


## Clinical Documentation

# Flesch-Kincaid Grade Level Readability Scores to Evaluate Readability of Clinical Documentation During an Electronic Health Record Transition

Douglas W Challener, MD, MS<sup>1</sup> , Andrew Wen, MS<sup>2</sup>, Jungwei W. Fan, PhD<sup>2</sup>, Hongfang Liu, PhD<sup>2</sup>, John O'Horo, MD, MPH<sup>1,3</sup>, Mark Nyman, MD<sup>4</sup>

<sup>1</sup> Public Health, Infectious Diseases, and Occupational Medicine, Mayo Clinic, <sup>2</sup> Artificial Intelligence and Informatics, Mayo Clinic, <sup>3</sup> Pulmonary and Critical Care Medicine, Mayo Clinic, <sup>4</sup> General Internal Medicine, Mayo Clinic

Keywords: readability, flesch-kincaid grade level, clinical documentation, EHR, clinical notes

<https://doi.org/10.63116/VBWY7913>

---

## Advances in Health Information Science and Practice

Vol. 1, Issue 1, 2025

---

### Background

Electronic health records (EHRs) serve as a critical tool for clinical documentation, yet their complexity has raised concerns about readability and usability for healthcare providers and patients. Readability scores, such as the Flesch-Kincaid (FK) Grade Level, offer a quantitative method to assess documentation clarity. In this study, the authors evaluated the impact of an EHR transition on the readability of clinical notes across multiple medical specialties.

### Methods

All clinical notes recorded between 2016 and 2019 at a large healthcare organization were evaluated across specialties. The organization transitioned EHR systems during the study period, in May 2018. FK Grade Level scores were calculated for each clinical consultation note and stratified by specialty and EHR.

### Results

Across 10 medical specialties, 630,246 clinical notes were evaluated. The median FK grade levels from notes composed in the new EHR were similar to those composed in the previous EHR (8.228 vs 8.240;  $p=0.58$ ). The variation in FK grade level as measured by standard deviation was lower in the new EHR compared with the old EHR (1.405 vs 1.764;  $p < 0.01$ ).

### Conclusions

Calculation of FK grade level of clinical notes is a feasible method to assess clinical note readability. Study results showed that clinical documentation readability may not be strongly associated with underlying EHR. Migrating to a new EHR alone should not be expected to improve readability of clinical notes.

## INTRODUCTION

Electronic health records (EHRs) serve as a unifying source of data for clinical care, research, administrative work, billing, and compliance. Clinical documentation in the EHR has become increasingly complex and challenging to read. Recognizing this problem, in 2015 the American College of Physicians convened their Medical Informatics Committee to publish a position paper on the goals of electronic health record (EHR) generated documentation. The committee stated, "The primary goal of EHR-generated documentation should be concise, history-rich notes that reflect the information gathered and are used to develop an impression, a diagnostic and/or treatment plan, and recom-

mended follow-up."<sup>1</sup> However, despite efforts to reinvent clinical documentation for the modern era, much remains unchanged from the time of paper charts.<sup>2</sup> Additionally, the digitization of health records has created a tension between structured data formats, which are more easily interpreted by computers, and flexible documentation formats, which may increase readability.<sup>3</sup> In this way, digital health records can be more difficult for a human to read than the paper charts they replaced. Integrating artificial intelligence to augment and support clinical documentation processes may relieve some of this burden, but this goal has not yet been fully achieved.<sup>4</sup> The competing priorities of the EHR lead to clinical documentation that is extensive and challenging for human end users to understand, a phenome-

non known as “note bloat.”<sup>5</sup> In attempting to serve many roles, clinical documentation often ends up serving none optimally.

For physicians, the record must document the patient’s presenting signs and symptoms along with diagnostic test results and clinical reasoning, ultimately leading to a diagnosis. For administrators, the clinical record must adequately document the level of patient care provided, to allow for appropriate financial reimbursement and compliance. As patient access to EHR portals has become more common (and even, in some cases, mandated by legislation), it has become increasingly important that personal medical records become more accessible, comprehensible, and portable, so that patients can effectively understand and coordinate their healthcare across multiple providers.<sup>6</sup> Metrics for evaluating how well clinical documentation can be understood by these stakeholders are lacking. Often, note length is used as a surrogate to detect “note bloat,” but the limitations of this as a single measure are self-evident.

An additional method of comprehension evaluation is a readability score; these scores evaluate texts to estimate the ease with which they can be understood by a reader. Readability can be assessed using multiple scales. A popular scale is the Flesch-Kincaid (FK) grade level, which was developed for the United States Armed Forces in the 1970s and has been used to judge the readability of many documents.<sup>7,8</sup> The FK grade level is calculated based on and is directly proportional to the ratio of total words to total sentences and the ratio of total syllables to total words; the result is a number indicating the grade level that would be required for a reader to understand the content. Therefore, a document with longer sentences comprised of words with more syllables will have a higher grade level score compared with a document with short sentences of monosyllabic words. A FK grade level of 8 or below is generally considered to be best practice for communication to the general public. In medicine, the FK grade level has been used to grade patient educational handouts and websites, dismissal summaries, questionnaires, and patient consent forms.<sup>9-16</sup> However, there has been little research into the readability of clinical documentation.

While not sparking a revolution in concise, easily readable, clinical documentation, the EHR can easily facilitate evaluation of clinical notes and related quality improvement initiatives. For example, in one study, clinical alerts were shown to improve compliance of anesthesia documentation.<sup>17</sup> Another study evaluated the “attractiveness” of clinical notes in an EHR.<sup>18</sup> Additionally, quality of clinical documentation was evaluated in one study using a standardized scale, the Physician Documentation Quality Instrument (PDQI).<sup>19</sup> Previous assessments of note quality largely required human adjudication. In this study, we evaluated the feasibility and utility of using the FK score as an automated tool for assessing clinical documentation quality and its association with the underlying electronic health record (EHR) software.

## METHODS

All clinical notes recorded between 2016 and 2019 at a large healthcare organization were evaluated across specialties. In May 2018, the organization transitioned from a highly-customized, “homegrown” EHR to a widely-used, commercial EHR. Notes were classified as either being created in the old or the new EHR. To limit variability, only labeled clinical consultation notes were included in the analysis; this note type was consistent between the two EHR systems. A custom model was used to tokenize and perform sentence splitting on each input document, yielding number of sentences and words per document.<sup>20</sup> Syllables in each word were calculated using a custom Python script that splits along vowels with special handling for multi-vowel syllables.<sup>21</sup> FK grade level was calculated<sup>7</sup> for every clinical note and summarized by author specialty. Summary statistics including minimum, maximum, 25<sup>th</sup> percentile, median, and 75<sup>th</sup> percentile were calculated. A paired T-test was performed to assess the relationship of each specialty’s notes before and after the EHR transition. A two-way analysis of variance (ANOVA) statistical test was performed to assess the relationship between author specialty and timing of authorship with median FK grade level. Two additional readability scores were calculated for comparison (Gunning Fog Index,<sup>22</sup> Dale-Chall readability<sup>23</sup>) and reported in the supplementary materials.

## RESULTS

A total of 630,246 clinical notes were evaluated across 10 medical specialties including anesthesiology, cardiology, family medicine, gastroenterology, general internal medicine, neurology, obstetrics-gynecology, pediatrics, psychiatry, and pulmonology (Table 1). The evaluated notes mostly came from cardiology (177,256; 28.1%), neurology (159,702; 25.3%), and gastroenterology practices (96,448; 15.3%). Anesthesia had the fewest notes evaluated (1,588; 0.3%). Note volume was relatively evenly split between the old EHR (294,625; 46.7%) and the new EHR (335,621; 53.2%).

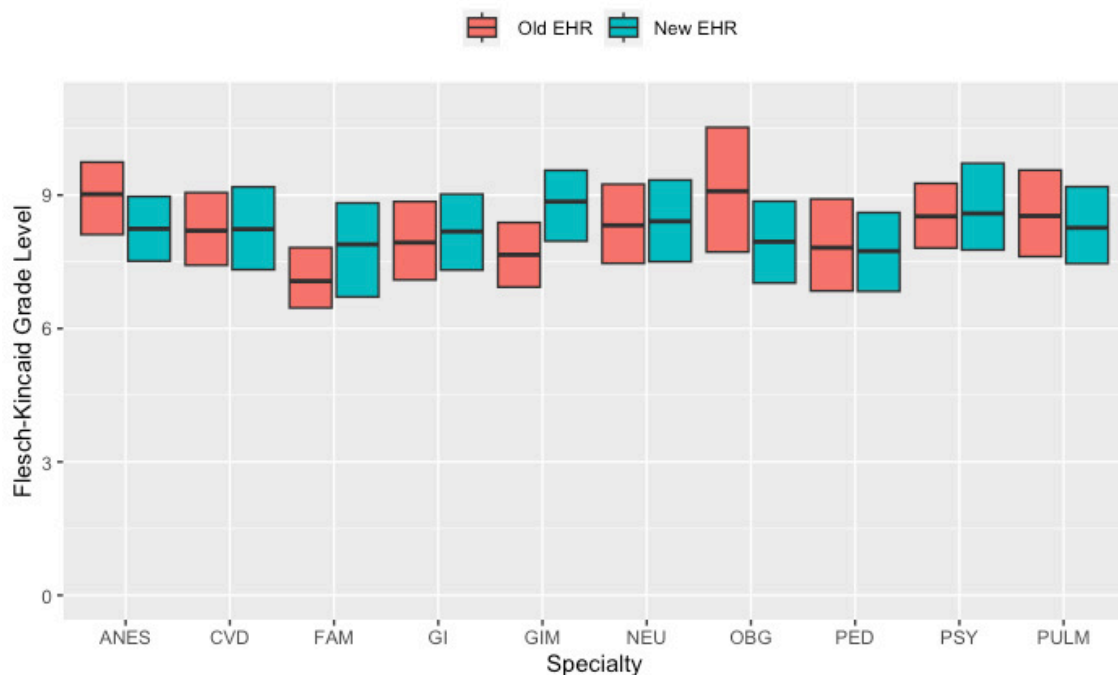
Overall, general internal medicine consultation notes had the highest median FK grade level (8.62) over the four-year study period. This was followed by the other specialties with median FK grade levels varying from 8.55 (psychiatry) to 7.72 (family medicine). Several specialties had a decrease in median FK grade levels following the EHR transition (anesthesiology, obstetrics and gynecology, pediatrics, pulmonology), though overall readability remained fairly stable. Figure 1 illustrates the changes in median, 25<sup>th</sup>, and 75<sup>th</sup> percentile FK grade level by specialty over the study period stratified by EHR.

In matched analysis by specialty, notes written in the new EHR had FK grade levels that were not statistically different those in the previous EHR ( $p=0.93$ ). Matched by specialty, the standard deviation of the FK grade levels of the notes significantly decreased from the previous EHR to the new EHR (mean difference in standard deviation of 0.29,  $p<0.01$ ). All specialties except family medicine had a de-

**Table 1. Summary of Clinical Note Evaluation.**

Specialty	Old EHR			New EHR		
	Total Notes (n = 294625)	Median FK Grade Level	Standard Deviation of FK Grade Level	Total Notes (n = 335621)	Median FK Grade Level	Standard Deviation of FK Grade Level
Anesthesia	465	9.02	1.44	1123	8.24	1.32
Cardiology	84758	8.20	1.71	92498	8.24	1.54
Family Medicine	1664	7.07	1.31	6437	7.89	1.50
Gastroenterology	50918	7.93	1.80	45530	8.18	1.35
General Internal Medicine	3328	7.66	1.74	13855	8.85	1.26
Neurology	72875	8.32	1.72	86827	8.41	1.43
Obstetrics/ Gynecology	3529	9.09	1.93	12919	7.95	1.47
Pediatrics	37350	7.82	1.85	37132	7.74	1.34
Psychiatry	25787	8.52	1.62	20741	8.59	1.46
Pulmonology	13951	8.53	1.86	18559	8.27	1.41

FK=Flesch-Kincaid

**Figure 1. Flesch-Kincaid Grade Level of Consultation Notes by Specialty and Electronic Health Record (EHR).**

ANES (Anesthesiology), CVD (Cardiology), FAM (Family Medicine), GI (Gastroenterology), GIM (General Internal Medicine), NEU (Neurology), OBG (Obstetrics and Gynecology), PED (Pediatrics), PSY (Psychiatry), PULM (Pulmonology)

crease in the standard deviation of their FK grade level note scores.

## DISCUSSION

In this study, the readability of notes written by physicians performing consultations before and after an EHR transition were not significantly affected by the underlying EHR system. There was similar distribution of FK grade level readability scores for clinical notes across specialties on

both the old and new EHRs. This finding should be reassuring and suggests that physicians will continue to write similarly readable notes regardless of the tool used for writing.

While the FK scores of consultation notes remained stable during the transition time, the variability of the scores decreased overall. This suggests that notes in the new EHR were generally more uniform in readability, possibly related to new note composition tools, as the new EHR included improved templating tools and provided easier access to copying components of previously written notes. Both fea-

tures could have contributed to the decrease in variability of FK scores following the EHR transition.

Some worry that increased use of note templating tools worsens the readability of clinical documents.<sup>24</sup> Our results showed that despite significant use of these features in the new EHR, readability scores overall were stable. While the stability of results can be considered comforting, a more desirable outcome would have been improved readability scores with the new EHR.

There are many potential strategies to improve readability of health documentation. Readability may be improved by decreasing stringent requirements, particularly for including structured data that is readily accessible in other areas of the EHR. Documenting vital signs, past medical history, medications administered makes it more challenging for a reader to find key information. Templates should be evaluated for readability principles before deployment. Adopting formal structures and vocabularies focused on describing the rationale behind a patient's care improves note quality and discourages the use of copy and paste.<sup>25</sup> <sup>26</sup> Some have proposed allowing physicians to simply annotate the record rather than repeatedly documenting the same findings.<sup>27</sup>

## LIMITATIONS

This study had several limitations. It did not include human assessment of FK score readability of clinical documentation. There is also some missing data in early 2016 due to a data platform change. Only clinical notes for consultations were considered in this study, which may not have provided a complete evaluation of the medical record. Additionally, the entire text of the clinical notes was evaluated. Limiting the evaluation to just the non-collapsed sections of the

notes may have improved readability scores and more accurately assessed the most-read sections of the note.

## CONCLUSIONS

Results from this study of clinical note readability scores pre- and post-EHR transition showed that clinical documentation readability remained stable and therefore may not be strongly associated with underlying EHR. Migrating to a new EHR alone should not be expected to improve readability of clinical notes. Our study results also suggested that FK grade level or other readability scores may be reliably used in the future evaluations of clinical documentation. The scores can easily be applied in an automated way across an entire health system's EHR. These evaluations could be used to identify individuals or specialties whose clinical notes are significantly different than others. Further work is needed to clarify the relationship between readability and documentation quality.

.....

## DISCLOSURES

The authors have nothing to disclose.

## FUNDING

The authors received no funding for this research.

Submitted: March 04, 2025 EDT. Accepted: April 28, 2025 EDT.

## REFERENCES

1. Kuhn T, Basch P, Barr M, Yackel T, Medical Informatics Committee of the American College of P. Clinical documentation in the 21st century: executive summary of a policy position paper from the American College of Physicians. *Ann Intern Med*. 2015;162(4):301-303. doi:[10.7326/M14-2128](https://doi.org/10.7326/M14-2128)
2. Gillum RF. From papyrus to the electronic tablet: a brief history of the clinical medical record with lessons for the digital age. *Am J Med*. 2013;126(10):853-857. doi:[10.1016/j.amjmed.2013.03.024](https://doi.org/10.1016/j.amjmed.2013.03.024)
3. Rosenbloom ST, Denny JC, Xu H, Lorenzi N, Stead WW, Johnson KB. Data from clinical notes: a perspective on the tension between structure and flexible documentation. *J Am Med Inform Assoc*. 2011;18(2):181-186. doi:[10.1136/jamia.2010.007237](https://doi.org/10.1136/jamia.2010.007237)
4. Lin SY, Shanafelt TD, Asch SM. Reimagining Clinical Documentation With Artificial Intelligence. *Mayo Clin Proc*. 2018;93(5):563-565. doi:[10.1016/j.mayocp.2018.02.016](https://doi.org/10.1016/j.mayocp.2018.02.016)
5. Kahn D, Stewart E, Duncan M, et al. A Prescription for Note Bloat: An Effective Progress Note Template. *Journal of Hospital Medicine*. 2018;13(6):378-382. doi:[10.12788/jhm.2898](https://doi.org/10.12788/jhm.2898)
6. Salmi L, Blease C, Hägglund M, Walker J, Desroches CM. US policy requires immediate release of records to patients. *BMJ*. Published online 2021:n426. doi:[10.1136/bmj.n426](https://doi.org/10.1136/bmj.n426)
7. Kincaid JPF, Robert RP Jr, Rogers RL, Chissom BS. Derivation Of New Readability Formulas (Automated Readability Index, Fog Count And Flesch Reading Ease Formula) For Navy Enlisted Personnel. Published online 1975. doi:[10.21236/ADA006655](https://doi.org/10.21236/ADA006655)
8. Kincaid JP, Braby R, Mears JE. Electronic Authoring and Delivery of Technical Information. *Journal of Instructional Development*. Published online 1988. doi:[10.1007/BF02904998](https://doi.org/10.1007/BF02904998)
9. Williamson JM, Martin AG. Analysis of patient information leaflets provided by a district general hospital by the Flesch and Flesch-Kincaid method. *Int J Clin Pract*. 2010;64(13):1824-1831. doi:[10.1111/j.1742-1241.2010.02408.x](https://doi.org/10.1111/j.1742-1241.2010.02408.x)
10. Paasche-Orlow MK, Taylor HA, Brancati FL. Readability Standards for Informed-Consent Forms as Compared with Actual Readability. *New Engl J Med*. Published online 2003. doi:[10.1056/NEJMsa021212](https://doi.org/10.1056/NEJMsa021212)
11. Grossman SA, Piantadosi S, Covahey C. Are informed consent forms that describe clinical oncology research protocols readable by most patients and their families? *J Clin Oncol*. 1994;12(10):2211-2215. doi:[10.1200/JCO.1994.12.10.2211](https://doi.org/10.1200/JCO.1994.12.10.2211)
12. Barbarite E, Shaye D, Oyer S, Lee LN. Quality Assessment of Online Patient Information for Cosmetic Botulinum Toxin. *Aesthetic Surgery Journal*. 2020;40(11):NP636-NP642. doi:[10.1093/asj/sjaa168](https://doi.org/10.1093/asj/sjaa168)
13. Margol-Gromada M, Sereda M, Baguley DM. Readability assessment of self-report hyperacusis questionnaires. *International Journal of Audiology*. 2020;59(7):506-512. doi:[10.1080/14992027.2020.1723033](https://doi.org/10.1080/14992027.2020.1723033)
14. Alas AN, Bergman J, Dunivan GC, et al. Readability of Common Health-Related Quality-of-Life Instruments in Female Pelvic Medicine. *Female Pelvic Medicine & Reconstructive Surgery*. 2013;19(5):293-297. doi:[10.1097/spv.0b013e31828ab3e2](https://doi.org/10.1097/spv.0b013e31828ab3e2)
15. Gulati R, Nawaz M, Lam L, Pysopoulos NT. Comparative Readability Analysis of Online Patient Education Resources on Inflammatory Bowel Diseases. *Can J Gastroenterol Hepatol*. 2017;2017:3681989. doi:[10.1155/2017/3681989](https://doi.org/10.1155/2017/3681989)
16. Choudhry AJ, Baghdadi YMK, Wagie AE, et al. Readability of discharge summaries: with what level of information are we dismissing our patients? *The American Journal of Surgery*. 2016;211(3):631-636. doi:[10.1016/j.amjsurg.2015.12.005](https://doi.org/10.1016/j.amjsurg.2015.12.005)
17. Tollinche LE, Shi R, Hannum M, et al. The impact of real-time clinical alerts on the compliance of anesthesia documentation: A retrospective observational study. *Computer Methods and Programs in Biomedicine*. 2020;191:105399. doi:[10.1016/j.cmpb.2020.105399](https://doi.org/10.1016/j.cmpb.2020.105399)
18. Payne TH, Patel R, Beahan S, Zehner J. The physical attractiveness of electronic physician notes. *AMIA Annu Symp Proc*. 2010;2010:622-626. <https://www.ncbi.nlm.nih.gov/pubmed/21347053>
19. Bakken S, Wrenn JO, Siegler EL, Stetson PD. Assessing Electronic Note Quality Using the Physician Documentation Quality Instrument (PDQI-9). *Applied Clinical Informatics*. 2012;03(02):164-174. doi:[10.4338/aci-2011-11-ra-0070](https://doi.org/10.4338/aci-2011-11-ra-0070)

20. Manning C, Surdeanu M, Bauer J, Finkel J, Bethard S, McClosky D. *The Stanford CoreNLP Natural Language Processing Toolkit*. Association for Computational Linguistics; 2014:55-60. doi:[10.3115/v1/P14-5010](https://doi.org/10.3115/v1/P14-5010)
21. Deshpande A. Counting syllables within a word. <https://stackoverflow.com/a/34209751>
22. Gunning R. *The Technique of Clear Writing*. McGraw-Hill; 1952.
23. Dale E, Chall J. A Formula for Predicting Readability. *Educational Research Bulletin*. 1948;21(1):11-20.
24. Savoy A, Frankel R, Weiner M. Clinical Thinking via Electronic Note Templates: Who Benefits? *Journal of General Internal Medicine*. 2021;36(3):577-579. doi:[10.1007/s11606-020-06376-y](https://doi.org/10.1007/s11606-020-06376-y)
25. Cimino JJ. Putting the “why” in “EHR”: capturing and coding clinical cognition. *Journal of the American Medical Informatics Association*. 2019;26(11):1379-1384. doi:[10.1093/jamia/ocz125](https://doi.org/10.1093/jamia/ocz125)
26. Tsou A, Lehmann C, Michel J, Solomon R, Possanza L, Gandhi T. Safe Practices for Copy and Paste in the EHR. *Applied Clinical Informatics*. 2017;26(01):12-34. doi:[10.4338/aci-2016-09-r-0150](https://doi.org/10.4338/aci-2016-09-r-0150)
27. Cimino JJ. Improving the Electronic Health Record---Are Clinicians Getting What They Wished For? *JAMA*. 2013;309(10):991. doi:[10.1001/jama.2013.890](https://doi.org/10.1001/jama.2013.890)