engine,[28] and Coh-Metrix.[29] These open-access tools rely on several NLP packages to process text including the Stanford Parser,[30] the British National Corpus,[31] the MRC psycholinguistic database,[32] Collins Birmingham University International Language Database frequency norms,[33] and Wordnet.[34] These tools have been developed using Python and Java. To generate word frequencies for medical terminology, we used medical corpora from HIstory of MEdicine coRpus Annotation[35] and Informatics for Integrating Biology and the Bedside.[36-39]

2.4 | Literacy profiles developed

Using the patients' secure messages, we applied natural language processing and machine learning techniques to develop five separate literacy profile prototypes for categorizing both patients' self-reported and expert-rated health literacy. As a result, the literacy profiles differed based on the dependent health literacy variable (self-reported or expert-rated) and the linguistic features that were used as independent variables to develop these literacy profiles. Each literacy profile is briefly discussed below, and the component linguistic indices used for each are summarized in Table 1.

2.4.1 | Literacy Profile Flesch-Kincaid (LP_FK)

As a "baseline" literacy profile, we calculated Flesch-Kincaid readability scores[40] for the secure messages. We used Flesch-Kincaid as a baseline measure because it is one of the most commonly used and widely available readability formulas in medical domain, including assessing the readability and comprehensibility of a broad range of medical information.[41-46] Flesch-Kincaid is based on average number of words per sentence and average number of syllables per word.

2.4.2 | Literacy Profile Lexical Diversity (LP_LD)

We used lexical diversity as an additional baseline measure because it is a commonly used and a straightforward method for assessing writing proficiency in the linguistics domain and it captures both lexical richness and text cohesion.[47,48] Both these features are consistent predictors of text sophistication and writing quality.[29,49] We calculated lexical diversity based on a type-token ratio (TTR) measure. TTR measures assess lexical variety based on the number of words produced (tokens) divided by the number of unique words produced (types), to evaluate writers' lexical production. TTR measure D[47] was used because it controls for text length by calculating probability curves that mathematically model how new words are introduced into increasingly large language samples.

2.4.3 | Literacy Profile Writing Quality (LP_WQ)

We used a previously validated model[50] to classify secure messages as either low or high in terms of writing quality. The model, derived from three linguistic indices of word frequency, syntactic complexity, and lexical diversity,[29,48] reveals that higher-level writers use more infrequent and lexical diverse words and more syntactically complex structures.

2.4.4 | Literacy Profile Self-Reported Health Literacy (LP_SR)

A set of 185 linguistic features was calculated from the patients' secure messages and used to predict patients' self-reported health literacy scores. A subset of the linguistic indices used for developing this literacy profile are provided in Table 1. The rationale, development, and experimental design for LP_SR have been briefly discussed in the Health Literacy "Gold Standards" section, and the details have also been previously reported.[14]

2.4.5 | Literacy Profile Expert-Rated Health Literacy (LP_Exp)

A set of eight linguistic indices, including lexical decision latencies, age of exposure, word naming response times, academic word lists, bigrams association strength, and dependency structures, were used as independent variables to predict human ratings of health literacy from the purposively sampled subset of 512 secure messages used in the LP_SR analysis (Table 1). Additional details related to the development and experimental design of LP_Exp have been previously reported[15] and can also be found in the Health Literacy "Gold Standards" section.

2.5 | Assessing performance of literacy profiles against gold standards

We compared the performance of the five literacy profiles using several supervised machine learning classification algorithms: linear discriminant analysis (LDA), random forests, support vector machine (SVM), naïve Bayes, and neural networks. In a supervised machine learning model, the algorithm learns from a labeled dataset, providing an answer key that the algorithm can use to classify unseen data and evaluate its accuracy. There are two main areas where supervised machine learning is useful: classification and regression. Classification problems ask the algorithm to predict a discrete value, identifying the input data as the member of a class, or group.[51-56] The models in this study were trained and tested using Weka (version 3.8.1) and R (version 3.3.2) implementations. We first examined performance between the self-reported health literacy and other literacy profiles, and then between the expert ratings of health literacy and all other literacy profiles. We report results for the models using support vector machines, because it yielded the best results for all the literacy profiles. Using a randomly allocated split-sample approach, we report discriminatory performance results using c-statistics (area under the receiver operator [ROC] curves), sensitivity, specificity, and positive and negative predictive values (PPV and NPV).

2.6 | Assessing criterion-related validity for literacy profiles

We examined associations between the health literacy classifications generated by the literacy profiles and known correlates of health literacy including patients' educational attainment, race/ethnicity, and age. Because of the known association between limited health literacy and suboptimal patient-provider communication,[7-10] we also examined relationships with patients' reports of physician communication using an adapted version of the most health literacy-relevant item from the 4-item CAHPS survey[7]: "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.[2] We also examined the extent to which each literacy profile was associated with diabetes-related outcomes previously found to

be associated with health literacy. These included adherence to cardio-metabolic medications based on continuous medication gaps (CMG),[57,58] a validated measure based on percent time with insufficient medication supply; hemoglobin A1c (HbA1c), an integrated measure of blood sugar control, measured both as optimal (HbA1c ≤ 7 percent) and poor control (HbA1c ≥ 9 percent); ≥1 clinically relevant hypoglycemic episodes (an adverse drug event associated with diabetes treatment)[59]; and comorbidities, using the Charlson index[60,61] (Deyo version).[62] HbA1c reflected the value collected after the first secure message was sent. CMG, hypoglycemia, and Charlson index were measured the year before the first secure message. The occurrence of one or more hypoglycemia-related ED visits or hospitalizations in the year prior was based on a validated algorithm that uses specific diagnostic codes.[63] Finally, we explored relationships between each literacy profile and outpatient, emergency room, and hospitalization utilization data 12 months prior to the first secure message date. For all analyses, we examined bivariate associations between each of the literacy profiles and sociodemographics, the single CAHPS item, and health outcomes using a two-sided p-value at the 0.05 level. Categorical variables such as education, race, adherence,[64] HbA1c levels, and hypoglycemia were analyzed using chi-square analysis. For comorbidity and health care utilization rates, mean comparisons were conducted using t tests.

3 | RESULTS

3.1 | Criterion-related validity of literacy profiles based on performance against gold standards

3.1.1 | Performance with respect to self-reported health literacy

When self-reported health literacy was the dependent variable, LP_SR performed well in terms of its ability to discriminate between those with limited vs. adequate health literacy, with a c-statistic of 0.86; sensitivity was high, specificity was modest (0.67), and PPV and NPV were acceptable. All other literacy profiles performed poorly (Figure 2). LP_Exp, while performing slightly better than LP_FK, LP_LD, and LP_WQ, yielded a c-statistic of 0.58 and sensitivity in the intermediate range; specificity, PPV, and NPV values were all low.

3.1.2 | Performance with respect to expert-rated health literacy

When expert-rated health literacy was the dependent variable, LP_Exp performed best in terms of its ability to discriminate
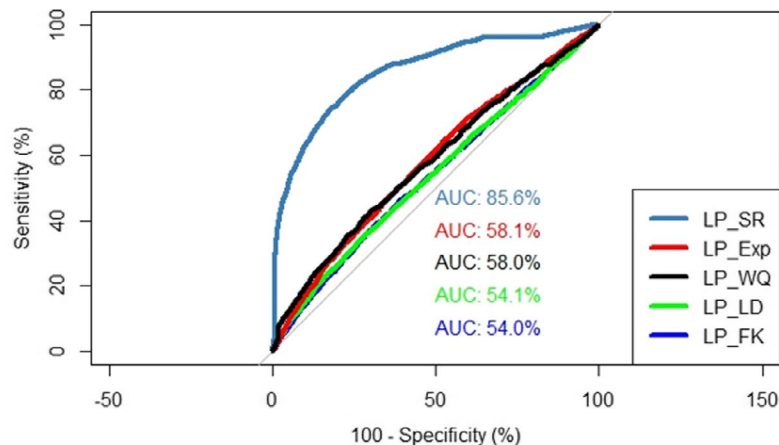
LP	Number of indices	ML Algorithm	C-statistic	Sensitivity	Specificity	PPV	NPV
LP_FK	Readability (1)	SVM	0.540	0.8297	0.2288	0.5496	0.5423
LP_LD	Lexical Diversity (1)	SVM	0.541	0.9485	0.0891	0.5528	0.5933
LP_WQ	Linguistic variables (3)	SVM	0.580	0.8479	0.2586	0.5647	0.5999
LP_SR	Linguistic variables (185)	SVM	0.856	0.8774	0.6650	0.7567	0.8204
LP_Exp	Linguistic variables (8)	SVM	0.581	0.7450	0.3636	0.5819	0.5506

FIGURE 2 ROCs and performance metrics for the literacy profiles relative to self-reported health literacy. AUC: Area Under Curve; LP_Exp: Literacy Profile Expert-Rated Health Literacy; LP_FK: Literacy Profile Flesch-Kincaid; LP_LD: Literacy Profile Lexical Diversity; LP_SR: Literacy Profile Self-Reported Health Literacy; LP_WQ: Literacy Profile Writing Quality; ML: Machine Learning; SVM: Support Vector Machine [Color figure can be viewed at wileyonlinelibrary.com]

between those with limited vs. adequate health literacy, with a c-statistic of 0.87, high sensitivity, moderate specificity, and PPV, and NPV in an acceptable range. LP_FK and LP_LD each performed poorly (Figure 3). LP_WQ performed better, with all the performance metrics > 0.70 except for specificity. Performance metrics for LP_SR were sub-optimal, with a c-statistic of 0.71, intermediate sensitivity, moderate specificity, and PPV, but low NPV.

3.2 | Predictive validity based on associations with sociodemographics, communication ratings, and health-related outcomes

3.2.1 | Sociodemographics

We found patterns that mirrored previously observed health literacy-related relationships, with considerable variation across patient characteristics and literacy profile type. Table 2 shows the educational attainment (% with college degree vs. less), race (% white vs. non-white), and the mean age among patients predicted to have adequate vs. inadequate health literacy for each of the five literacy profiles. All literacy profiles generated classifications in which limited health literacy was associated with non-white race. LP_LD, LP_SR,

and LP_Exp each were associated with lower education, with the strongest effects observed for LP_SR and LP_Exp. Only LP_SR was associated with older patient age.

3.2.2 | Provider communication

The proportion of patients identified as having limited or adequate health literacy and who reported poor physician communication is shown in Table 3. Those patients predicted to have limited health literacy by LP_SR and LP_Exp only were significantly more likely to rate their health care providers as "poor" on the CAHPS item, with somewhat more robust findings for LP_SR.

3.2.3 | Health outcomes

Limited health literacy as categorized only by the three literacy profiles (LP_FK, LP_SR, and LP_Exp) was associated with poor cardiometabolic medication adherence, serious hypoglycemia and greater comorbidity. Poor medication adherence was most robustly associated with LP_FK and LP_Exp. Limited health literacy as measured only by LP_FK and LP_Exp was associated with both optimal and poor diabetes control.
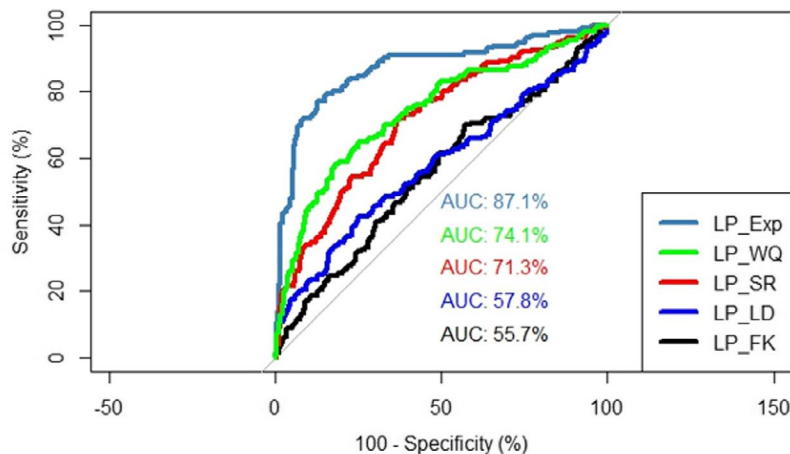
LP	Number of Indices	ML Algorithm	C-statistic	Sensitivity	Specificity	PPV	NPV
LP_FK	Readability (1)	LDA	0.557	0.9968	0.0050	0.6098	0.5000
LP_LD	Lexical Diversity (1)	SVM	0.578	0.9839	0.1000	0.6296	0.6296
LP_WQ	Linguistic variables (3)	SVM	0.741	0.8718	0.4950	0.7292	0.7122
LP_SR	Linguistic variables (185)	SVM	0.713	0.6945	0.6000	0.7297	0.5581
LP_Exp	Linguistic variables (8)	SVM	0.871	0.9199	0.7150	0.8343	0.8512

FIGURE 3 ROCs and performance metrics for the literacy profiles relative to expert-rated literacy. AUC: Area Under Curve; LDA: Linear Discriminant Analysis; LP_Exp: Literacy Profile Expert-Rated Health Literacy; LP_FK: Literacy Profile Flesch-Kincaid; LP_LD: Literacy Profile Lexical Diversity; LP_SR: Literacy Profile Self-Reported Health Literacy; LP_WQ: Literacy Profile Writing Quality; ML: Machine Learning; SVM: Support Vector Machine [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Prevalence of sociodemographic characteristics by literacy profile

	Education—College degree %			Race—White %			Age at Survey—Mean (SD)		
Literacy profile	Limited health literacy	Adequate health literacy	P-value	Limited health literacy	Adequate health literacy	P-value	Limited health literacy	Adequate health literacy	P-value
LP_FK	66.3	60.4	.076	22.0	33.7	<.001	56.60 (11.4)	57.29 (9.89)	.305
LP_LD	68.9	59.9	.002	21.4	32.6	<.001	56.72 (9.73)	56.74 (10.0)	.966
LP_WQ	60.4	57.4	.070	31.7	34.8	.056	56.71 (9.89)	57.44 (10.2)	.032
LP_SR	72.3	53.7	<.001	23.9	36.5	<.001	58.88 (9.98)	55.74 (9.74)	<.001
LP_Exp	71.2	57.4	<.001	23.1	33.1	<.001	56.70 (10.2)	57.80 (10.0)	<.001

Abbreviations: LP_Exp, Literacy Profile Expert-Rated Health Literacy; LP_FK, Literacy Profile Flesch-Kincaid; LP_LD, Literacy Profile Lexical Diversity; LP_SR, Literacy Profile Self-Reported Health Literacy; LP_WQ, Literacy Profile Writing Quality; SD, Standard Deviation.

TABLE 3 Associations between five literacy profiles and single-item CAHPS ratings of poor physician communication, diabetes-related health outcomes (%), and annual health care service utilization—mean visits (SD)

Health outcomes	Literacy profile	LP_FK	LP_LD	LP_WQ	LP_SR	LP_Exp
Poor Physician Communication (%)	Limited health literacy	12.2	10.6	9.2	13.8	15.5
	Adequate health literacy	8.8	9.6	10.7	7.3	11.3
	P-value	.0919	.5610	.1372	<.001	<.001
Poor medication adherence (%)	Limited health literacy	29.0	26.6	23.9	25.6	27.9
	Adequate health literacy	22.9	23.8	25.3	23.4	22.9
	P-value	.043	.277	.364	.047	<.001
≥1 Severe Hypoglycemia (%)	Limited health literacy	8.9	4.3	2.9	5.1	4.3
	Adequate health literacy	3.1	3.3	3.9	2.0	3.4
	P-value	<.001	.318	.119	<.001	.02
HbA1c ≤ 7%	Limited health literacy	40.2	44.8	47.4	45.9	43.3
	Adequate health literacy	48.8	48.5	45.2	47.7	47.4
	P-value	.011	.202	.193	.141	<.001
HbA1c ≥ 9%	Limited health literacy	19.1	13.3	13.8	14.6	16.4
	Adequate health literacy	13.7	13.5	14.6	13.5	13.0
	P-value	.02	.91	.499	.24	<.001
Charlson Index	Limited health literacy	2.61 (1.84)	2.36 (1.69)	2.20 (1.61)	2.65 (1.91)	2.42 (1.79)
	Adequate health literacy	2.28 (1.68)	2.31 (1.69)	2.40 (1.72)	2.02 (1.41)	2.32 (1.70)
	P-value	.004	.636	<.001	<.001	.006
Outpatient clinic visits	Limited health literacy	9.10 (7.37)	9.01 (8.98)	9.45 (9.75)	10.29 (10.7)	9.83 (10.6)
	Adequate health literacy	9.53 (9.33)	9.61 (10.3)	9.42 (9.53)	9.01 (9.16)	9.68 (9.57)
	P-value	.479	.301	.931	<.001	.499
ED visits	Limited health literacy	0.48 (1.00)	0.47 (1.15)	0.38 (0.94)	0.53 (1.20)	0.47 (1.14)
	Adequate health literacy	0.39 (0.94)	0.38 (0.88)	0.43 (0.96)	0.31 (0.76)	0.42 (1.01)
	P-value	.170	.102	.085	<.001	.016
Hospitalization	Limited health literacy	0.23 (0.71)	0.21 (0.61)	0.17 (0.60)	0.25 (0.73)	0.20 (0.65)
	Adequate health literacy	0.18 (0.62)	0.19 (0.65)	0.19 (0.67)	0.13 (0.54)	0.20 (0.67)
	P-value	.243	.604	.503	<.001	.713

Abbreviations: Exp, Expert-Rated; FK, Flesch-Kincaid; LD, Lexical Diversity; LP, Literacy Profile; SD, Standard Deviation; SR, Self-Reported; WQ, Writing Quality.

3.2.4 | Health care utilization

Utilizations rates associated with each of the five literacy profiles are given in Table 3. For those classified as having limited health literacy, LP_SR was the only model that associated inadequate health literacy with higher rates of outpatient visits and hospitalizations. Higher annual emergency room utilization rates were observed for limited health literacy when assessed by both LP_SR

and LP_Exp, with health literacy-related differences more robust for LP_SR.

4 | DISCUSSION

Generating accurate information on a population's health literacy, or on an individual patient's health literacy, through the use of an automated literacy profile efficiently provides new avenues that can both inform health services research as well as improve health services delivery and population management. The main added value of our approach is that it supplants the requirement to assess patients' health literacy one at a time; any effort required to operationalize our system provides tremendous economies of scale. As such, a scalable, automated measure of health literacy has the potential to enable health systems to (a) efficiently determine whether quality of care and health outcomes vary by patient health literacy; (b) identify populations and/or individual patients at risk of miscommunication so as to better target and deliver tailored health communications and self-management support interventions; and (c) inform clinicians so as to promote improvements in individual-level care. A 2012 report from the National Academy of Medicine called for health systems to measure the extent to which quality and outcomes differ across patient health literacy level so that systems can take steps to reduce such disparities and track the success of these quality improvement efforts.[65] However, to date, no measure of health literacy has been available to enable such comparisons. In addition, prior research has shown that delivering health literacy-appropriate communication interventions can disproportionately benefit those with limited health literacy skills or narrow extant health literacy-related disparities in such common conditions such as diabetes, heart failure, asthma, and end-of-life care.[9,13,66-69] But translation of this work into real-world settings has been hampered, in part, by the inability to efficiently scale the identification of limited health literacy so as to facilitate targeting those most in need. Health systems are increasingly interested in incorporating predictive analytics as a means of risk stratifying and targeting care. Harnessing "big (linguistic) data" by using natural language processing and machine learning approaches to classify levels of health literacy 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 health literacy-related disparities.[6] Finally, prior studies have demonstrated that clinicians often overestimate the health literacy status of their patients.[12] However, when their patients have been screened for health literacy, primary care physicians have been shown to be receptive to this information and, once they have learned that a patient has limited health literacy, physicians have been shown to engage in a range of communication behaviors that can promote better comprehension and adherence. The translational implications of the research on physician behavior have been limited due, in part, to the lack of efficient and scalable measures of health literacy, as well as physicians' reports that in order for them to best respond, they would need additional system-level support.

The current study compared the performance of five literacy profiles generated from linguistic features extracted using natural language processing and trained using machine learning techniques. While natural language processing and machine learning tools have previously been employed in a variety of health care research applications,[70-80] our research is one of the first to attempt to do so to classify patients' health literacy. We determined that, by applying innovative computational linguistics approaches we were able to generate two automated literacy profiles that (a) have sufficient accuracy in classifying levels of either self-reported health literacy (LP_SR) or expert-rated health literacy (LP_Exp), and (b) reveal confirmatory patterns with sociodemographic, communication and health variables previously shown to be associated with health literacy. The findings that LP_SR and LP-Exp were weakly correlated suggest that these measures reflect different aspects of the broader construct of health literacy.

Several limitations should be noted. First, while our patient sample was large and diverse, and we studied a very large number of patients and secure messages, we only were able to analyze those patients who had engaged in secure message with their physicians, likely excluding patients with severe health literacy limitations. However, in a related analysis (previously unpublished data), we found that patients with limited health literacy are accelerating in their use of patient portals and secure messaging relative to those with adequate HL. Based on the DISTANCE cohort, between 2006 and 2015, the proportion of those with inadequate health literacy who used the portal to engage in two or more secure message threads increased nearly ten-fold (from 6% to 57%), as compared to -a five-fold increase among those with adequate health literacy (13% to 74%). By 2018, 99 percent of those with both limited and adequate health literacy who had registered for the portal had used the portal for secure messaging. Furthermore, we found no significant health literacy-related differences in exclusions at the patient or secure message level as shown in the flow diagram in Figure 1.

Second, the setting in which we carried out this research raises questions about external validity. While limited health literacy is more concentrated in safety net health care settings, it is still common in this fully insured population. KPNC has a sizable Medicaid population and over 1/3 of their diabetes patients have limited health literacy.[21,59] From an internal validity standpoint, this setting provided access to a mature patient portal and availability of extensive linguistic and health-related data, and the fully integrated care and closed pharmacy system of KPNC ensured complete capture of health care utilization and medication refills. Relatedly, we excluded proxy secure messages (ie, those written by another individual on behalf of the patient) to enhance accuracy and limited the study to secure messages written in English. Third, the single items CAHPS measure is a subjective measure of provider communication and is subject to recall bias similar to that of self-reported health literacy, potentially over- or underestimating the strength of the association between LP_SR and provider communication. Fourth, although our literacy profiles were trained on self-reported health literacy and expert-rated health literacy, the absence of a universally accepted, comprehensive, "true" gold standard for health literacy, and the fact

that we used linguistic indices validated before email exchange became so prevalent, may limit our categorization of health literacy. Finally, additional research in other settings and patient populations with different conditions may provide a more definitive answer as to the optimal literacy profile to use in classifying health literacy. Our current work to develop and evaluate a measure of discordance in secure message exchange that takes into account both patients' and physicians' linguistic complexity may provide further insights.

The ECLIPPSE Project set out to harness secure messages sent by diabetes patients to their primary care physician(s) to develop literacy profiles that can identify patients with limited health literacy in an automated way that avoids time-consuming and potentially sensitive questioning of the patient. Given the time and personnel demands intrinsic to current health literacy instruments, measuring health literacy has historically been extremely challenging. An automated literacy profile could provide an efficient means to identify subpopulations of patients with limited health literacy. Identifying patients likely to have limited health literacy could prove useful for alerting clinicians about potential difficulties in comprehending written and/or verbal instructions. Additionally, patients identified as having limited health literacy could be supported better by receiving follow-up communications to ensure understanding of critical communications, such as new medication instructions, and promote adherence and increased shared meaning.[13] As such, our research to develop automated methods for health literacy assessment represents a significant accomplishment with potentially broad clinical and population health benefits in the context of health services delivery.

However, there may be privacy and ethical issues that researchers and health systems planners need to consider before employing the automated literacy profiles for a new generation of health literacy research, or for scaling health system and clinical applications to reduce health literacy-related disparities. To generate literacy profiles, patients' own written words are harnessed, raising potential concerns about confidentiality. Further, having one's own linguistic data analyzed to estimate individual-level health literacy in the absence of an explicit consent process may be perceived as problematic given the prior literature on literacy screening and stigma.[81,82] The fact that electronic health data—both clinical and administrative—are commonly used by patients' health systems to identify populations and individuals at risk and to target associated interventions, both at the clinician-patient dyad and health system-population levels, suggests that efforts to introduce the literacy profile methodology could be met with acceptance, based on these precedents. Further, the fact that literacy profiles are generated using computational linguistic methods that require no human engagement with the actual content of the messages can provide additional reassurance to the public. Nevertheless, health systems interested in employing literacy profiles should consider adding linguistic data to the patient-related electronic health data for which they already obtain blanket, advanced informed consent. In addition, researchers and health systems should develop policy guidance that permits usage of the literacy profiles that promote population health and reduce health literacy-related disparities while not undermining patient well-being, as

well as practical guidance for clinicians as to how they might use the literacy profile in patient-centered and sensitive ways.[65] Developing this guidance would benefit from the inclusion of, and input from, advisory members who have limited health literacy skills.

In summary, the two top-performing literacy profiles (LP_SR, LP_Exp) revealed associations consistent with previous health literacy research across a range of outcomes related to quality, safety, comorbidity, and utilization. Future implementation and dissemination research is needed. This research should include evaluating the transportability of our approach to deriving literacy profiles from patients' secure messages to diverse health care delivery settings, the development of provider workflow and/or novel population management approaches when patients with limited health literacy are identified, and the effects of interventions that harness this novel source of information on health-related outcomes. We conclude that applying innovative NLP and machine learning approaches[14,15,51-56,83] to generate literacy profiles from patients' secure messages is a novel, feasible, automated, and scalable strategy to identify patients and subpopulations with limited health literacy, thus providing a health IT tool to enable tailored communication support and other targeted interventions with the potential to reduce health literacy-related disparities.

ORCID

Renu Balyan https://orcid.org/0000-0003-1393-2416
Andrew J. Karter https://orcid.org/0000-0001-5527-316X

REFERENCES

1. Stewart MA. Effective physician-patient communication and health outcomes: a review. *Can Med Assoc J.* 1995;152(9):1423.
2. Ratanawongsa N, Karter AJ, Parker MM, et al. Communication and medication refill adherence: the Diabetes Study of Northern California. *JAMA Intern Med.* 2013;173(3):210-218.
3. Centers for Disease Control and Prevention. *Diabetes Report Card 2017.* Atlanta, GA: Centers for Disease Control and Prevention, US Dept of Health and Human Services; 2018.
4. Bailey SC, Brega AG, Crutchfield TM, et al. Update on health literacy and diabetes. *Diabetes Educator.* 2014;40(5):581-604.
5. Bauer AM, Schillinger D, Parker MM, et al. Health literacy and antidepressant medication adherence among adults with diabetes: the diabetes study of Northern California (DISTANCE). *J Gen Intern Med.* 2013;28(9):1181-1187.
6. Karter AJ, Parker MM, Duru OK, et al. Impact of a pharmacy benefit change on new use of mail order pharmacy among diabetes patients: the Diabetes Study of Northern California (DISTANCE). *Health Serv Res.* 2015;50(2):537-559.
7. Schillinger D, Bindman A, Wang F, Stewart A, Piette J. Functional health literacy and the quality of physician–patient communication among diabetes patients. *Patient Educ Couns.* 2004;52(3):315-323.

8. Castro CM, Wilson C, Wang F, Schillinger D. Babel babble: physicians' use of unclarified medical jargon with patients. *Am J Health Behavior*. 2007;31(1):S85-S95.

9. Schillinger D, Piette J, Grumbach K, et al. Closing the loop: physician communication with diabetic patients who have low health literacy. *Arch Intern Med*. 2003;163(1):83-90.

10. Sarkar U, Piette JD, Gonzales R, et al. Preferences for self-management support: findings from a survey of diabetes patients in safety-net health systems. *Patient Educ Couns*. 2008; 70(1):102-110.

11. Schillinger D, McNamara D, Crossley S, et al. The Next Frontier in Communication and the ECLIPPSE Study: bridging the linguistic divide in secure messaging. *J Diabetes Res*. 2017;2017:1348242.

12. Seligman HK, Wang FF, Palacios JL, et al. Physician notification of their diabetes patients' limited health literacy: a randomized, controlled trial. *J Gen Intern Med*. 2005;20(11):1001-1007.

13. DeWalt DA, Baker DW, Schillinger D, et al. A multisite randomized trial of a single-versus multi-session literacy sensitive self-care intervention for patients with heart failure. *J General Int Med*. 2011;26:S57-S58.

14. Balyan R, Crossley SA, Brown W III, et al. Using natural language processing and machine learning to classify health literacy from secure messages: The ECLIPPSE study. *PLoS One*. 2019;14(2):e0212488.

15. Crossley SA, Balyan R, Liu J, Karter AJ, McNamara D, Schillinger D. Developing and testing automatic models of patient communicative health literacy using linguistic features: findings from the ECLIPPSE study. *Health Communication*. 2020;2:1.

16. Chew LD, Griffin JM, Partin MR, et al. Validation of screening questions for limited health literacy in a large VA outpatient population. *J Gen Intern Med*. 2008;23(5):561-566.

17. Moffet HH, Adler N, Schillinger D, et al. Cohort Profile: The Diabetes Study of Northern California (DISTANCE)—objectives and design of a survey follow-up study of social health disparities in a managed care population. *Int J Epidemiol*. 2008;38(1):38-47.

18. Semere W, Crossley S, Karter AJ, et al. Secure messaging with physicians by proxies for patients with diabetes: findings from the ECLIPPSE Study. *J Gen Intern Med*. 2019;19:1-7.

19. Crossley S, Kostyuk V. Letting the Genie out of the Lamp: using natural language processing tools to predict math performance. In *International Conference on Language, Data and Knowledge 2017 Jun 19*. Cham: Springer; 2017:330-342.

20. Crossley S, Paquette L, Dascalu M, McNamara DS, Baker RS. Combining click-stream data with NLP tools to better understand MOOC completion. In *Proceedings of the sixth international conference on learning analytics & knowledge 2016 Apr 25*. ACM; 2016:6-14.

21. Sarkar U, Schillinger D, López A, Sudore R. Validation of self-reported health literacy questions among diverse English and Spanish-speaking populations. *J Gen Intern Med*. 2011;26(3):265-271.

22. Crossley SA, Kyle K, McNamara DS. To aggregate or not? Linguistic features in automatic essay scoring and feedback systems. *Grantee Submission*. 2015;8(1).

23. Kyle K, Crossley SA. Automatically assessing lexical sophistication: Indices, tools, findings, and application. *Tesol Quarter*. 2015;49(4):757-786.

24. Kyle K, Crossley S, Berger C. The tool for the automatic analysis of lexical sophistication (TAALES): version 2.0. *Behav Res Methods*. 2018;50(3):1030-1046.

25. Crossley SA, Kyle K, McNamara DS. The tool for the automatic analysis of text cohesion (TAACO): automatic assessment of local, global, and text cohesion. *Behav Res Methods*. 2016;48(4):1227-1237.

26. Kyle K. Measuring syntactic development in L2 writing: fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication.

27. Crossley SA, Skalicky S, Dascalu M, et al. Predicting text comprehension, processing, and familiarity in adult readers: new approaches to readability formulas. *Discourse Process*. 2017;54(5-6):340-359.

28. Crossley SA, Kyle K, McNamara DS. Sentiment Analysis and Social Cognition Engine (SEANCE): an automatic tool for sentiment, social cognition, and social-order analysis. *Behav Res Methods*. 2017;49(3):803-821.

29. McNamara DS, Graesser AC, McCarthy PM, Cai Z. *Automated Evaluation of Text and Discourse with Coh-Metrix*. Cambridge, UK: Cambridge University Press; 2014.

30. De Marneffe MC, MacCartney B, Manning CD. Generating typed dependency parses from phrase structure parses. In *Proceedings of LREC 2006 May 28*, vol. 6, No. 2006, 2006:449-454.

31. BNC Consortium. *The British National Corpus, version 2 (BNC world)*. Distributed by Oxford University Computing Services; 2001.

32. Coltheart M. The MRC psycholinguistic database. *Quarter J Exp Psychol*. 1981;33(4):497-505.

33. Baayen RH, Piepenbrock R, Gulikers L. *The CELEX lexical database (release 2)*. Distributed by the Linguistic Data Consortium, University of Pennsylvania; 1995.

34. Miller GA. WordNet: a lexical database for English. *Commun ACM*. 1995;38(11):39-41.

35. Thompson P, Batista-Navarro RT, Kontonatsios G, et al. Text mining the history of medicine. *PLoS One*. 2016;11(1):e0144717.

36. Uzuner Ö, Juo Y, Szolovits P. Evaluating the state-of-the-art in automatic de-identification. *J Am Med Inform Assoc*. 2007;14(5):550-563.

37. Uzuner Ö, Goldstein I, Luo Y, Kohane I. Identifying patient smoking status from medical discharge records. *J Am Med Inform Assoc*. 2008;15(1):15-24.

38. Uzuner Ö. Recognizing obesity and comorbidities in sparse data. *J Am Med Inform Assoc*. 2009;16(4):561-570.

39. Uzuner Ö, Solti I, Cadag E. Extracting medication information from clinical text. *J Am Med Inform Assoc*. 2010;17(5):514-518.

40. Flesch R. A new readability yardstick. *J Appl Psychol*. 1948;32(3):221.

41. Jindal P, MacDermid J. Assessing reading levels of health information: uses and limitations of flesch formula. *Education for Health*. 2017;30(1):84.

42. Munsour EE, Awaisu A, Hassali MA, Darwish S, Abdoun E. Readability and comprehensibility of patient information leaflets for Antidiabetic medications in Qatar. *J Pharm Technol*. 2017;33(4):128-136.

43. Paasche-Orlow MK, Taylor HA, Brancati FL. Readability standards for informed-consent forms as compared with actual readability. *N Engl J Med*. 2003;348(8):721-726.

44. Piñero-López MÁ, Modamio P, Lastra CF, Mariño EL. Readability analysis of the package leaflets for biological medicines available on the internet between 2007 and 2013: an analytical longitudinal study. *J Med Internet Res*. 2016;18(5):e100.

45. Wilson M. Readability and patient education materials used for low-income populations. *Clin Nurse Spec*. 2009;23(1):33-40.

46. Zheng J, Yu H. Assessing the readability of medical documents: a ranking approach. *JMIR Med Inform*. 2018;6(1):e17.

47. Malvern DD, Richards BJ, Chipere N, Durán P. *Lexical Diversity and Language Development: Quantification and Assessment*. Houndmills: Palgrave Macmillan; 2004.

48. McCarthy PM. An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD). Dissertation Abstracts International, 66, UMI No. 3199485; 2005.

49. McNamara DS, Crossley SA, Roscoe RD, Allen LK, Dai J. A hierarchical classification approach to automated essay scoring. *Assessing Writing*. 2015;1(23):35-59.

50. McNamara DS, Crossley SA, McCarthy PM. Linguistic features of writing quality. *Written Commun*. 2010;27(1):57-86.

51. Balyan R, McCarthy KS, McNamara DS. Combining Machine Learning and Natural Language Processing to Assess Literary Text Comprehension. In Hershkovitz A, Paquette L, editors. *Proceedings of the 10th International Conference on Educational Data Mining*

(EDM), Wuhan, China. International Educational Data Mining Society; 2017.

52. Han J, Pei J, Kamber M. *Data Mining: Concepts and Techniques*. Amsterdam, The Netherlands: Elsevier; 2011.

53. Joachims T. Text categorization with support vector machines: Learning with many relevant features. In *European Conference on Machine Learning 1998 Apr 21*. Berlin, Heidelberg: Springer; 1998:137-142.

54. Mitchell TM. *Machine Learning*. Burr Ridge, IL: McGraw Hill. 1997;45(37):870-877.

55. Schölkopf B, Smola AJ. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA: MIT Press; 2002.

56. Balyan R, McCarthy KS, McNamara DS.Comparing machine learning classification approaches for predicting expository text difficulty. In The Thirty-First International Flairs Conference 2018 May 10.

57. Steiner JF, Koepsell TD, Fihn SD, Inui TS. A general method of compliance assessment using centralized pharmacy records: description and validation. *Med Care*. 1988;1:814-823.

58. Steiner JF, Prochazka AV. The assessment of refill compliance using pharmacy records: methods, validity, and applications. *J Clin Epidemiol*. 1997;50(1):105-116.

59. Sarkar U, Karter AJ, Liu JY, Moffet HH, Adler NE, Schillinger D. Hypoglycemia is more common among type 2 diabetes patients with limited health literacy: the Diabetes Study of Northern California (DISTANCE). *J Gen Intern Med*. 2010;25(9):962-968.

60. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373-383.

61. Charlson M, Szatrowski TP, Peterson J, Gold J. Validation of a combined comorbidity index. *J Clin Epidemiol*. 1994;47(11):1245-1251.

62. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol*. 1992;45(6):613-619.

63. Ginde AA, Blanc PG, Lieberman RM, Camargo CA. Validation of ICD-9-CM coding algorithm for improved identification of hypoglycemia visits. *BMC Endocrine Disorders*. 2008;8(1):4.

64. Raebel MA, Schmittdiel J, Karter AJ, Konieczny JL, Steiner JF. Standardizing terminology and definitions of medication adherence and persistence in research employing electronic databases. *Med Care*. 2013;51:S11-S21.

65. Brach C, Keller D, Hernandez LM, et al. *Ten Attributes of Health Literate Health Care Organizations*. NAM Perspectives. Discussion Paper, National Academy of Medicine, Washington, DC; 2012.

66. Sheridan SL, Halpern DJ, Viera AJ, Berkman ND, Donahue KE, Crotty K. Interventions for individuals with low health literacy: a systematic review. *J Health Commun*. 2011;16(Supp. 3):30-54.

67. Machtinger EL, Wang F, Chen LL, Rodriguez M, Wu S, Schillinger D. A visual medication schedule to improve anticoagulation control: a randomized, controlled trial. *Joint Commission J Qual Patient Saf*. 2007;33(10):625-635.

68. Paasche-Orlow MK, Riekert KA, Bilderback A, et al. Tailored education may reduce health literacy disparities in asthma self-management. *Am J Respir Crit Care Med*. 2005;172(8):980-986.

69. Sudore RL, Schillinger D, Katen MT, et al. Engaging diverse English-and Spanish-speaking older adults in advance care planning: the PREPARE randomized clinical trial. *JAMA Intern Med*. 2018;178(12):1616-1625.

70. Friedman C, Johnson SB, Forman B, Starren J. Architectural requirements for a multipurpose natural language processor in the clinical environment. In *Proceedings of the Annual Symposium on Computer Application in Medical Care 1995*. American Medical Informatics Association:347.

71. Savova GK, Masanz JJ, Ogren PV, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *J Am Med Inform Assoc*. 2010;17(5):507-513.

72. Soysal E, Wang J, Jiang M, et al. CLAMP–a toolkit for efficiently building customized clinical natural language processing pipelines. *J Am Med Inform Assoc*. 2017;25(3):331-336.

73. Denny JC, Irani PR, Wehbe FH, Smithers JD, Spickard A III. The KnowledgeMap project: development of a concept-based medical school curriculum database. In *AMIA Annual Symposium Proceedings 2003*, vol. 2003. American Medical Informatics Association:195.

74. Aronson AR, Lang FM. An overview of MetaMap: historical perspective and recent advances. *J Am Med Inform Assoc*. 2010;17(3):229-236.

75. Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic Acids Res*. 2004;32:D267-D270.

76. Thorn CF, Klein TE, Altman RB. PharmGKB: the pharmacogenomics knowledge base. In *Pharmacogenomics 2013*. Totowa, NJ: Humana Press; 2013:311-320.

77. Van Gurp M, Decoene M, Holvoet M, dos Santos MC.LinKBase, a Philosophically-Inspired Ontology for NLP/NLU Applications. In KR-MED 2006 Nov 8.

78. Smith B, Fellbaum C. Medical WordNet: a new methodology for the construction and validation of information resources for consumer health. In *Proceedings of the 20th International Conference on Computational Linguistics 2004 Aug 23*. Association for Computational Linguistics:371.

79. Nobata C, Cotter P, Okazaki N, et al. Kleio: a knowledge-enriched information retrieval system for biology. In *Proceedings of the 31st Annual International ACM SIGIR Conference on RESEARCH and Development in Information Retrieval 2008 Jul 20*. ACM: 787-788.

80. Tsuruoka Y, Tsujii JI, Ananiadou S. FACTA: a text search engine for finding associated biomedical concepts. *Bioinformatics*. 2008;24(21):2559-2560.

81. Easton P, Entwistle VA, Williams B. How the stigma of low literacy can impair patient-professional spoken interactions and affect health: insights from a qualitative investigation. *BMC Health Serv Res*. 2013;13(1):319.

82. Mackert M, Donovan EE, Mabry A, Guadagno M, Stout PA. Stigma and health literacy: an agenda for advancing research and practice. *Am J Health Behav*. 2014;38(5):690-698.

83. Balyan R, McCarthy KS, McNamara DS. Applying natural language processing and hierarchical machine learning approaches to text difficulty classification. *International Journal of Artificial Intelligence in Education*. 2020;1–34. https://doi.org/10.1007/s40593-020-00201-7

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.