Original Paper

Structured MRI Reporting as an In-Workflow Educational Tool for Rectal Cancer Staging During Radiology Residency: An Exploratory Pilot Study

MIHAI-ALEXANDRU ENE[1], LUCIAN-MIHAI FLORESCU[2], ANDREI-GABRIEL NASCU[3], CRISTINA-MIHAELA CIOFIAC[2], ROSSY-VLĂDUȚ TEICĂ[2], IOANA-ANDREEA CÎRLIG[1], AURELIA-ȘTEFANIA DOMENCO[1], IOANA-ANDREEA GHEONEA[2]

[1]*Doctoral School, University of Medicine and Pharmacy of Craiova*
[2]*Department of Radiology and Medical Imaging, University of Medicine and Pharmacy of Craiova*
[3]*Department of Informatics, University of Craiova*

ABSTRACT: Objectives: This exploratory pilot study assessed whether structured reporting (SR) could improve primary rectal cancer MRI reports among radiology residents in a low-exposure training setting. Methods: A pre-study survey was administered to 56 residents at a single training center; 45 responded (80.4%), and only five reported more than 15 prior rectal cancer MRI cases. Twelve residents were stratified by prior exposure and randomized into two balanced groups. In Phase A, Group 1 reported three cases using free text (FT), while Group 2 used an SR template. Afterward, Group 1 attended a focused mini-course, whereas Group 2 received individualized case-based feedback. In Phase B, both groups reported three new cases using SR. For each case, 15 descriptors were classified as omitted, incorrect, or correct. Results: In Phase A, SR was associated with fewer omissions than FT reporting (5.6% vs. 24.4%; p=0.004). Correctness was descriptively higher with SR (65.6% vs. 47.8%; p=0.108), while error rate was also higher (30.0% vs. 22.2%; p=0.044). In Phase B, no significant between-group differences were observed. In Group 1, switching from FT to SR reduced omissions (24.4% to 0.0%; p=0.031). Conclusions: SR was associated with immediate improvements in report completeness among radiology residents. Its educational value may derive from converting unreported descriptors into explicit interpretative attempts, creating opportunities for targeted feedback, and serving as an educational scaffold. Larger multicenter studies with delayed follow-up are needed to confirm these findings and assess long-term learning effects.

KEYWORDS: *Structured reporting, rectal cancer, magnetic resonance imaging, radiology residency, medical education, pilot study.*

Introduction

The rationale for this study emerged from three empirical observations at a single radiology residency training center that may reflect broader patterns in rectal MRI training practice.

First, adherence to current guidelines for primary rectal cancer MRI staging was markedly inconsistent among residents and non-subspecialized radiologists, with reports ranging from sparse, unstructured free-text (FT) to individually improvised hybrid formats and frequent omissions of key staging elements.

Second, standardized templates were absent from routine local practice. Third, cumulative experience with rectal cancer MRI accumulated slowly and unevenly during residency.

These observations converged on a practical question: in a setting where residents encounter rectal cancer MRI findings infrequently, report them inconsistently, and lack a standardized reporting tool, could the introduction of a structured reporting (SR) template simultaneously serve as both reporting support and an educational instrument?

To our knowledge, no prior study has examined SR as an educational intervention for residents in rectal cancer MRI reporting. Prior work in other subspecialties, including head and neck imaging, neuroradiology, and prostate MRI, suggests that SR may improve trainee reporting performance [1-3], but its role in rectal MRI training remains insufficiently studied.

SR has been proposed to reduce reporting variability, improve completeness, and standardize radiological workflows [4-8].

Multiple professional societies have endorsed template-based reporting and structured staging lexicons for rectal cancer MRI [4, 9-18].

Implementation nonetheless remains incomplete, constrained in part by workflow integration challenges and concerns regarding template rigidity [7,19].

Rectal cancer MRI is a high-complexity staging task. Significant interobserver

10.12865/CHSJ.52.03.05

This is an open-access article distributed under the terms of a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Public License, which permits unrestricted use, adaptation, distribution and reproduction in any medium, non-commercially, provided the new creations are licensed under identical terms as the original work and the original work is properly cited.

variability has been documented even among experienced readers [20], and expert interpretation has been shown to influence staging and treatment decisions [21].

Formal educational programs can improve staging performance [23], but they are logistically demanding and not routinely integrated into many residency curricula. SR templates, by contrast, can be embedded directly into existing reporting workflows, making them a practically accessible option for educational support.

Objective

In this specific context, we designed a compact exploratory pilot experiment with two related objectives.

Phase A was designed to provide a direct comparison between SR and FT reporting: both groups were balanced at baseline and received no prior educational intervention, so any performance differences were attributable primarily to the reporting format.

Phase B compared two pragmatic training architectures: SR integrated into the prevailing standard of feedback-based residency practice (Group 2, repeated SR use after individualized feedback) versus SR introduced following a formal mini-course not routinely available in the local program (Group 1).

The experiment was structured to simulate the individual reporting session as it occurs in residency, rather than to accumulate exposure equivalent to months of practice.

We hypothesized that integrating SR into a routine-like reporting workflow would produce immediate reductions in omissions and improve short-term reporting performance in residents.

Methods

This study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the University of Medicine and Pharmacy of Craiova.

Participation was voluntary, and written informed consent was obtained from all participants. M.A.E. was the only team member with access to participant identities; all evaluators involved in scoring (L.M.F. and I.A.G.) were blinded to participant identity.

The proposed workflow of the study is depicted in Figure 1.

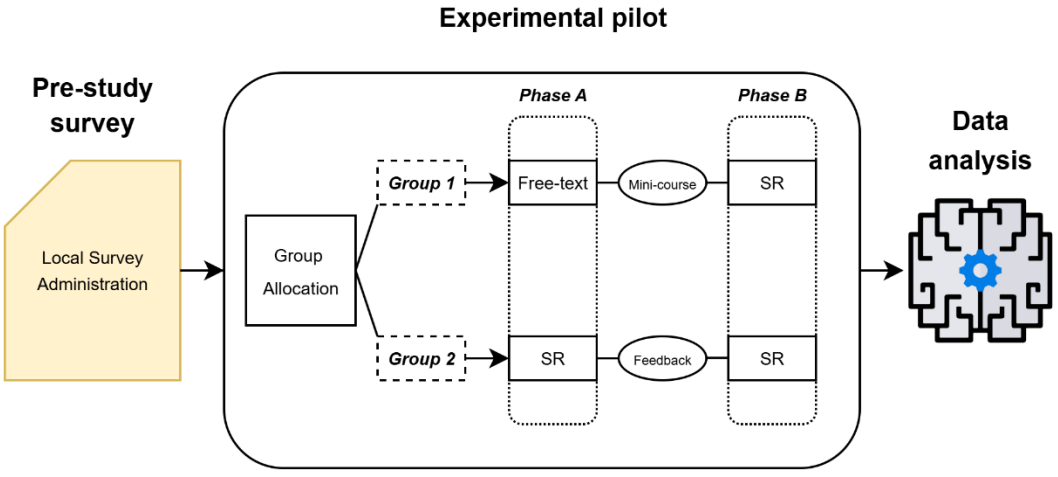


Figure 1. Proposed workflow of the study.

Pre-study experience survey and stratified allocation

Before the experimental phase, all 56 radiology residents at the local training center were asked to self-report their prior exposure to rectal cancer MRI using predefined categories: none (0 cases), low (1-5 cases), intermediate (6-15 cases), high (16-30 cases), and very high (>30 cases). 45 residents responded (80.4% uptake).

Residents were stratified into three eligible experience strata: low (1-5 cases), intermediate (6-15 cases), and high (16-30 cases).

Residents with no prior exposure and those with more than 30 cases were excluded.

Eligible residents were randomly selected within each stratum, with proportions preserved in both groups. Each group included three low-, two intermediate-, and one high-experience resident (n=6 per group).

10.12865/CHSJ.52.03.05

Experimental workflow and reference standard

Six rectal cancer MRI cases were selected and divided into two series of three cases each. Within each series, cases were chosen to represent low, intermediate, and high interpretative difficulty. The low-difficulty case was placed centrally in each series to reduce expectation bias. The two case sets were matched in overall difficulty distribution.

A composite reference standard was developed through investigator consensus.

Three radiologists independently reviewed all cases: a senior radiology resident specializing in rectal MRI (M.A.E.), an abdominal imaging specialist (L.M.F.), and an expert radiologist with over 20 years of clinical practice (I.A.G.). Disagreements were resolved by consensus.

An SR template for rectal MRI primary staging was developed using validated models proposed by ESGAR [4] and SAR [22], translated and adapted to local reporting practices. The SR template is shown in Figure 2.

MRI Structured Reporting Template for Rectal Cancer Primary Staging

(Adapted to local practice for educational and research purposes within the present study)

Morphology

☐ Polypoid – at [...] o'clock
☐ Semicircumferential – between [...] and [...] o'clock
☐ Circumferential

Mucinous component

☐ No mucinous component
☐ Partially mucinous
☐ Predominantly mucinous

Tumor Location

☐ Upper rectum
☐ Mid rectum
☐ Lower rectum

Distance from anal verge: [...] cm
Distance from anorectal junction: [...] cm

Relation to the anterior peritoneal reflection

☐ Above
☐ Below
☐ Extending both above and below

Cranio-caudal tumor **length**: [...] mm
Maximum tumor **thickness**: [...] mm

T Stage

☐ T1/2
☐ T3a (<1 mm)
☐ T3b (1–5 mm)
☐ T3c (>5–15 mm)
☐ T3d (>15 mm)
☐ T4a
☐ T4b – involved organs: [...]

Sphincter Complex

☐ No invasion
☐ Internal sphincter only
☐ Internal sphincter + intersphincteric space
☐ External sphincter involvement

If applicable – level involved:
☐ Upper anal canal
☐ Mid anal canal
☐ Distal anal canal

External sphincter invasion description (if present): [...]

Mesorectal Fascia (MRF)
☐ Free – minimum tumor-to-MRF distance: [...] mm, at [...] o'clock
☐ Involved, at [...] o'clock
☐ MRF+ due to tumor deposit/lymph node/EMVI at [...] o'clock

Peritoneal Reflection
☐ Free
☐ Invaded

Suspicious Lymph Nodes

N Stage
☐ N0
☐ N+
Number of suspicious regional lymph nodes*: [...]

For mesorectal and presacral/inferior mesenteric N+:
[...] with short axis ≥ 9 mm
[...] with short axis 5–8 mm + at least 2 morphological criteria**
[...] with short axis < 5 mm + all 3 morphological criteria**

High regional location (presacral/inferior mesenteric): [...]

Regional extramesorectal N+
☐ Right internal iliac: [...] mm
☐ Right obturator: [...] mm
☐ Right inguinal†: [...] mm
☐ Left internal iliac: [...] mm
☐ Left obturator: [...] mm
☐ Left inguinal†: [...] mm
†*Only if tumor is below the dentate line*

M1 Lymph Nodes (location + short-axis diameter)
☐ Right common iliac: [...] mm
☐ Right external iliac: [...] mm
☐ Right inguinal: [...] mm
☐ Right retroperitoneal: [...] mm
☐ Left common iliac: [...] mm
☐ Left external iliac: [...] mm
☐ Left inguinal: [...] mm
☐ Left retroperitoneal: [...] mm

Tumor Deposits
☐ No
☐ Yes – description, number, location: [...]

Extramural Vascular Invasion (EMVI)
☐ No
☐ Yes

Additional Findings [...]

Conclusion
Primary tumor location: [...]
MRI staging: T [...] N [...] M [...]
Sphincter invasion: No / Yes [...]
MRF: Negative / Positive
EMVI: Negative / Positive

Figure 2. Structured reporting template.

All sessions approximated routine clinical reporting conditions. Online educational resources were accessible during reporting, consistent with everyday practice; communication between participants and the use of AI-based tools were not permitted.

Reports were completed sequentially without interrupting the overall timing, and the total phase reporting time was recorded for each participant. All cases were anonymized.

Participants were identified by study-specific IDs, and the blinded evaluators (L.M.F. and I.A.G.) remained unaware of participant identity throughout.

In Phase A, Group 1 reported the first case series using FT, while Group 2 used the SR template. Phase A was designed to isolate the immediate effect of the reporting format.

Then, Group 1 attended a focused mini-course (~ 2 hours) delivered by M.A.E. on primary rectal MRI staging that systematically covered key imaging descriptors and SR principles. Group 2 received individualized, case-based feedback on their Phase A reports from M.A.E., who reviewed correction criteria and addressed errors or uncertainties, with sessions lasting approximately 30 to 60 minutes per participant. This feedback-based approach sought to reflect the standard of practice in residency at the local center.

In Phase B, both groups reported the second case series using SR. Group 2 used SR for the second consecutive time, in continuity with the individualized feedback received; Group 1 used SR for the first time following the mini-course.

Phase B, therefore, compared individualized feedback (Group 2) with formal teaching (Group 1) to test the feasibility of integrating SR into the routine residency workflow in the absence of more costly approaches (e.g., lecture-based).

Statistical analysis

Statistical analyses and graphical outputs were performed by A.G.N. using Python in a Jupyter Notebook environment.

For each case, 15 key findings (Table 1) were encoded as 0 (omitted), 1 (incorrectly reported), or 2 (correctly reported). Omission, error, and correctness rates were calculated per participant per phase by aggregating outcomes across the three cases and 15 descriptors, yielding 45 descriptor-level decisions per resident per phase.

Table 1. Evaluated descriptors for each case.

No.	Finding
1	Tumor shape
2	Tumor T2 signal
3	Cranio-caudal tumor length
4	Localization
5	Anal canal involvement
6	T stage
7	MRF status
8	Minimum MRF distance/involvement localization
9	Peritoneal reflection status and relation
10	EMVI status
11	Tumor deposits status
12	N stage
13	N count (approx.)
14	Lateral lymph nodes status
15	Most cranially located suspicious lymph node

As a primary descriptive measure of individual-level performance, each participant's case-level metric was compared with the median across all 12 participants for the corresponding case and metric. Performance was classified as favorable when at or below the median for omissions, errors, and reporting time, or at or above the median for correctness; the number of favorable assessments per group across all three cases was reported. This approach was used to convey the direction and consistency of individual-level group separation while preserving participant-level resolution.

Group-level comparisons used the Mann-Whitney U test for between-group comparisons and the Wilcoxon signed-rank test for within-group phase comparisons, applied to participant-level phase summaries.

All p-values are reported as exploratory, given the small sample size. Median differences serve as effect size estimates.

Participant-level distributions were visualized using strip charts, with each point representing one resident's phase-level summary and with median and interquartile range displayed for each phase/group combination. In addition, paired plots were generated to illustrate within-resident changes in omission rate from Phase A to Phase B, separately for each group.

Additional descriptive analyses included paired plots for correctness rate, error rate, and total reporting time; binary median-threshold matrices providing case-level views of performance consistency; and exploratory principal component analysis (PCA) of the full 15-descriptor space. The non-omission paired

10.12865/CHSJ.52.03.05

plots and median-threshold matrices were used for visualization only.

PCA was performed as a descriptive, hypothesis-generating dimensionality-reduction analysis and was not used for inferential testing.

Results

Pre-study experience survey

Before the experimental phase, 45 of 56 residents responded (80.4% uptake).

Nine respondents reported no prior MRI experience and were excluded from eligibility assessment, representing 20.0% of respondents.

Among the remaining 36 residents undergoing MRI training, exposure to rectal cancer MRI was generally limited (Figure 3): 5 residents (13.9%) reported no rectal MRI experience, 18 (50.0%) reported 1-5 cases, 8 (22.2%) reported 6-15 cases, 3 (8.3%) reported 16-30 cases, and only 2 (5.6%) reported more than 30 cases. Among 11 fifth-year residents, 2 (18.2%) reported 1-5 cases, 4 (36.4%) reported 6-15 cases, 3 (27.3%) reported 16-30 cases, and 2 (18.2%) reported more than 30 cases. These findings confirmed that rectal MRI experience accumulated slowly and unevenly across the training cohort.

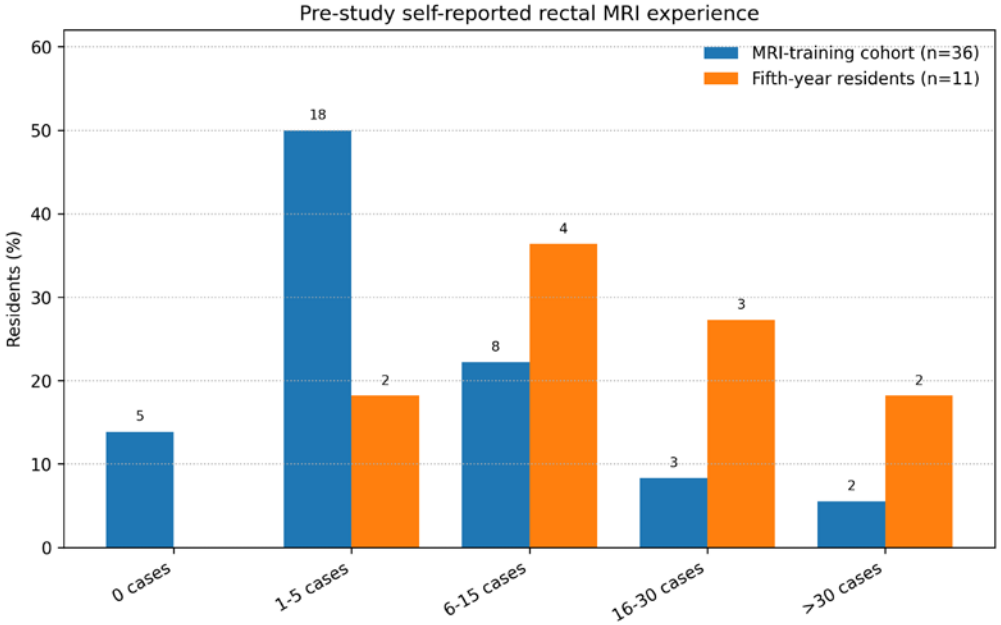


Figure 3. Pre-study self-reported rectal MRI experience among residents undergoing MRI training (n=36) and final-year residents (n=11); the y-axis indicates percentages within each subgroup.

Educational pilot study

In Phase A, SR was associated with a marked reduction in the omission rate compared with FT reporting. Median omission rate was 24.4% (IQR, 22.2-31.7%) in the FT group and 5.6% (IQR, 1.1-6.7%) in the SR group (median difference, −18.9 percentage points; p=0.004).

Error rate was significantly higher in the SR group (30.0% [IQR, 27.2-41.1%] vs 22.2% [IQR, 20.6-25.6%]; median difference, +7.8 pp; p=0.044), while correctness was descriptively higher but not significantly different (65.6% [IQR, 52.2-70.6%] vs 47.8% [IQR, 46.7-53.9%]; p=0.108).

In Phase B, no statistically significant between-group differences were observed.

Omission rates were very low in both groups (median 0.0% [IQR, 0.0-1.7%] in Group 1 and 1.1% [IQR, 0.0-2.2%] in Group 2; p=0.718).

The feedback/repeated-SR group showed a descriptively more favorable correctness/error profile - higher correctness (77.8% [IQR, 73.9-80.0%] vs 68.9% [IQR, 65.6-72.2%]) and lower error rate (21.1% [IQR, 17.8-26.1%] vs 28.9% [IQR, 26.7-32.8%])-but these differences were not statistically supported (p=0.106 for both outcomes).

Group-level medians, interquartile ranges, median differences, and exploratory p-values are summarized in Table 2.

Table 2. Group-level main statistical results.

Outcome	Group 1 Phase A median (IQR)	Group 2 Phase A median (IQR)	Δ Phase A (G2−G1)	p Phase A	Group 1 Phase B median (IQR)	Group 2 Phase B median (IQR)	Δ Phase B (G2−G1)	p Phase B
Correctness rate (%)	47.8 (46.7-53.9)	65.6 (52.2-70.6)	+17.8	p=0.108	68.9 (65.6-72.2)	77.8 (73.9-80.0)	+8.9	p=0.106
Error rate (%)	22.2 (20.6-25.6)	30.0 (27.2-41.1)	+7.8	p=0.044*	28.9 (26.7-32.8)	21.1 (17.8-26.1)	−7.8	p=0.106
Omission rate (%)	24.4 (22.2-31.7)	5.6 (1.1-6.7)	−18.9	p=0.004**	0.0 (0.0-1.7)	1.1 (0.0-2.2)	+1.1	p=0.718
Reporting time (min)	79 (58-97)	100 (73-108)	+21	p=0.630	72 (49-95)	85 (62-100)	+13	p=0.748

*Values are medians with interquartile ranges. Differences are expressed as Group 2 minus Group 1, in percentage points for rates and minutes for reporting time. P-values are exploratory and were calculated using Mann-Whitney U tests for between-group comparisons within each phase. * p <0.05; ** p <0.01.*

The participant-level analytical dataset underlying the group-level comparisons is presented in Table 3 to support transparency and reproducibility.

Table 3. Participant-level analytical dataset used for the main group-level comparisons.

Case	G1.1	G1.2	G1.3	G1.4	G1.5	G1.6	G2.1	G2.2	G2.3	G2.4	G2.5	G2.6
A1_Omis	33.33%	33.33%	33.33%	33.33%	26.67%	33.33%	0.00%	0.00%	6.67%	6.67%	6.67%	6.67%
A2_Omis	46.67%	6.67%	33.33%	6.67%	13.33%	26.67%	0.00%	0.00%	6.67%	6.67%	0.00%	6.67%
A3_Omis	26.67%	26.67%	33.33%	26.67%	26.67%	20.00%	0.00%	0.00%	6.67%	6.67%	6.67%	6.67%
A1_Cor	40.00%	26.67%	40.00%	6.67%	33.33%	33.33%	66.67%	60.00%	46.67%	26.67%	33.33%	33.33%
A2_Cor	20.00%	86.67%	40.00%	93.33%	73.33%	33.33%	80.00%	80.00%	53.33%	40.00%	93.33%	66.67%
A3_Cor	66.67%	53.33%	66.67%	40.00%	66.67%	73.33%	80.00%	73.33%	46.67%	66.67%	80.00%	86.67%
A1_Err	26.67%	40.00%	26.67%	60.00%	40.00%	33.33%	33.33%	40.00%	46.67%	66.67%	60.00%	60.00%
A2_Err	33.33%	6.67%	26.67%	0.00%	13.33%	40.00%	20.00%	20.00%	40.00%	53.33%	6.67%	26.67%
A3_Err	6.67%	20.00%	0.00%	33.33%	6.67%	6.67%	20.00%	26.67%	46.67%	26.67%	13.33%	6.67%
A1_Time	36	64	53	16	19	26	51	59	30	10	57	45
A2_Time	26	15	28	27	10	24	16	20	19	11	20	19
A3_Time	22	22	37	10	13	24	31	49	16	14	33	37
B1_Omis	0.00%	6.67%	0.00%	0.00%	0.00%	0.00%	0.00%	6.67%	6.67%	0.00%	0.00%	0.00%
B2_Omis	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	6.67%	0.00%
B3_Omis	6.67%	6.67%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	6.67%	0.00%	0.00%	0.00%
B1_Cor	66.67%	40.00%	60.00%	40.00%	53.33%	66.67%	80.00%	73.33%	46.67%	66.67%	86.67%	80.00%
B2_Cor	93.33%	93.33%	100.00%	93.33%	93.33%	100.00%	86.67%	100.00%	93.33%	86.67%	86.67%	93.33%
B3_Cor	33.33%	73.33%	60.00%	46.67%	60.00%	60.00%	60.00%	66.67%	40.00%	66.67%	66.67%	80.00%
B1_Err	33.33%	53.33%	40.00%	60.00%	46.67%	33.33%	20.00%	20.00%	46.67%	33.33%	13.33%	20.00%
B2_Err	6.67%	6.67%	0.00%	6.67%	6.67%	0.00%	13.33%	0.00%	6.67%	13.33%	6.67%	6.67%
B3_Err	60.00%	20.00%	40.00%	53.33%	40.00%	40.00%	40.00%	33.33%	53.33%	33.33%	33.33%	20.00%
B1_Time	40	43	60	20	18	25	41	33	20	16	35	27
B2_Time	16	16	17	11	5	11	27	25	16	11	11	14
B3_Time	25	40	44	13	12	26	38	46	20	16	35	47

Abbreviations and coding: *G1.1-G1.6 and G2.1-G2.6 denote individual participants in Groups 1 and 2, respectively. A and B indicate study phase, and the following number (1-3) indicates case number. Omis=omission rate; Cor=correctness rate; Err=error rate; Time=reporting time (min). Rates are expressed as percentages. For example, A1_Omis denotes the omission rate for Case 1 in Phase A.*

Figure 4 illustrates the individual-level distributions underlying the group-level results presented in Table 2, including omission rate, correctness rate, error rate, and total reporting time. Between-group comparisons within each phase were assessed using the Mann-Whitney U test.

Statistically significant differences were observed only in Phase A for omission rate and error rate, while all other between-group comparisons were not statistically significant, as also reported in Table 2.

10.12865/CHSJ.52.03.05

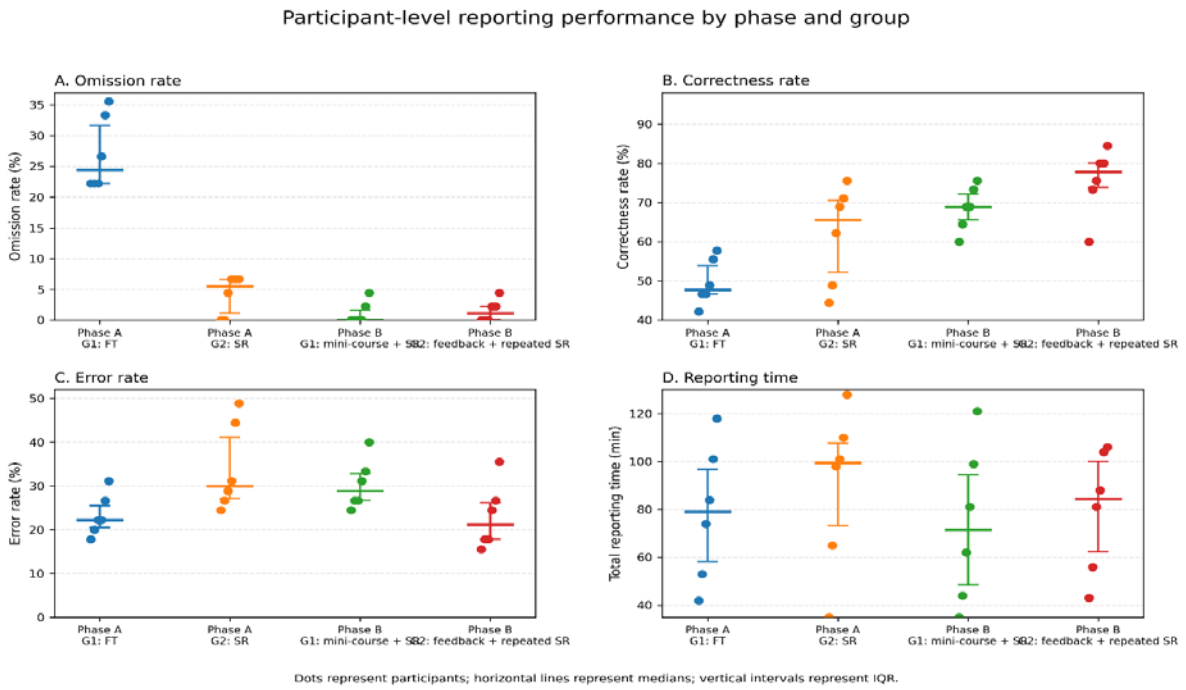


Figure 4. Participant-level reporting performance by phase and group. Strip charts show resident-level phase summaries for omission rate, correctness rate, error rate, and total reporting time. Dots represent individual participants; horizontal lines represent medians; vertical intervals represent interquartile ranges. Between-group comparisons within each phase used the Mann-Whitney U test; statistically significant differences were observed only in Phase A for omission rate and error rate. FT=free-text; SR=structured reporting.

To further describe the consistency of these participant-level patterns across individual cases, favorable median-threshold assessments were aggregated by phase and metric (Figure 5).

This analysis most clearly supported the main findings for omissions: in Phase A, the SR group achieved favorable omission performance across all 18 case-level assessments, compared with 2/18 in the FT group.

In Phase B, favorable omission assessments were frequent in both groups (15/18 in Group 1 and 14/18 in Group 2), consistent with the very low omission rates observed once both groups used SR.

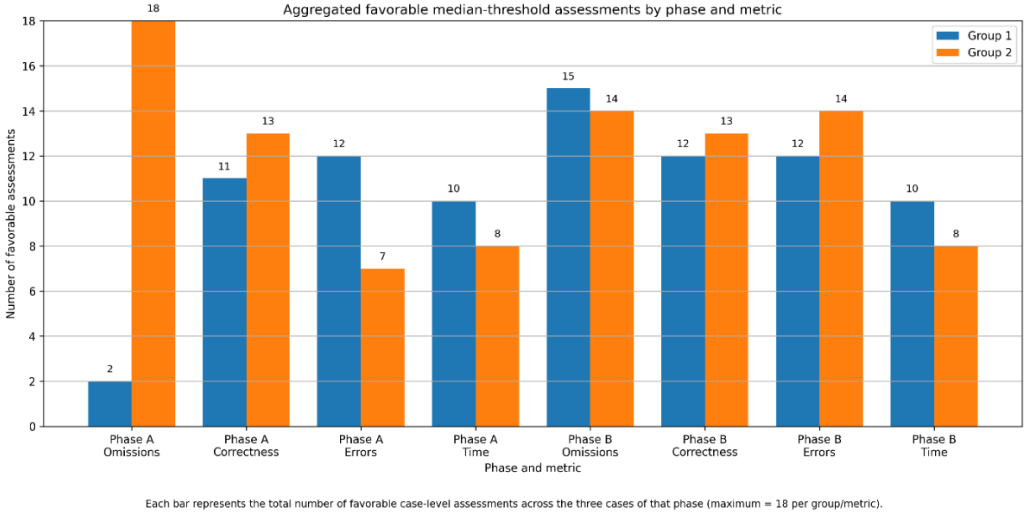


Figure 5. Aggregated favorable median-threshold assessments by phase and metric. Each bar represents the total number of favorable case-level assessments within a group (maximum 18 per phase/metric). Values at or below the overall median were classified as favorable for omissions, errors, and reporting time, whereas values at or above the median were classified as favorable for correctness. This analysis is descriptive and was not subjected to inferential testing.

For correctness, errors, and reporting time, favorable assessments were less clearly separated between groups. To display the participant- and case-level structure underlying the aggregated results in Figure 5, binary median-threshold matrices are shown for Phase A (Figure 6) and Phase B (Figure 7).

Each marker indicates that the participant reached the favorable median threshold for the corresponding case-level metric.

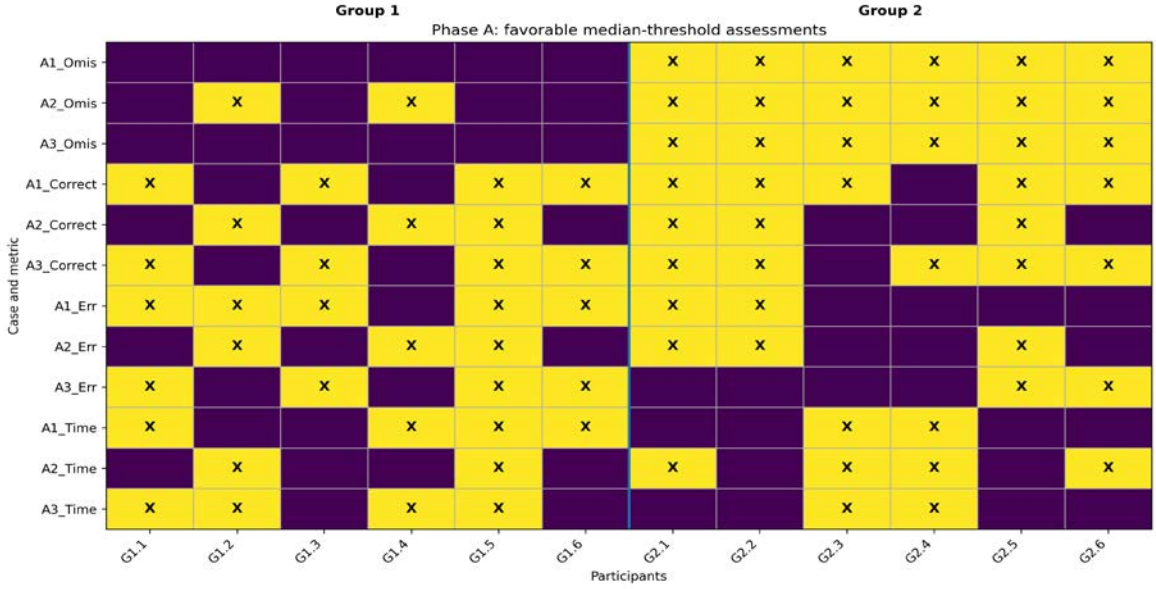


Figure 6. Binary median-threshold matrix for Phase A. Each marker denotes a favorable case-level assessment for the corresponding participant and metric; favorable thresholds were defined as values at or below the median for omissions, errors, and reporting time and at or above the median for correctness. Descriptive visualization only; no inferential test was applied.

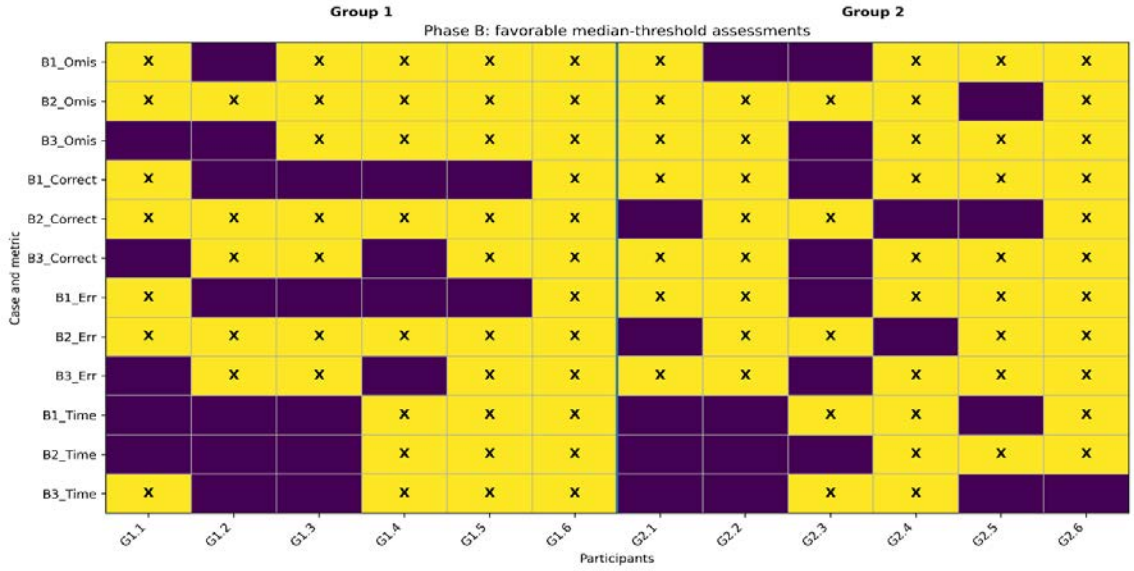


Figure 7. Binary median-threshold matrix for Phase B. Each marker denotes a favorable case-level assessment for the corresponding participant and metric; favorable thresholds were defined as values at or below the median for omissions, errors, and reporting time and at or above the median for correctness. Descriptive visualization only; no inferential test was applied.

Regarding the within-group phase transition, omission rate decreased from 24.4% (IQR, 22.2-31.7%) in Phase A to 0.0% (IQR, 0.0-1.7%) in Phase B following transition to SR after the mini-course (median paired difference, −24.4 pp; p=0.031).

In Group 2, the omission rate was already low in Phase A and remained low in Phase B (5.6% [IQR, 1.1-6.7%] vs 1.1% [IQR, 0.0-2.2%]; median paired difference, −2.2 pp; p=0.250). The within-participant omission dynamics are shown in Figure 8.

10.12865/CHSJ.52.03.05

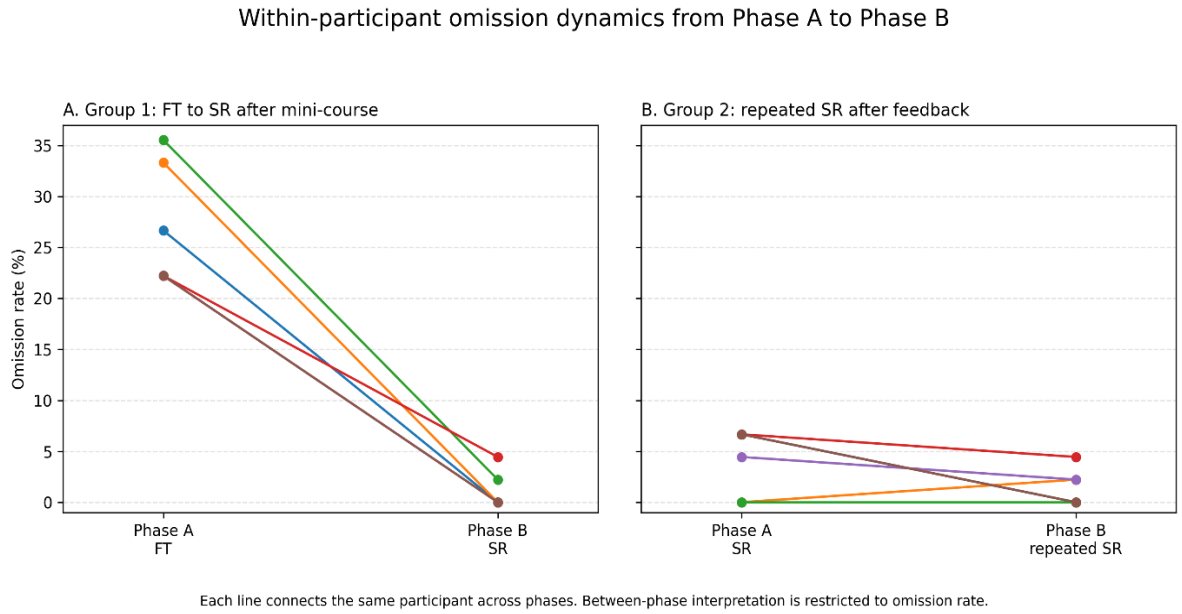


Figure 8. Within-participant omission dynamics from Phase A to Phase B. Each line connects the same participant across phases. In Group 1, the transition was from FT reporting to SR after the mini-course; in Group 2, SR was used in both phases, with individualized feedback before Phase B. Within-group phase comparisons used the Wilcoxon signed-rank test; the decrease was statistically significant in Group 1 (p=0.031) but not in Group 2 (p=0.250).

Paired plots for correctness rate, error rate, and total reporting time are shown in Figure 9 to illustrate within-resident dynamics from Phase A to Phase B. These plots are descriptive only and were not used for formal between-phase inference because the two phases used distinct case sets.

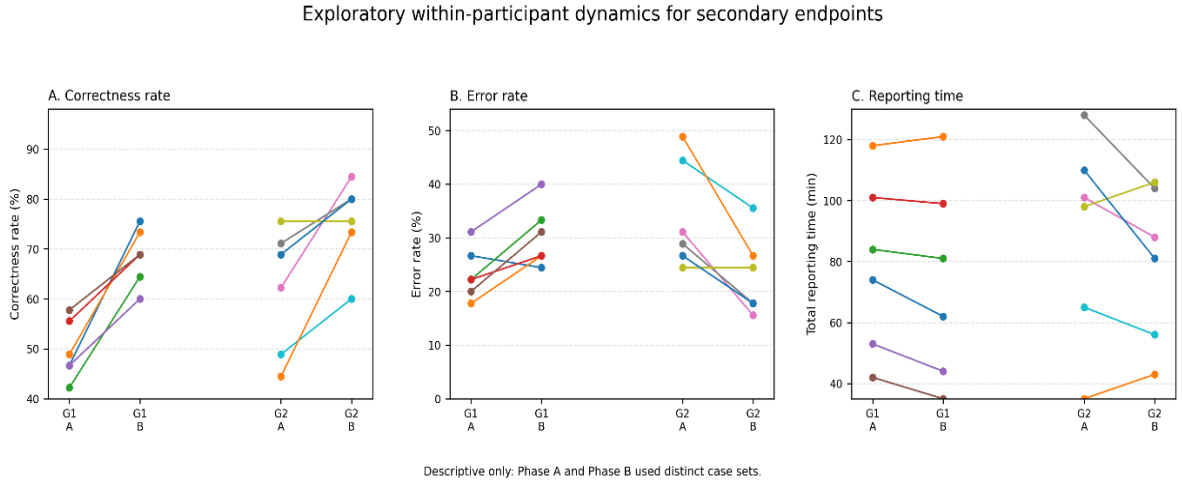


Figure 9. Exploratory paired plots for non-omission endpoints.
Each line connects the same participant between Phase A and Phase B for correctness rate, error rate, and total reporting time. The plots are descriptive only; no formal between-phase statistical testing was performed for these endpoints because the two phases used distinct case sets.

A descriptor-level analysis was performed by pooling all participants, groups, and phases. The most frequently omitted descriptors were: the most cranially located suspicious lymph node description (37.5%), tumor deposits (19.4%), morphology T2 signal (16.7%), peritoneal reflection status (12.5%), and lateral lymph nodes (11.1%). The most frequently misreported descriptors were T stage (47.2%), cranio-caudal tumor length (43.1%), peritoneal reflection status (43.1%), tumor localization (41.7%), and suspicious lymph node count (36.1%). The complete descriptor-level distribution is presented in Table 4.

Table 4. Descriptor-level omission and error distribution across the full dataset (N=72 observations).

Descriptor	Omissions, n/72 (%)	Errors, n/72 (%)	Total omissions + errors, n/72 (%)	Dominant pattern
Peritoneal reflection relation and status	9/72 (12.5%)	31/72 (43.1%)	40/72 (55.6%)	Mixed, predominantly interpretation-driven
T stage	1/72 (1.4%)	34/72 (47.2%)	35/72 (48.6%)	Interpretation-driven
Most cranially located suspicious lymph node	27/72 (37.5%)	7/72 (9.7%)	34/72 (47.2%)	Omission-driven
Cranio-caudal tumor length	2/72 (2.8%)	31/72 (43.1%)	33/72 (45.8%)	Interpretation-driven
Tumor localization	0/72 (0.0%)	30/72 (41.7%)	30/72 (41.7%)	Interpretation-driven
Suspicious lymph node count	3/72 (4.2%)	26/72 (36.1%)	29/72 (40.3%)	Interpretation-driven
Minimum distance to MRF / invasion location	5/72 (6.9%)	23/72 (31.9%)	28/72 (38.9%)	Interpretation-driven
Tumor T2 signal	12/72 (16.7%)	14/72 (19.4%)	26/72 (36.1%)	Mixed
Lateral lymph nodes	8/72 (11.1%)	18/72 (25.0%)	26/72 (36.1%)	Mixed, predominantly interpretation-driven
Anal canal involvement	7/72 (9.7%)	17/72 (23.6%)	24/72 (33.3%)	Interpretation-driven
MRF status	0/72 (0.0%)	21/72 (29.2%)	21/72 (29.2%)	Interpretation-driven
Tumor deposits	14/72 (19.4%)	6/72 (8.3%)	20/72 (27.8%)	Omission-driven
EMVI status	3/72 (4.2%)	16/72 (22.2%)	19/72 (26.4%)	Interpretation-driven
Tumor shape	0/72 (0.0%)	16/72 (22.2%)	16/72 (22.2%)	Interpretation-driven
N stage	0/72 (0.0%)	9/72 (12.5%)	9/72 (12.5%)	Interpretation-driven

Total session reporting time showed high inter-individual variability in both groups and phases, with no statistically significant between-group differences in either Phase A ($p=0.630$) or Phase B ($p=0.748$).

Case-level reporting times followed the intended difficulty calibration: the easy case (Case 2) produced the shortest median times in both Phase A (19.5 min) and Phase B (15 min), whereas the intermediate- and high-difficulty cases produced longer and more variable times.

Median times across matched case positions were broadly comparable between the two case series, supporting the premise that the two phases were similar in overall reporting effort.

Additional exploratory multivariate analysis

Exploratory PCA of the full 15-descriptor space was used to provide a low-dimensional visualization of overall reporting-behavior patterns. Within-phase PCA plots (Figure 10) showed clearer visual separation between groups in Phase A, when Group 1 used FT and Group 2 used SR.

In Phase B, after both groups were using SR, the between-group separation appeared less pronounced, with greater overlap among points.

This descriptive pattern suggests that the main visible distinction in reporting behavior was associated with SR versus FT rather than persistent group-level differences after both groups adopted SR. No inferential testing was applied to the PCA results.

10.12865/CHSJ.52.03.05

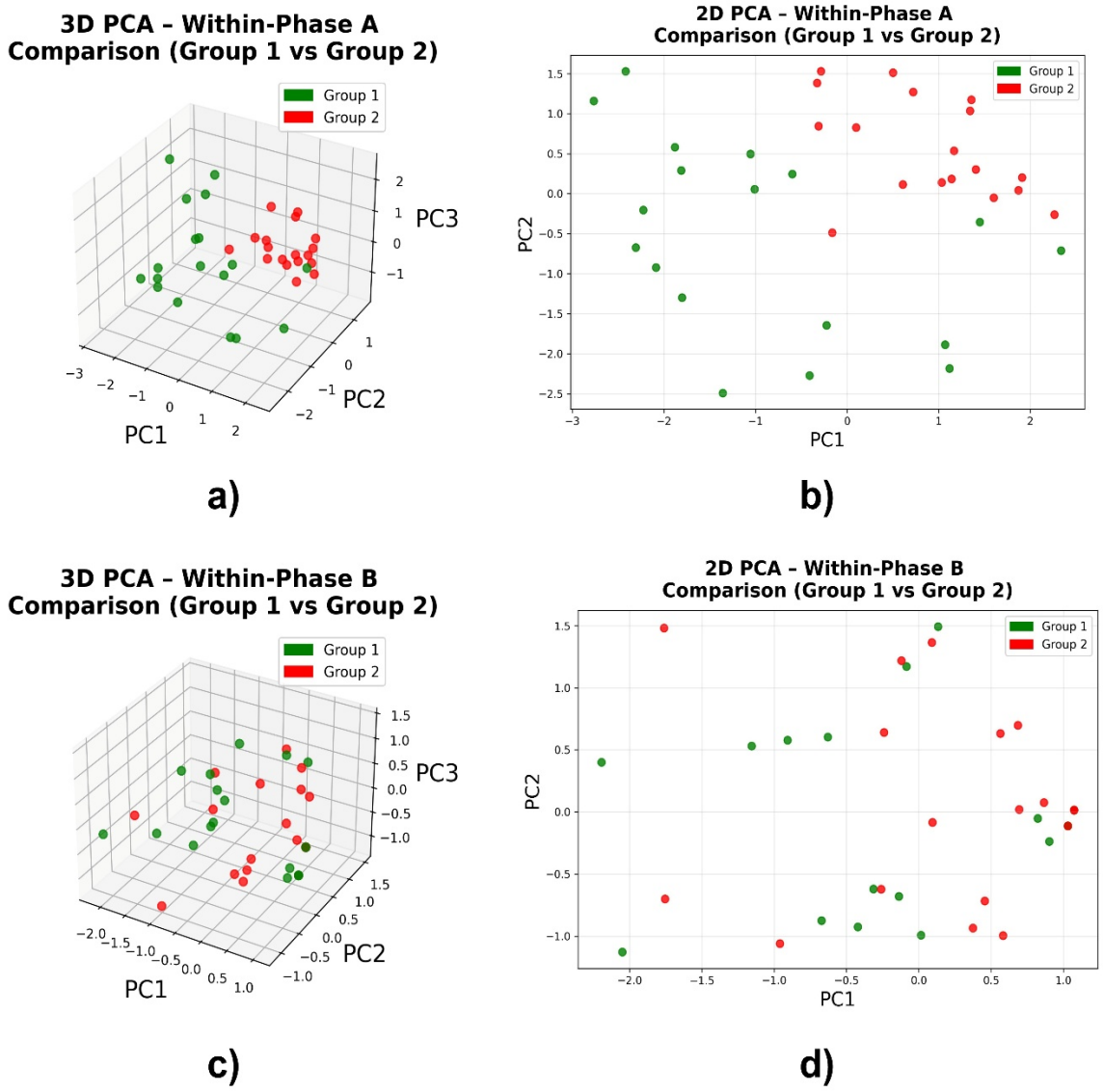


Figure 10. Within-phase principal component analysis (PCA) plots comparing Group 1 and Group 2 separately in Phase A and Phase B. PCA was performed on the full 15-descriptor space as a descriptive, hypothesis-generating visualization; no inferential statistical testing was applied.

Within-group PCA plots (Figure 11) showed a more pronounced phase-related shift in Group 1 than in Group 2.

In Group 1, Phase A and Phase B appeared more clearly separated, and Phase B showed a more compact distribution, suggesting more homogeneous reporting behavior after introduction of SR.

In Group 2, phase separation was less pronounced and showed greater overlap, consistent with SR having been used from the beginning.

These patterns are descriptive and align with the main finding that the largest behavioral shift occurred when SR replaced FT.

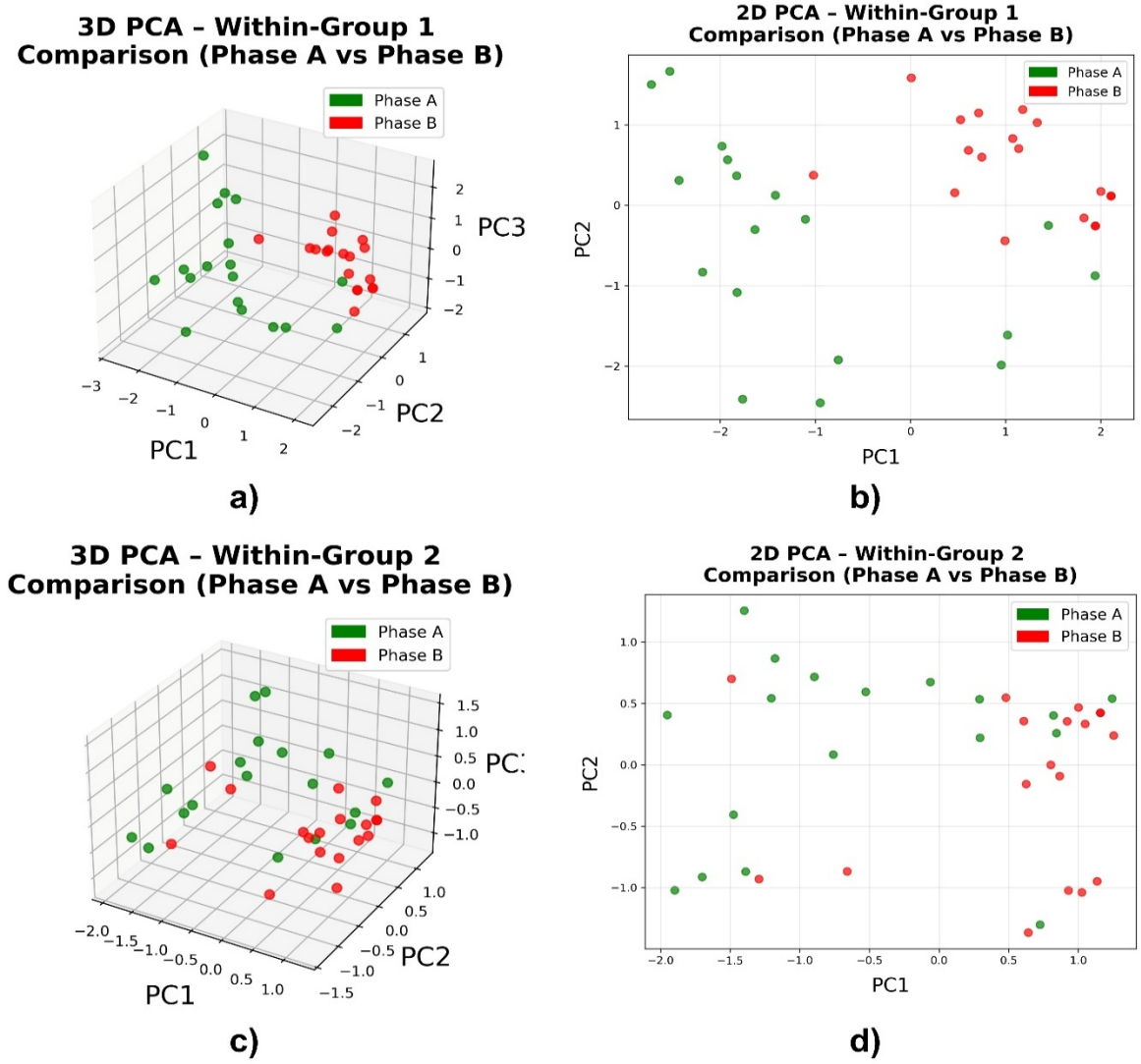


Figure 11. Within-group principal component analysis (PCA) plots comparing Phase A and Phase B separately for Group 1 and Group 2. PCA was used as a descriptive visualization of reporting-behavior patterns; no inferential statistical testing was applied.

Discussion

This exploratory pilot study evaluated SR as an in-workflow educational tool for primary rectal cancer MRI reporting during radiology residency. The main finding was the immediate reduction in omissions associated with SR in Phase A. Even without prior teaching, the template prompted residents to address staging descriptors that were frequently omitted in FT reports.

A particularly relevant observation was the relationship between omissions, correctness, and errors. In Phase A, the SR group had markedly lower omissions but significantly higher error rates. This should not be interpreted as a disadvantage of SR. In FT reporting, difficult or uncertain descriptors may remain unreported because residents are unaware of their importance, unsure how to assess them, or avoid committing to an interpretation. Under SR, these same descriptors enter the reporting workflow and are actively attempted. Some are correct, improving correctness; others are incorrect, increasing the visible error rate. From an educational perspective, this shift may be valuable: an incorrectly reported descriptor can be identified and corrected through feedback, whereas an omitted descriptor remains outside the learning process altogether. In other words, SR may expose knowledge gaps that are otherwise masked by FT reporting.

The descriptor-level analysis supports the distinction between omission-driven and interpretation-driven reporting gaps.

10.12865/CHSJ.52.03.05

Descriptors such as the most cranially located suspicious lymph node, tumor deposits, and lateral lymph nodes were more frequently omitted than incorrectly reported, suggesting they may be overlooked or inconsistently incorporated into the reporting framework.

Descriptors such as T stage, cranio-caudal tumor length, and tumor localization were rarely omitted but frequently incorrect, indicating that their main limitation was interpretative accuracy rather than completeness. Peritoneal reflection assessment combined both patterns, appearing among the most-omitted and most-incorrect descriptors.

This distinction is relevant for educational planning: SR may directly address omission-driven gaps through structural prompting, while interpretation-driven gaps require targeted teaching and case-based feedback.

This finding also has implications for how experience accumulates during residency. If residents routinely omit specific staging elements in FT reports, their practical exposure to those elements remains limited even when they encounter relevant cases. SR may increase the number of meaningful interpretative decisions made during each case by forcing engagement with the full staging checklist. This may be particularly relevant for rectal MRI, where cumulative case exposure during residency is often limited.

In Phase B, omission rates were very low in both groups, and no statistically significant between-group differences were observed for any endpoint. This is a relevant practical observation: both educational pathways-SR with a mini-course and SR with individualized feedback - were compatible with near-complete omission coverage. The feedback/repeated-SR group showed a descriptively more favorable correctness/error profile in Phase B, which may reflect the advantage of iterative, case-based learning for refining descriptor-level accuracy.

This pattern is consistent with prior evidence that individualized feedback contributes meaningfully to performance improvement in rectal MRI training [23].

Reporting time findings provide a practical perspective on SR implementation. Although SR required residents to address a broader and more explicit set of descriptors, no statistically supported time penalty between groups was observed in either phase. Descriptively, slightly longer times during initial SR use may reflect novelty and accommodation to the template rather than a sustained disruption of the reporting workflow. Because the study was short and did not include repeated longitudinal use, no firm conclusion can be drawn regarding efficiency, adaptation, or long-term reporting speed. Nevertheless, the absence of a clear increase in time in this pilot setting suggests that SR can be integrated into residency practice without workflow disruption.

Although the experiment's compact design was intentional, several limitations should be acknowledged. The sample was small, particularly in terms of the number of cases per phase, and most findings are exploratory.

Similar educational reader-performance experiments in oncologic MRI have used comparable numbers of participants, but larger case sets [3]; in this study, we aimed to simulate short, realistic reporting sessions while preserving ecological validity in a voluntary training setting. In environments where rectal MRI exposure accumulates slowly, a large-volume design would risk estimating performance under conditions most residents do not reach. The trade-off was reduced statistical power, which is acknowledged throughout. The two phases used difficulty-matched but distinct case sets; between-phase comparisons were therefore restricted to the omission rate, where the influence of reporting format was considered more direct than that of case-specific interpretive difficulty. The study captures only immediate within-session reporting behavior; no delayed follow-up was conducted, and no conclusions about long-term retention or transfer to clinical practice can be drawn.

Future studies should replicate a similar compact session model in multicenter cohorts, incorporate delayed reassessment, and evaluate clinically oriented endpoints, such as treatment-relevant risk-stratification accuracy and multidisciplinary decision categories.

Conclusions

Structured reporting was associated with an immediate and consistent reduction in omission rate among radiology residents reporting primary rectal cancer MRI, without a statistically supported time penalty.

The concurrent increase in visible errors reflects an educationally relevant mechanism: SR converts omitted descriptors into explicit interpretive attempts, making knowledge gaps accessible to targeted feedback.

Both training pathways evaluated in Phase B achieved near-complete omission coverage,

suggesting that SR can be feasibly integrated into existing residency workflows.

Larger multicenter studies with delayed follow-up are needed to determine whether these immediate effects translate into durable learning outcomes.

Acknowledgement

None to declare.

Author Contributions

Conceptualization, M.A.E., C.M.C. and R.V.T.; Methodology, L.M.F. and R.V.T.; Formal Analysis, A.G.N.; Investigation, M.A.E., L.M.F. and I.A.G.; Resources, L.M.F., C.M.C., A.Ș.D. and I.A.G.; Data Curation, M.A.E. and I.A.C.; Writing - Original Draft Preparation, M.A.E. and A.G.N.; Writing - Review & Editing, L.M.F., R.V.T. and I.A.G.; Visualization, A.G.N., I.A.C. and A.Ș.D.; Supervision, I.A.G.; Project Administration, M.A.E.

All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the grant POCU/993/6/13/153178, "Research performance", co-financed by the European Social Fund within the Sectorial Operational Program Human Capital 2014-2020. The program provided administrative support and academic mentorship. The funding body had no role in the design of the study, data collection, data analysis, or writing of the manuscript.

Conflicts of interest

The authors declare no competing interests.

Institutional Review Board

This study was approved by the Ethics Committee of the University of Medicine and Pharmacy of Craiova (IRB No. 425/02.12.2025) and conducted in accordance with the Declaration of Helsinki.

Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data availability

The participant-level analytical dataset supporting the statistical analyses is included within the article. Additional anonymized data may be made available by the corresponding author upon reasonable request, subject to institutional and ethical restrictions.

References

1. Johnson TF, Brinjikji W, Doolittle DA, Nagelschneider AA, Welch BT, Kotsenas AL. Structured Head and Neck CT Angiography Reporting Reduces Resident Revision Rates. Curr Probl Diagn Radiol, 2019, 48(2):114-116.
2. Vosshenrich J, Brantner P, Cyriac J, Jadczak A, Lieb JM, Blackham KA, Heye T. Quantifying the Effects of Structured Reporting on Report Turnaround Times and Proofreading Workload in Neuroradiology. Acad Radiol, 2023, 30(4):727-736.
3. Ponsiglione A, Stanzione A, Minieri A, Musella R, D'Elia AC, Negroni D, Sacco M, Brancaccio D, Sicignano E, Muto F, Crocetto F, Cuocolo R, Imbriaco M. Impact of software-assisted structured reporting on radiology residents approaching prostate MRI. Eur J Radiol, 2025, 183:111889.
4. Beets-Tan RGH, Lambregts DMJ, Maas M, Bipat S, Barbaro B, Curvo-Semedo L, Fenlon HM, Gollub MJ, Gourtsoyianni S, Halligan S, Hoeffel C, Kim SH, Laghi A, Maier A, Rafaelsen SR, Stoker J, Taylor SA, Torkzad MR, Blomqvist L. Magnetic resonance imaging for clinical management of rectal cancer: Updated recommendations from the 2016 European Society of Gastrointestinal and Abdominal Radiology (ESGAR) consensus meeting. Eur Radiol, 2018, 28(4):1465-1475.
5. Sahni VA, Silveira PC, Sainani NI, Khorasani R. Impact of a Structured Report Template on the Quality of MRI Reports for Rectal Cancer Staging. AJR Am J Roentgenol, 2015, 205(3):584-588.
6. Nörenberg D, Sommer WH, Thasler W, D'Haese J, Rentsch M, Kolben T, Schreyer A, Rist C, Reiser M, Armbruster M. Structured Reporting of Rectal Magnetic Resonance Imaging in Suspected Primary Rectal Cancer: Potential Benefits for Surgical Planning and Interdisciplinary Communication. Invest Radiol, 2017, 52(4):232-239.
7. Alvfeldt G, Aspelin P, Blomqvist L, Sellberg N. Radiology reporting in rectal cancer using magnetic resonance imaging: Comparison of reporting completeness between different reporting styles and structure. Acta Radiol Open, 2024, 13(7):20584601241258675.
8. Ganeshan D, Duong PAT, Probyn L, Lenchik L, McArthur TA, Retrouvey M, Ghobadi EH, Desouches SL, Pastel D, Francis IR. Structured Reporting in Radiology. Acad Radiol, 2018, 25(1):66-73.
9. Kahn CE Jr, Heilbrun ME, Applegate KE. From guidelines to practice: how reporting templates promote the use of radiology practice guidelines. J Am Coll Radiol, 2013, 10(4):268-273.
10. Morgan TA, Heilbrun ME, Kahn CE Jr. Reporting initiative of the Radiological Society of North America: progress and new directions. Radiology, 2014, 273(3):642-645.
11. European Society of Radiology. ESR paper on structured reporting in radiology. Insights Imaging, 2018, 9(1):1-7.
12. KSAR Study Group for Rectal Cancer. Essential Items for Structured Reporting of Rectal Cancer MRI: 2016 Consensus Recommendation from the Korean Society of Abdominal Radiology. Korean J Radiol, 2017, 18(1):132-151.

13. Granata V, Caruso D, Grassi R, Cappabianca S, Reginelli A, Rizzati R, Masselli G, Golfieri R, Rengo M, Regge D, Lo Re G, Pradella S, Fusco R, Faggioni L, Laghi A, Miele V, Neri E, Coppola F. Structured Reporting of Rectal Cancer Staging and Restaging: A Consensus Proposal. Cancers (Basel), 2021, 13(9):2135.
14. Nougaret S, Rousset P, Gormly K, Lucidarme O, Brunelle S, Milot L, Salut C, Pilleul F, Arrivé L, Hordonneau C, Baudin G, Soyer P, Brun V, Laurent V, Savoye-Collet C, Petkovska I, Gerard JP, Rullier E, Cotte E, Rouanet P, Beets-Tan RGH, Frulio N, Hoeffel C. Structured and shared MRI staging lexicon and report of rectal cancer: A consensus proposal by the French Radiology Group (GRERCAR) and Surgical Group (GRECCAR) for rectal cancer. Diagn Interv Imaging, 2022, 103(3):127-141.
15. Canadian Association of Radiologists, 2023, Position Statement on Structured Reporting [online]. Available at: https://car.ca/wp-content/uploads/2025/03/CAR-Position-Statement-Structured-Reporting_FINAL2.pdf [Accessed 01.04.2026].
16. European Society of Radiology. ESR paper on structured reporting in radiology-update 2023. Insights Imaging, 2023, 14(1):199.
17. ESGAR Rectal Imaging Guideline Group. MRI to guide clinical management of rectal cancer: updated consensus recommendations from the European Society of Gastrointestinal and Abdominal Radiology (ESGAR)-PART I primary staging. Eur Radiol, 2026, 36(6):4592-4607.
18. ESGAR Rectal Imaging Guideline Group. MRI to guide clinical management of rectal cancer: updated consensus recommendations from the European Society of Gastrointestinal and Abdominal Radiology (ESGAR): PART II-Restaging and response evaluation. Eur Radiol, 2026, 36(6):4608-4621.
19. Reschke P, Booz C, Gruenewald LD, Koch V, Höhne E, Gökduman A, Eichler K, Schlüchtermann J, Vogl TJ, Gotta J. Rethinking radiology reports: A survey of referring physicians' perspectives. Eur J Radiol, 2025, 187:112068.
20. Lambregts DMJ, Bogveradze N, Blomqvist LK, Fokas E, Garcia-Aguilar J, Glimelius B, Gollub MJ, Konishi T, Marijnen CAM, Nagtegaal ID, Nilsson PJ, Perez RO, Snaebjornsson P, Taylor SA, Tolan DJM, Valentini V, West NP, Wolthuis A, Lahaye MJ, Maas M, Beets GL, Beets-Tan RGH. Current controversies in TNM for the radiological staging of rectal cancer and how to deal with them: results of a global online survey and multidisciplinary expert consensus. Eur Radiol, 2022, 32(7):4991-5003.
21. Golia Pernicka JS, Kim TH, Nougaret S, Bates DDB, El Homsi M, Fuqua JL 3rd, Gangai N, Horvat N, Pappou E, Paroder V, Widmar M, Petkovska I, Gollub MJ. Added value and clinical impact of second-opinion subspecialist radiologist interpretations of baseline rectal MRI in patients with rectal cancer. Eur Radiol, 2026, 36(4):2864-2875.
22. Lee S, Kassam Z, Baheti AD, Hope TA, Chang KJ, Korngold EK, Taggart MW, Horvat N. Rectal cancer lexicon 2023 revised and updated consensus statement from the Society of Abdominal Radiology Colorectal and Anal Cancer Disease-Focused Panel. Abdom Radiol, 2023, 48(9):2792-2806.
23. Bregendahl S, Bondeven P, Grønborg TK, Brown G, Laurberg S, Pedersen BG. Training of radiology specialists in local staging of primary rectal cancer on MRI: a prospective intervention study exploring the impact of various educational elements on the interpretive performance. BMJ Open Qual, 2022, 11(3):e001716.

Corresponding Author: Lucian-Mihai Florescu, Department of Radiology and Medical Imaging, University of Medicine and Pharmacy of Craiova, e-mail: lucian.florescu@umfcv.ro