

Supplementary Materials for

Precision communication: Physicians' linguistic adaptation to patients' health literacy

Dean Schillinger*, Nicholas D. Duran, Danielle S. McNamara, Scott A. Crossley,
Renu Balyan, Andrew J. Karter

*Corresponding author. Email: dean.schillinger@ucsf.edu

Published 17 December 2021, *Sci. Adv.* 7, eabj2836 (2021)
DOI: 10.1126/sciadv.abj2836

This PDF file includes:

Table S1

A.			B.		
Physician Complexity Profile Linguistic Indices			Patient Health Literacy Profile Linguistic Indices		
Label (from NLP tool)	Category	High CP	Label (from NLP tool)	Category	High HL
Argument overlap, sentences	Cohesion	-	Argument overlap, paragraph	Cohesion	+
Lexical diversity (D)	Lexical diversity	+	Attested lemmas (COCA magazine)	Lexical diversity	+
Word range score, CW, SUBTLEXus	Lexical sophistication	-	Lexical decision response time (CW)	Lexical sophistication	+
Bigram association strength (delta p), COCA academic	Lexical sophistication	+	Number of associations (EAT, FW)	Lexical sophistication	+
Trigram frequency, COCA newspaper	Lexical sophistication	+	Phonological Neighborhood Frequency (FW)	Lexical sophistication	+
Trigram range, COCA fiction	Lexical sophistication	+	Determiner per nominal phrase	Syntax	+
Word concreteness, AW	Lexical sophistication	-	Passive auxiliary verbs per clause	Syntax	+
Average number of direct object dependencies	Syntax	+	Dependents per nominal subject	Syntax	+
Construction frequency, SD, COCA fiction	Syntax	+	Adjective complements per clause	Syntax	+
Terms related to action	Cognition	-	NOTES: CW=Content words, EAT = Edinburgh association thesaurus, FW=Function words		
Incidence of words related to objects	Cognition	-			
Fear and disgust words	Sentiment	+			
Words related to friends and family	Sentiment	-			
Words related to joy	Sentiment	-			

NOTES: CW=Content words, AW=All words, SD=Standard deviation

Table S1. Linguistic indices used to generate the patient health literacy (HL) measure and physician complexity profile (CP). Provided is the name of each index as labeled in the natural language processing (NLP) tool used to compute the index. These indices correspond to broad language categories of cohesion, lexical diversity, lexical sophistication, syntax, and other factors (e.g., cognition, sentiment). The table also

indicates whether each index was more prevalent (positive sign, +) or less prevalent (negative sign, -) in **A**: high CP texts relative to low CP texts and **B**: high HL texts relative to low HL texts. The concordance between CP and HL does not depend on the presence or direction of any single index or small set of indices. Concordance is based on the high-level emergent representations of CP and HL, derived from the totality of indices and their interactions in machine learning models trained to predict expert ratings.