Educational exchange in thyroid core needle biopsy diagnosis: enhancing pathological interpretation through guideline integration and peer learning

Agnes Stephanie Harahap¹, Chan Kwon Jung²,³

¹Department of Anatomical Pathology, Faculty of Medicine Universitas Indonesia - Dr. Cipto Mangunkusumo Hospital, Jakarta, Indonesia;
²Department of Hospital Pathology, College of Medicine, The Catholic University of Korea, Seoul;
³Cancer Research Institute, College of Medicine, The Catholic University of Korea, Seoul, Korea

Background: While fine needle aspiration cytology (FNAC) plays an essential role in the screening of thyroid nodules, core needle biopsy (CNB) acts as an alternative method to address FNAC limitations. However, diagnosing thyroid CNB samples can be challenging due to variations in background and levels of experience. Effective training is indispensable to mitigate this challenge. We aim to evaluate the impact of an educational program on improving the accuracy of CNB diagnostics. Methods: The 2-week observational program included a host mentor pathologist with extensive experience and a visiting pathologist. The CNB classification by The Practice Guidelines Committee of the Korean Thyroid Association was used for the report. Two rounds of reviewing the case were carried out, and the level of agreement between the reviewers was analyzed. Results: The first-round assessment showed a concordance between two pathologists for 247 thyroid CNB specimens by 84.2%, with a kappa coefficient of 0.74 (indicating substantial agreement). This finding was attributed to the discordance in the use of categories III and V. After peer learning, the two pathologists evaluated 30 new cases, which showed an overall improvement in the level of agreement. The percentage of agreement between pathologists on thyroid CNB diagnosis was 86.7%, as measured by kappa coefficient of 0.80. Conclusions: This educational program, consisting of guided mentorship and peer learning, can substantially enhance the diagnostic accuracy of thyroid CNB. It is useful in promoting consistent diagnostic standards and contributes to the ongoing development of global pathology practices.

Key Words: Thyroid neoplasms; Large-core needle biopsy; Practice guideline; Education
Education program

The educational program was conducted at the Department of Pathology, Seoul St. Mary’s Hospital for 2 weeks. During the program, pathologist 2 participated in observation sessions, reviewed diagnostic cases, and engaged in interactive learning sessions under the guidance of pathologist 1. The primary aim of the program was to teach standard diagnostic techniques and guidelines through hands-on case reviews, discussions on guideline-based diagnosis, and feedback sessions.

The program’s curriculum began with discussions to achieve a consensus utilizing the established guidelines by The Practice Guidelines Committee of the Korean Thyroid Association [9]. The educational materials for thyroid CNB consisted of 247 cases for education and training in the first round and 30 CNB cases for validation of educational effect in the second round. All cases with diagnostic categories IV and VI had confirmed pathologic diagnoses after surgery. Non-thyroidal lesions were immunohistochemically confirmed on biopsy specimens. Category II CNB cases without a surgically confirmed diagnosis were considered benign thyroid nodules if the size of the nodule remained stable or reduced throughout a 2-year observation period or if repeated FNAC and/or CNB yielded a benign diagnosis.

Pathologist 2 initially provided the diagnosis for all cases us-

MATERIALS AND METHODS

Participants

Two pathologists with varying backgrounds participated in an educational program. The program included a senior pathologist from The Catholic University of Korea, Seoul St. Mary’s Hospital, in Seoul (pathologist 1), and a visiting pathologist from Universitas Indonesia - Dr. Cipto Mangunkusumo Hospital in Jakarta (pathologist 2). Pathologist 1 had extensive expertise in thyroid pathology and cytopathology with 20 years of clinical practice. Pathologist 2 had 6 years of expertise in endocrine pathology from Jakarta, Indonesia. She possesses two years of experience in thyroid CNB diagnosis, handling 10 cases in the first year and 112 cases in the second year.

Digital whole slide image

All CNB glass slides were digitally scanned at a 40× magnification using a Hamamatsu NanoZoomer S360 Digital Slide Scanner (Hamamatsu Photonics, Shizuoka Prefecture, Japan). The scanned images did not contain any meta-information or slide label details that could be used to identify patients. For research purposes, each image file was assigned a new identifier, which ensured patient anonymity and data privacy. Whole slide images were viewed and diagnosed using NDPview2 Image viewing software (Hamamatsu Photonics).

Pathologist 2 reviewed the deidentified whole slide images of CNB and then examined the histologic slides of the matched tumor in a surgical specimen. To do this, we used the Philips IntelliSite Pathology Solution (Philips, Amsterdam, Netherlands), which is the primary diagnostic tool at the Department of Pathology in Seoul St. Mary’s Hospital.
ing her own diagnostic experience on the Korean CNB reporting system. The diagnostic results were then compared with those of pathologist 1. Afterward, discussions of the findings were held for each case, giving both pathologists an opportunity to provide feedback on their results, which helped to enrich the final diagnosis. Subsequently, discussions connecting the CNB findings with the gold standard post-operative results were conducted. As part of the educational program, a comprehensive validation was held at the end. This involved a second-round individual review of 30 entirely new cases.

### Statistical analysis

A statistical analysis was performed to evaluate the level of agreement between two pathologists before and after an educational program. The agreement was measured as a percentage and also by using Cohen’s kappa coefficient. The analysis was conducted by the SPSS software for Windows ver. 23.0 (IBM Corp., Armonk, NY, USA). The agreement percentage indicates the frequency of agreement between the pathologists, whereas Cohen’s kappa coefficient takes into account the agreement that may occur by chance, providing a more precise measure of concordance. An increase in agreement percentage before and after the educational program indicates an improvement in the concordance between pathologists, which was achieved through educational exchange.

### RESULTS

#### Characteristics of CNB cases

In the first round of evaluation, a total of 247 cases of thyroid CNB consisted of 30% benign thyroid lesions, 8.1% cases of non-invasive thyroid neoplasm with papillary-like nuclear features (NIFTP), 57.1% cases of thyroid cancers, 3.6% cases of non-thyroid cancers, two parathyroid lesions, and one schwannoma (Table 1). The benign thyroid lesions included follicular adenoma, oncocytic adenoma, and nodular hyperplasia. The cases of thyroid cancer included papillary thyroid carcinoma (PTC), invasive encapsulated follicular variant of PTC (IEFVPTC), follicular thyroid carcinoma, oncocytic thyroid carcinoma, poorly differentiated thyroid carcinoma, anaplastic thyroid carcinoma, medullary thyroid carcinoma (MTC), and lymphoma. Non-thyroid cancers included cases of metastases and other malignancies that did not fall within the specified group. In the second round of evaluation, a total of 30 independent cases were included, comprising 16 benign lesions, 13 cases of thyroid cancer, and one parathyroid lesion.

### Table 1. The distribution of thyroid core needle biopsy samples used for review according to the final diagnosis

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>First round (n = 247)</th>
<th>Second round (n = 30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benign</td>
<td>74 (30.0)</td>
<td>16 (53.3)</td>
</tr>
<tr>
<td>Follicular adenoma</td>
<td>49 (19.8)</td>
<td>2 (6.7)</td>
</tr>
<tr>
<td>Oncocytic adenoma</td>
<td>12 (4.9)</td>
<td>2 (6.7)</td>
</tr>
<tr>
<td>Nodular hyperplasia</td>
<td>13 (5.3)</td>
<td>12 (40.0)</td>
</tr>
<tr>
<td>NIFTP</td>
<td>20 (8.1)</td>
<td>0</td>
</tr>
<tr>
<td>Thyroid cancer</td>
<td>141 (57.1)</td>
<td>13 (43.3)</td>
</tr>
<tr>
<td>Papillary carcinoma</td>
<td>103 (41.7)</td>
<td>8 (26.6)</td>
</tr>
<tr>
<td>IEFVPTC</td>
<td>13 (5.3)</td>
<td>3 (10.0)</td>
</tr>
<tr>
<td>Follicular carcinoma</td>
<td>4 (1.6)</td>
<td>0</td>
</tr>
<tr>
<td>Oncocytic carcinoma</td>
<td>4 (1.6)</td>
<td>0</td>
</tr>
<tr>
<td>Poorly differentiated carcinoma</td>
<td>2 (0.8)</td>
<td>0</td>
</tr>
<tr>
<td>Anaplastic carcinoma</td>
<td>3 (1.2)</td>
<td>0</td>
</tr>
<tr>
<td>Medullary carcinoma</td>
<td>4 (1.6)</td>
<td>1 (3.3)</td>
</tr>
<tr>
<td>Lymphoma</td>
<td>8 (3.2)</td>
<td>1 (3.3)</td>
</tr>
<tr>
<td>Non-thyroid cancer</td>
<td>9 (3.6)</td>
<td>0</td>
</tr>
<tr>
<td>Metastatic carcinoma</td>
<td>8 (3.2)</td>
<td>0</td>
</tr>
<tr>
<td>Adenoid cystic carcinoma</td>
<td>2 (0.8)</td>
<td>0</td>
</tr>
<tr>
<td>Parathyroid</td>
<td>2 (0.8)</td>
<td>1 (3.3)</td>
</tr>
<tr>
<td>Schwannoma</td>
<td>1 (0.4)</td>
<td>0</td>
</tr>
</tbody>
</table>

Values are presented as number (%).
NIFTP, non-invasive thyroid neoplasm with papillary-like nuclear features; IEFVPTC, invasive encapsulated follicular variant of papillary thyroid carcinoma.

#### First-round assessment

The first-round assessment showed that two pathologists agreed on their diagnostic categories for 247 thyroid CNB specimens by 84.2% (Table 2). The Cohen’s kappa coefficient was 0.74, which indicates a substantial level of agreement. Pathologist 1 did not diagnose categories III and V, while pathologist 2 diagnosed category III in 16 cases (6.5%) and category V in nine cases (3.6%). These were the main causes of discordance (Figs. 1–3). In the analysis of concordance based on the final diagnostic group, the following concordance rates were found: 87.8% for benign lesions, 90% for NIFTP, 81.6% for thyroid cancers, 88.9% for non-thyroid cancer, 50% for parathyroid lesion, and 100% for schwannoma.

#### Diagnostic accuracy comparison between pathologists

We compared the CNB diagnoses made by two pathologists with the final diagnoses to determine which pathologist was more accurate. The pathologists used their own diagnostic criteria to make the CNB diagnoses. We calculated the percentage of cases in each final diagnosis group that each pathologist diagnosed. Subsequently, we aggregated these percentages to determine each pathologist’s overall accuracy. We used the following policy for categorizing the CNB diagnoses: Benign thyroid lesions
should be classified as category II, while NIFTP cases should be classified as category IV and cancers should be classified as category VI. Based on the classification policy, pathologist 1 correctly diagnosed approximately 63.9% of the relevant cases, while pathologist 2 correctly diagnosed approximately 52.9% of the relevant cases. Therefore, pathologist 1 had a higher accuracy rate in diagnosing the cases according to the given criteria, making him more aligned with the specified diagnostic expectations for benign, NIFTP, and follicular adenoma cases.

Second-round assessment

After the interactive learning sessions, a subsequent evaluation of thyroid CNB diagnosis was conducted. The two pathologists unanimously agreed to diagnose IEFVPTC, PTC, MTC, lymphoma, and parathyroid lesions. However, the agreement for nodular hyperplasia, follicular adenoma, and oncocytic adenoma was 83.3%, 50%, and 50%, respectively. During the second-round assessment, pathologist 1 consistently classified 12 cases of nodular hyperplasia as category 2, 2 cases of follicular adenoma as category 4, and two cases of oncocytic adenoma as category 4. Pathologist 2, on the other hand, diagnosed two cases of nodular hyperplasia as category 4, along with a diagnosis of category 2. Additionally, two cases of oncocytic adenoma and follicular adenoma were each diagnosed with category 2 and category 4.

Assessment of peer learning impact on diagnosis

To evaluate the impact of interactive peer learning sessions on thyroid CNB diagnosis, we focused on thyroid and parathyroid lesions, excluding metastatic cancers and other tumors (Fig. 4). During the first round, the concordance rate and Cohen’s kappa coefficient between pathologist 1 and pathologist 2 were 84.0% and 0.73, respectively. In the second round, the concordance rate and Cohen’s kappa coefficient between the same pathologists were 86.7% and 0.80, respectively (Table 3). This suggests an improvement in the level of agreement between pathologists compared to the first-round evaluation.

**DISCUSSION**

Pathology diagnosis is influenced by various factors that contribute to a degree of variability among observers. In the field of thyroid pathology, even the most widely accepted screening method, known as FNAC, continues to exhibit variability among different observers [12-14]. Diagnosing thyroid CNB samples can be challenging due to different levels of experience and exposure to the cases. This variability is common, especially when diagnosing indeterminate cases with category III, IV, or V. This is mainly due to the subtle nature of atypia, a limited amount of tumor cells, and uncertainty regarding the tumor capsule’s status in CNB [15]. The complexity of these diagnoses highlights the need for continuous education and peer learning.

We demonstrated that consistent exposure and feedback are essential for improving diagnostic skills. In the first-round assessment, our analysis revealed an overall 84.2% concordance and a substantial agreement between the two pathologists. In addition to the propensity of using categories III and V, pathologist 2 also diagnosed fewer cases as category VI. The diagnosis of category III is considered appropriate when a follicular proliferative lesion displays focal nuclear atypia, characterized by nuclear enlarge-
ment with pale chromatin, irregular nuclear membrane, and nuclear grooves amidst a predominantly benign background of follicles [9]. The challenges associated with the diagnosis of category III might be attributable to several factors. The subtlety of nuclear atypia, the presence of oncocytic nuclei, vacuoles, lymphocyte infiltration, and the occurrence of macrofollicles complicated accurate diagnosis. Artifacts introduced during the biopsy procedure, such as tissue distortion or fragmentation, further con-

Fig. 1. A case of discrepant diagnosis between two pathologists in a core needle biopsy specimen with papillary thyroid carcinoma. (A) The biopsy specimen presents a fibrotic lesion with few tumor cells. On high-power views (B, C), a few tumor cells with nuclear features of papillary carcinoma were observed. One pathologist interpreted this case as an indeterminate follicular lesion (category III), while the other pathologist diagnosed it as papillary carcinoma (category VI). (D) A matched tumor in a surgical specimen reveals paucicellular papillary carcinoma.

Fig. 2. A case of discrepant diagnosis between two pathologists in a core needle biopsy specimen with medullary thyroid carcinoma. (A) The biopsy specimen shows a tumor area with a mixed follicular and solid growth pattern, which has no fibrotic capsule and is detached from the surrounding normal thyroid tissue. (B) The tumor cells exhibit round to oval hyperchromatic nuclei and ill-defined amphophilic cytoplasm. One pathologist interpreted this case as an indeterminate follicular lesion (category III), while the other pathologist diagnosed it as medullary thyroid carcinoma (category VI). In the matched tumor in a surgical specimen (C, D), a well-defined medullary thyroid carcinoma is observed.
Over an extended period, histological assessment of the PTC nuclear features has been associated with potential interobserver disagreement, even in cases utilizing surgical specimens \[16\]. Previous studies have reported changes in nuclear features in CNB specimens \[17,18\]. Seok et al. \[17\] reported that the nuclei of PTC in CNB specimens are smaller and less irregular with less chromatin clearing than in thyroidectomy specimens. In a study conducted by Haq et al. \[18\], the nuclear morphology of CNB was compared to that of surgical specimens using a deep-learning model. The results showed a significant reduction in nuclear size, darker nuclear staining, and frequent nuclear vacuole artifacts in the CNB specimens of PTC. Although the PTC nuclear features were still detectable in microscopic appearance, they were more difficult to notice due to their smaller size and darker chromatin in the CNB specimens.

The microscopic appearance of PTC subtypes, particularly tall cell, follicular, and oncocytic, may vary between CNB and surgical specimens. The tall cell subtype is defined by the presence of more than 30% of tumor cells exhibiting a height at least three times greater than their width \[19,20\]. The presence of ‘shrunken cells’ in CNB samples poses a challenge to obtaining accurate measurements. Similarly, in cases where there is a clear-cut oncocytic appearance, the oncocytes may have enlarged nuclei, occasionally with bizarre shapes, and abundant granular eosinophilic cytoplasm. However, certain tumors displayed minimal oncocytic alterations, which posed a challenge in evaluating CNB samples. The absence of the usual prominent nucleoli observed in oncocytes further complicated the assessment. Large-sized follicles, which are more commonly encountered in nodular hyperplasia, require meticulous examination of nuclear atypia to exclude.

Category IV follicular neoplasm is diagnosed when there is an encapsulated follicular-patterned tumor, with or without nuclear atypia. The presence of a fibrous capsule in the CNB specimen plays a crucial role in determining category IV. Therefore, adequate sampling of the tumor, capsule, and adjacent non-neoplastic area is essential. Ideally, the tumor would be classified as category IV based on its microfollicular pattern, presence of atypical follicular cells distinct from adjacent normal follicles, and a portion of the fibrous capsule. In situations where a tumor is not accompanied by the capsule, it can be difficult to determine whether it should be categorized as III or IV \[9\]. This could be the reason why pathologist 2 used category III. In such cases, ultrasound images can be helpful in ensuring accurate intratumoral sampling and providing a clearer diagnosis.

The diagnosis of thyroid cancer cases can vary greatly depending on the pathologist. Pathologist 2 tends to have the most var-
ied categorical diagnoses, while non-thyroidal cancer and parathyroid lesions show the lowest agreement. Pathologist 1, on the other hand, has a more distinct diagnosis for these entities. It is possible that this is due to the small number of tumor cells in the specimens or a lower level of confidence from pathologist 2 during the initial diagnosis.

The peer learning educational program allowed pathologist 2 to engage in discussions about challenging cases and to receive immediate feedback, which contributed to an increase in consistency and concordance in CNB interpretation. The peer learning education program...
aspect of the educational program was instrumental in addressing these challenges. By participating in discussions, observing cases, and receiving immediate feedback from pathologist 1, pathologist 2 was able to refine their diagnostic skills. This collaborative approach not only improved the overall concordance rate but also emphasized the value of standardizing diagnostic criteria to minimize ambiguity.

In addition to skill enhancement, our study also highlights the importance of a robust classification system for CNB diagnoses. The Korean Thyroid Association’s guidelines provide a solid foundation for effective communication between pathologists and clinicians. After receiving appropriate training, pathologists from different backgrounds can successfully apply this diagnostic system. This framework has been proven effective in achieving accurate and reproducible diagnoses, even for those with varying levels of experience. This standardized approach plays a crucial role in reducing discrepancies and ensuring that patients receive appropriate and consistent care.

The educational program, designed to bridge these gaps, proved to be highly effective. Pathologist 2’s interaction with pathologist 1 through discussions, case reviews, and immediate feedback facilitated a significant improvement in diagnostic concordance, reducing the variability in interpretation. The peer learning model adopted in the program provided a platform for addressing challenges and refining diagnostic skills, ultimately leading to better consistency and reliability in CNB interpretation.

A notable outcome from our program was the enhanced consensus achieved between the two pathologists, reflected in the second-round assessment. The increase in Cohen’s kappa coefficient from 0.73 to 0.80 demonstrated the success of the peer learning approach in improving diagnostic accuracy. The consensus in diagnosing specific thyroid cancers, such as PTC and MTC, further validated the effectiveness of this collaborative model.

This study suggests that implementing similar educational exchanges and standardized classification systems in other areas of pathology can promote consistent diagnostic standards and reduce diagnostic discrepancies. By encouraging ongoing education, peer feedback, and standardization, the pathology community can improve the quality of diagnosis and patient care. These efforts contribute to the continued development of reliable diagnostic practices in thyroid CNB and beyond.

In conclusion, our educational program demonstrated that consistent exposure to complex cases, guided mentorship, and immediate feedback can substantially enhance the diagnostic skills of pathologists and improve the diagnostic accuracy of thyroid CNB. The variability in diagnostic outcomes, particularly in indeterminate cases, can be significantly reduced through structured educational programs and standardization of diagnostic frameworks. The success of this program in aligning diagnostic interpretations underscores the importance of collaborative learning and standardization. This approach can contribute to reducing diagnostic discrepancies and improving the quality of patient care in thyroid CNB. This model can be replicated in other areas of pathology to promote consistent diagnostic standards and facilitate knowledge exchange among professionals, contributing to the ongoing development of global pathology practices.

Ethics Statement

This study was performed in accordance with the principles of the Declaration of Helsinki and was approved by the Catholic Medical Center, The Catholic University of Korea (XC21ENDI0031K). Informed consent was waived because the study was conducted retrospectively.

Availability of Data and Material

All data analyzed during this study are included in this published article.

Code Availability

Not applicable.

ORCID

Agnes Stephanie Harahap https://orcid.org/0000-0001-8920-7873
Chan Kwon Jung https://orcid.org/0000-0001-6843-3708
Author Contributions
Conceptualization: CKJ. Data curation: CKJ. Formal analysis: CKJ, ASH. Funding acquisition: CKJ. Investigation: CKJ. Methodology: ASH, CKJ. Project administration: CKJ. Resources: CKJ. Supervision: CKJ. Validation: ASH, CKJ. Visualization: CKJ. Writing—original draft: ASH, CKJ. Writing—review & editing: ASH, CKJ. Approval of final manuscript: ASH, CKJ.

Conflicts of Interest
C.K.J., the editor-in-chief of the *Journal of Pathology and Translational Medicine*, was not involved in the editorial evaluation or decision to publish this article. The remaining author has declared no conflicts of interest.

Funding Statement
This research was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: RS-2021-KH113146).

References
3. Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid 2016; 26: 1-133.