Fluoroscopy/IR Skin-Reaction Follow-up Handout: Constrained LLM Rewrites Improve Readability with Preserved Expert-Rated Fidelity and Actionability

Shahab Ahmad Khan, MS[1] · Muhannad N. Fadhel, PhD, DABR[2] · Rasha S. Makkia, PhD, DABR[3] · Dania Daye, MD, PhD[1] · **Mosa Alhamami, PhD, MCCPM[1,4*]**

[1]University of Wisconsin–Madison, Madison, WI · [2]Northwestern University, Chicago, IL · [3]Medical College of Georgia, Augusta University, Augusta, GA [4]Radiology Physics, LLC, Houston, TX *Presenting & corresponding author

2026 Joint AAPM | COMP
Annual Meeting & Exhibition
July 19–22, 2026 · Vancouver, BC, Canada
Digital Poster Viewing · Abstract #26503

BACKGROUND & PURPOSE

- Radiation-induced skin reactions are a recognized risk of fluoroscopically guided interventional (FGI) procedures; substantial-dose cases require structured follow-up with clear patient instructions. [1–3]
- Patient-facing materials often exceed recommended reading levels (≤6th grade), risking reduced comprehension and adherence. [4]
- Large language models (LLMs) can simplify health information, but unconstrained rewrites risk introducing unsupported clinical claims. [5,6]
- Purpose: evaluate whether a constrained, source-text-only LLM workflow improves readability of a fluoroscopy skin-reaction follow-up handout while preserving safety-critical content and actionability. [11]
    - Note: source handout evaluated here was already relatively readable at baseline, allowing us to test whether constrained rewriting could produce additional gains without compromising content fidelity or required patient actions.

INNOVATION & IMPACT

- Demonstrates a constrained, source-text-only LLM workflow that improves readability of substantial-dose follow-up instructions without introducing new clinical claims.
- Pairs readability metrics with multidisciplinary review of technical fidelity/appropriateness and a safety-critical actionability checklist, supporting clinically defensible adoption. [10]
- Shows prompt intent matters: readability- and fidelity-focused prompts achieved uniformly high appropriateness and perfect checklist alignment, informing future protocolized prompting.

METHODS

1. **INPUT** Baseline patient handout HF#6089 (V0) · UW Health "Health Facts for You" patient-education document (publicly available). [11]
2. **CONSTRAINT HEADER** Source text only · no new clinical claims · missing essential safety information flagged as [NEEDS CLINICIAN INPUT].
3. **THREE INDEPENDENT REWRITES** Generated with ChatGPT (OpenAI) · V1 readability-focused · V2 actionability-optimized · V3 fidelity-focused.
4. **OUTCOMES** Readability (FKGL, FRE, GFI) · blinded panel appropriateness · 10-item safety-critical actionability checklist.

Blinded multidisciplinary review

- Panel of four raters: two diagnostic medical physicists, one interventional radiologist, one data scientist working in IR physics.
- Rewrites were blinded to raters; the corresponding author did not participate in scoring.
- Rater agreement summarized using rating distributions and checklist score ranges.

Readability metrics [7–9]

$$\mathrm{FKGL} = 0.39\cdot(W/S) + 11.8\cdot(Syl/W) - 15.59$$

$$\mathrm{FRE} = 206.835 - 1.015\cdot(W/S) - 84.6\cdot(Syl/W)$$

$$\mathrm{GFI} = 0.4\cdot[(W/S) + 100\cdot(CW/W)]$$

W words · S sentences · Syl syllables · CW complex words (≥3 syllables).
FKGL and GFI estimate the U.S. school-grade level needed to understand the text on first reading; FRE is a 0–100 ease score (higher = easier; ≥80 ≈ conversational English). These validated, empirically derived regression formulas quantify sentence length and word complexity [7–9], enabling objective, reproducible comparison against the ≤6th-grade target for patient materials [4].

RESULTS – READABILITY

Version	FKGL ↓	FRE ↑	GFI ↓
V0 (baseline)	4.68	80.70	7.06
V1 · readability	**3.48**	**83.38**	**5.75**
V2 · actionability	4.15	82.76	6.67
V3 · fidelity	4.10	81.97	6.44

Table 1. Readability of the baseline handout (V0) and constrained LLM rewrites (V1–V3). Best values in bold.

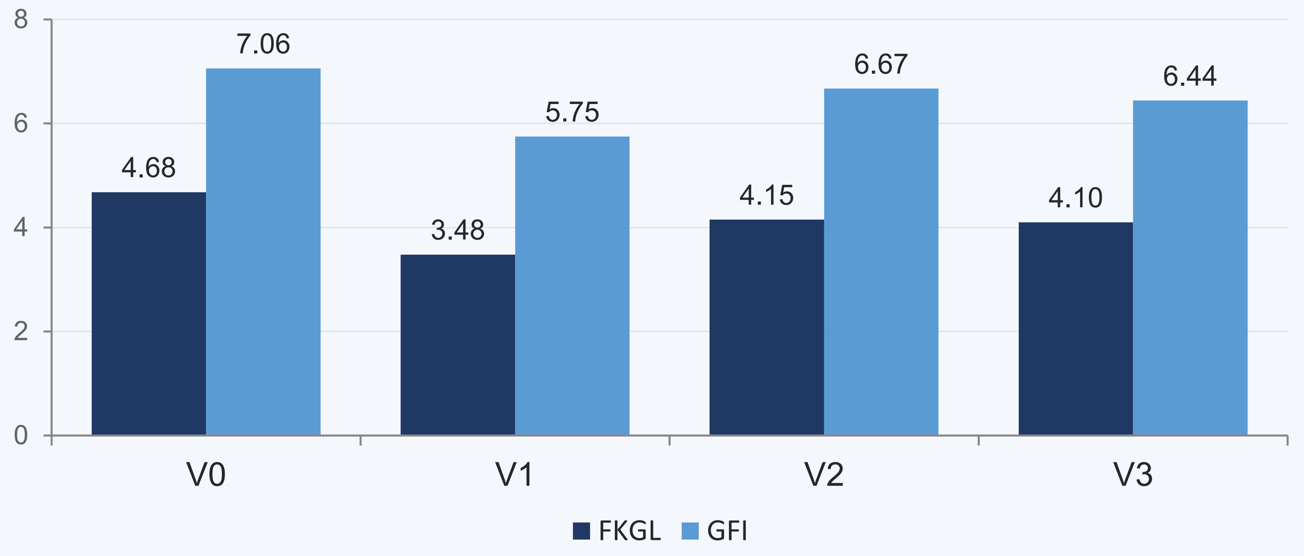

Figure 1. FKGL and GFI by version (grade-level scales; lower = easier reading).

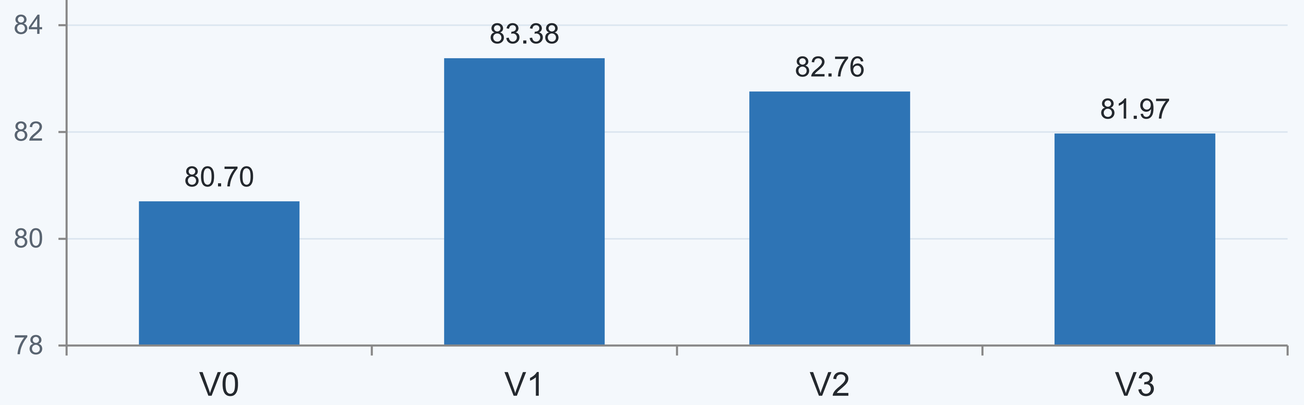

Figure 2. Flesch Reading Ease (FRE) by version (0–100 scale; higher = easier; axis 78–85).

- Readability improved on all three metrics for every rewrite (V1–V3).
- Mean change vs baseline (±SD): ΔFKGL −0.76 ± 0.38 · ΔFRE +2.00 ± 0.71 · ΔGFI −0.77 ± 0.48.

RESULTS – EXPERT REVIEW

Rewrite	Fully approp.	Somewhat	Not	Actionability (0–10), mean (range)
V1	**4/4**	0/4	0/4	**10.0 (10–10)**
V2	3/4	1/4	0/4	9.5 (8–10)
V3	**4/4**	0/4	0/4	**10.0 (10–10)**

Table 2. Blinded multidisciplinary panel (4 raters): technical fidelity/appropriateness and 10-item HF#6089-derived safety-critical actionability checklist.

4/4
Fully appropriate · V1 & V3 · all raters

10/10
Actionability checklist · V1 & V3 · every rater

CONCLUSIONS

- A constrained, source-text-only LLM workflow produced consistent readability gains while preserving expert-rated appropriateness and high actionability.
- Prompt intent influenced performance: readability- (V1) and fidelity-focused (V3) prompting yielded the most consistent balance.
- Findings support constrained LLM prompting to refine substantial-dose follow-up materials without adding new clinical claims, informing future protocolized prompting for patient-facing radiation-safety materials.

Limitations

Single source handout (HF#6089); four-rater panel; three rewrite variants. Generalizability to other materials and institutions requires further evaluation.

References

1. Balter S, Hopewell JW, Miller DL, Wagner LK, Zelefsky MJ. Fluoroscopically guided interventional procedures: a review of radiation effects on patients' skin and hair. Radiology. 2010;254(2):326–341.
2. National Council on Radiation Protection and Measurements. Radiation Dose Management for Fluoroscopically-Guided Interventional Medical Procedures. NCRP Report No. 168. Bethesda, MD: NCRP; 2010.
3. Stecker MS, Balter S, Towbin RB, et al. Guidelines for patient radiation dose management. J Vasc Interv Radiol. 2009;20(7 Suppl):S263–S273.
4. Weiss BD. Health Literacy: A Manual for Clinicians. Chicago, IL: American Medical Association Foundation; 2003.
5. Kirchner GJ, Kim RY, Weddle JB, Bible JE. Can artificial intelligence improve the readability of patient education materials? Clin Orthop Relat Res. 2023;481(11):2260–2267.
6. Ayre J, Mac O, McCaffery K, et al. New frontiers in health literacy: using ChatGPT to simplify health information for people in the community. J Gen Intern Med. 2024;39(4):573–577.
7. Flesch R. A new readability yardstick. J Appl Psychol. 1948;32(3):221–233.
8. Kincaid JP, Fishburne RP, Rogers RL, Chissom BS. Derivation of New Readability Formulas for Navy Enlisted Personnel. Research Branch Report 8-75. Millington, TN: Naval Technical Training Command; 1975.
9. Gunning R. The Technique of Clear Writing. New York, NY: McGraw-Hill; 1952.
10. Shoemaker SJ, Wolf MS, Brach C. Development of the Patient Education Materials Assessment Tool (PEMAT). Patient Educ Couns. 2014;96(3):395–403.
11. UW Health. Skin Reactions After Radiation Exposure. Health Facts for You #6089. Madison, WI. patient.uwhealth.org/healthfacts/6089.

Presenting & Corresponding Author
Mosa Alhamami, PhD MCCPM · mosa@RadiologyPhys.com

Disclosures: none · No NIH funding · Category: Diagnostic & Interventional Radiology Physics