# Experience Reviewing Digital Pap Tests using a Gallery of Images

Liron Pantanowitz[1], Sarah Harrington[2]

[1]Department of Pathology, University of Michigan, Ann Arbor, MI, USA, [2]Hologic Inc., Marlborough, MA, USA

## Abstract

**Introduction:** Hologic is developing a digital cytology platform. An educational website was launched for users to review these digitized Pap test cases. The aim of this study was to analyze data captured from this website. **Materials and Methods:** ThinPrep® Pap test slides were scanned at ×40 using a volumetric (14 focal plane) technique. Website cases consisted of an image gallery and whole slide image (WSI). Over a 13 month period data were recorded including diagnoses, time participants spent online, and number of clicks on the gallery and WSI. **Results:** 51,289 cases were reviewed by 918 reviewers. Cytotechnologists spent less time (M [Median] = 65.0 s) than pathologists (M = 82.2 s) reviewing cases ($P < 0.001$). Longer times were associated with incorrect diagnoses and cases with organisms. Cytotechnologists matched the reference diagnoses in 85% of cases compared to pathologists who matched in 79.8%. While in 62% of cases reviewers only examined the gallery, they attained the correct diagnosis 92.7% of the time. Pathologists made more clicks on the gallery and WSI than cytotechnologists ($P < 0.001$). Diagnostic accuracy decreased with increasing clicks. **Conclusions:** Website participation provided feedback about how cytologists interact with a digital platform when reviewing cases. These data suggest that digital Pap test review when comprised of an image gallery displaying diagnostically relevant objects is quick and easy to interpret. The high diagnostic concordance of digital Pap tests with reference diagnoses can be attributed to high image quality with volumetric scanning, image gallery format, and ability for users to freely navigate the entire digital slide.

**Keywords:** Artificial intelligence, cytology, digital pathology, education, Pap test, whole slide imaging

## INTRODUCTION

Digital pathology is being increasingly used by pathology laboratories worldwide, especially as some of the barriers to widespread adoption (e.g., regulations, limited interoperability, opposing pathologist mindset) are slowly being overcome. In particular, whole slide imaging (WSI) has been available for over two decades and is now being used for several clinical use cases (e.g., primary diagnosis, telepathology) as well as non-clinical applications (e.g., education, proficiency testing, research). Digital imaging has also been employed in the field of cytopathology.[1-3] However, WSI technology has been less pervasive in cytopathology because of unique challenges related to image acquisition (e.g., focus problems) and cytology workflow (e.g., screening requirements).[4-7] Focus problems arise because cytology material frequently contains three-dimensional (3D) cell clusters and obscuring material such as blood and mucus. Even liquid-based cytology (LBC) slides with monolayers of cells may still have 3D cell clustering.[8] This consequently warrants scanning cytology

slides using multiple focal planes, which can undesirably increase scan times and increase file size. Screening digital slides, typically performed by a cytotechnologist, is tedious and time consuming when using a computer mouse.[9] Hence, to be more useful in routine cytology practice a dedicated digital system that overcomes these setbacks is necessary.

Hologic are currently developing a digital cytology system that is devoted to addressing the aforementioned imaging and workflow needs in cytology. This system incorporates a WSI scanner capable of scanning cytology slides with multiple focal planes simultaneously while maintaining high scan speeds at high magnification.[10] It is expected that the digital cytology scanner will be coupled with an artificial

---

**Access this article online**

**Quick Response Code:**

**Website:**
www.jpathinformatics.org

**DOI:**
10.4103/jpi.jpi_96_20

**How to cite this article:** Pantanowitz L, Harrington S. Experience reviewing digital pap tests using a gallery of images. J Pathol Inform 2021;12:7.

Available FREE in open access from: http://www.jpathinformatics.org/text.asp?2021/12/1/7/310042

intelligence (AI)-based algorithm to analyze ThinPrep Pap test slides. The data output from this algorithm will no longer direct cytologists to use a microscope workstation to examine particular field of views on a slide, but instead will allow them to view select digital images collectively displayed in a gallery on a computer monitor. As a result, cytologists will only have to examine a gallery of images which should simplify the task of screening slides. Cytologists will still be given the option to review the entire ThinPrep cell spot (WSI) if desired. As long as this new digital approach does not compromise the accuracy of Pap test screening, the benefits of untethering cytologists from their microscopes should greatly improve workflow. To provide end users with access to the high quality digital images as an educational tool, Hologic launched a digital cytology education website (https://digitalcytologyeducation.com) globally in March 2019. The website currently includes ThinPrep Pap test cases, with new content added regularly. These online cases are comprised of all diagnostic categories and range in difficulty, allowing users with different levels of experience and background to benefit from the educational experience.

As adoption of digital pathology and AI increases, it is important to understand the acceptability and requirements of a digital cytology solution for clinical use. Knowledge of user preference, acceptability of a digital review process, and diagnostic accuracy using a gallery of images provides important information to guide development, educational and regulatory activities. Therefore, the aim of this study was to analyze at least one year's worth of data captured from Hologic's digital cytology education website.

## MATERIALS AND METHODS

### Hologic digital pathology system

ThinPrep glass slides of Pap tests stained with a ThinPrep stain were digitally scanned in full color at the equivalent of 40x magnification (0.25 μm/pixel) using a Digital Imager (whole slide scanner) designed for cytology applications (Hologic, Inc., In-development). Using a proprietary volumetric scanning technique, 14 focal planes were captured and merged into a single focal layer of optimal focus. Acquired digital slides were subsequently analyzed using an AI-based algorithm (Hologic, Inc., In-development) to detect and classify: (i) presence of a normal endocervical cell component (ECC), used for Pap test adequacy assessment, (ii) squamous and glandular epithelial abnormalities, according to The Bethesda System for reporting cervical cytology, and (iii) microorganisms (e.g., Candida, etc.).

### Education website

ThinPrep vials were acquired from US reference labs with accompanying patient demographics and lab diagnosis (cytology and histology, if available). The reference (original) diagnosis from the donating labs for abnormal cases were rendered by a pathologist. All cases were reviewed by two senior cytotechnologists. A ThinPrep slide was made at Hologic and this new glass slide was screened unblinded by a senior

cytotechnologist to confirm the reference diagnosis. The slides were then digitally scanned at Hologic. Scanned and analyzed cases were uploaded onto an education website and hosted globally through Amazon Web Services. Each case consisted of (i) a gallery of tiles (i.e., image thumbnails) containing diagnostically relevant objects (e.g., epithelial cells, organisms) that are displayed on the left of the screen, and (ii) the corresponding cell spot (WSI) displayed on the right of the screen [Figure 1]. The gallery of 24 diagnostically relevant tiles are presented at ×10 magnification. An additional 24 optional tiles containing similar objects/cells of interest were also available to the user by clicking on the tile, but only if the user wanted to see more images. Clicking any tile in the gallery zoomed to that exact area within the WSI cell spot for the user to see this specific object/cell within the context of the entire slide [Figure 2]. The user was free to lighten the image, zoom in/out, and pan around the rest of the WSI. After cases were uploaded to the website, they were reviewed and edited by a minimum of 15 senior cytotechnologists from Hologic to ensure there was diagnostic consensus and that the gallery of tiles was representative of the reference diagnosis. The cases selected were not meant to be representative of any screening population, but rather as an education website was meant to allow users to see many abnormal cases that they may not regularly see in clinical practice. The education website was optimized for the Google Chrome web browser. Users were accordingly strongly encouraged to use Chrome, but this was not enforced. Other browsers worked, but may have less functionality. To ensure good image quality, users were instructed to use a high definition monitor (preferably 27 inch) if possible. Users were also encouraged to use high speed Internet when reviewing cases.

The education website was available for anyone to register. The website was advertised at several 2019 cytology conferences and Hologic promoted the website to customers. The website assigned each user an arbitrary identification to maintain anonymity. During registration, all users consented for their data to be used for development and education purposes. Users were asked to self-declare their occupation. Additional information about the users (e.g., education, years of practice) was not captured. Once registered, users were provided with a quick reference guide explaining website functionality, and provided with a set of known reference cases. While some participants were provided with a demo, there was no formal on-site training performed. At the beginning of each week, participants were emailed a set of new cases for voluntary educational review. Cases were provided in study test sets of five, which contained a variety of diagnoses.

There was no requirement to complete the sets in order, however within each set the user could not move on to the next case without completing the prior case. There was no time limit for completing a case. To complete a case, the user entered their interpretation by selecting the diagnostic category in the review form [Figure 3]. This form required the user to enter the following information: (i) ECC (absent or present), (ii) epithelial

cell interpretation/result (e.g., unsatisfactory, negative for intraepithelial lesion or malignancy [NILM], atypical squamous cells of undetermined significance [ASC-US], low-grade squamous intraepithelial lesion [LSIL], etc.), and (iii) organisms present (Trichomonas vaginalis, Candida, bacterial vaginosis, actinomyces, herpes virus, cytomegalovirus). Notes and additional details could be added to the comment box at the bottom of the review form. Once the Complete Review form was submitted, feedback [Figure 4] was immediately provided to the user about whether their interpretation was correct compared to the reference diagnosis. This feedback also included up to four images of diagnostic cells from the case and relevant diagnostic criteria.

For each reviewer, the website software tracked the time users spent reviewing each case, the number of times they clicked on objects of interest (OOI) (i.e., images of cells) in the gallery, the number of times they clicked on the cell spot (WSI) and panned, as well as the interpretations they submitted. The amount of time(s) was recorded, which began when the reviewer opened the case and ended when the review was completed or the browser was closed. The amount of time spent entering the diagnosis and comments in the form were excluded from analysis. The interpretation of these cases was compared to the reference diagnosis to determine accuracy. Discrepancies were not adjudicated.
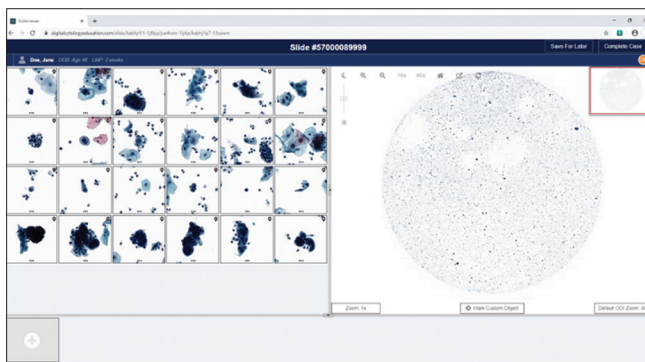
## Statistical analysis

Data collected from the education website over the course of 13 months (March 2019 to April 2020) were analyzed. Overall analyses included all reviewers, with separate comparisons of cytotechnologists and pathologists, when noted. The Pearson's correlation was used to determine if a relationship exists between the number of seconds spent on the review and the number of OOI clicks, number of WSI clicks, and the number of organisms identified. The Kendall rank correlation was used to determine if there were associations between accuracy (identified as matching diagnostic category, and separately as matching exact diagnosis) and the presence of ECC, number of OOI clicks, number of WSI clicks, final cytologic diagnosis, presence of organisms, reviewer (cytotechnologist or pathologist), as well as time spent on the review and presence of these recorded variables. The Pearson's Chi-square test was used to determine if the proportion of cases accurately diagnosed (either matching category or exact diagnostic match) was different based on ECC presence, diagnosis, organism presence, and reviewer type (cytotechnologist or pathologist). Average number of OOI clicks, WSI clicks, and seconds spent on assessment were compared using a $t$-test for comparisons with dichotomous categories (user job type, diagnostic match, organisms) and an



**Figure 1:** Screenshot from the education website showing the (left) gallery comprised of 24 image objects of interest and corresponding (right) entire ThinPrep cell spot (i.e., interactive whole slide image)
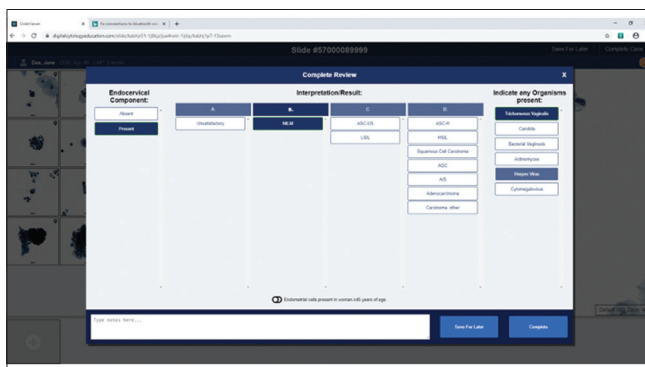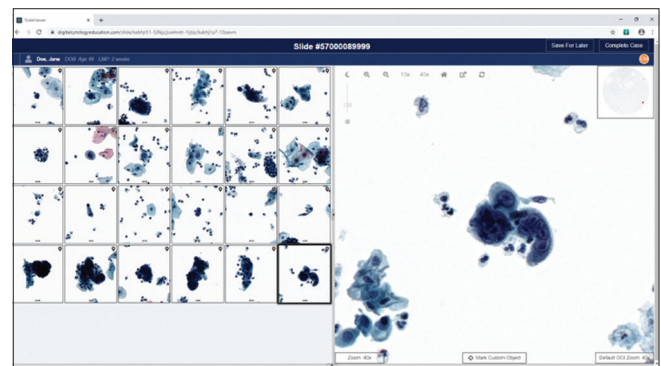


**Figure 2:** Screenshot from the education website showing (left) the bottom right cell in the gallery selected by the user being displayed (right) in much higher magnification within the context of the entire slide. This Pap test was a study case with herpes infection



**Figure 3:** Screenshot from the education website showing the review form that reviewers were asked to complete online



**Figure 4:** Screenshot from the education website showing feedback to the user. In this example the correct interpretation for the study case was negative for intraepithelial lesion or malignancy with associated trichomoniasis and herpes infection

analysis of variance for comparisons of variables with three or more categories (ECC presence, absence, or not recorded; and diagnosis). Statistical significance was assumed at $P < 0.05$. Analyses was performed using IBM SPSS Statistics 22 (IBM, New York, USA).

## RESULTS

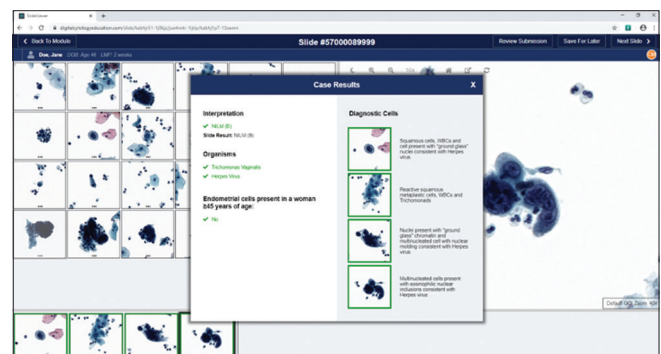A total of 51,289 digital slides were reviewed by 918 different reviewers including cytotechnologists, pathologists, and other laboratory personnel (e.g., trainees, histotechnologists, etc.). The majority of the cases were reviewed by cytotechnologists (81.5%). Pathologists reviewed 15.5% of cases and the rest (2.9%) of the participants included pathology residents (1.16%) and other types of reviewers (1.77%).

During the study time period, individual reviewers assessed between 1 and 320 cases (range) and spent between 1 second and 8 days (range) per review. The excessive review times are likely indicative of a distracted reviewer (e.g., starting a review and forgetting to close their browser before returning). Figure 5 shows the cumulative percent of time spent for diagnostic assessments. With times extending into days, the x-axis was capped at 1000 second, capturing more than 97% of recorded times. A cutoff of 300 second was determined to be a reasonable amount of time for a review even with realistic distractions, and therefore times recorded beyond 300 s were ignored for comparisons; this permitted capturing 93% of the times for cytotechnologists and 91% of times for pathologists. Most of the cases that took longer than 300 s had a higher proportion of NILM cases than those that took less than 300 s. Overall, these data demonstrate that 75% of cytotechnologists spent 100 s or less on their reviews, compared to 140 s for pathologists at the same benchmark. Cytotechnologists as a group spent overall less time reviewing cases (M = 65.0, standard deviation [SD] = 60.05) than pathologists (M = 82.2, SD = 66.74); $t(46309) = -22.088$, $P < 0.001$. A Pearson's

correlation analysis demonstrated that more time was spent on cases containing more organisms ($r = 0.038$, $P < 0.001$), more OOI clicks ($r = 0.569$, $P < 0.001$) and more WSI clicks ($r = 0.464$, $P < 0.001$). Further, longer review times were also associated with incorrect diagnoses (M = 87.9, SD = 68.21) compared to cases with an exact diagnosis match (M = 64.2, SD = 59.50); $t(47643) = 30.686$, $P < 0.001$. One-way ANOVA results showed a statistically significant difference in the amount of time spent between diagnostic groups($F(10,47634) = 151.839$, $P < 0.001$), where LSIL cases had the lowest average review time (52.1 s), and cases called unsatisfactory had the highest (113.5 s).

The overall diagnostic summary of reviews is shown in Table 1. While all of the enrolled study cases were satisfactory (per the reference diagnosis), only a few (0.1%) reviewers did feel that the cases were unsatisfactory. Table 2 shows a contingency table of cytotechnologist diagnoses. Cytotechnologists matched the reference diagnosis in 35,531 cases (85.0%), and they matched the diagnostic category in 38,768 cases (92.7%). Table 3 shows a similar contingency table for pathologists' diagnoses. Pathologists matched the reference diagnosis in 6363 cases (79.8%), and matched the diagnostic category in 7034 cases (88.2%). Cytotechnologists demonstrated a greater diagnostic match for each category as well for most of the individual Bethesda system interpretations compared to pathologists (each diagnostic category significant at $P < 0.001$) [Table 4]. However, pathologists did have a higher diagnostic match for atypical squamous cells cannot exclude high grade (ASC-H) and atypical glandular cells (AGC), but this was not statistically significant (ASC-H $\chi^2 = 0.035$, $P = 0.851$; AGC $\chi^2 = 0.380$, $P = 0.538$) [Table 5]. Most disagreements between cytotechnologists and pathologists occurred with atypical cases (i.e., ASC-US, ASC-H, AGC) and adenocarcinoma *in situ* (AIS). Pathologists interacted significantly more with the WSI (M = 7.5, SD = 29.45) in
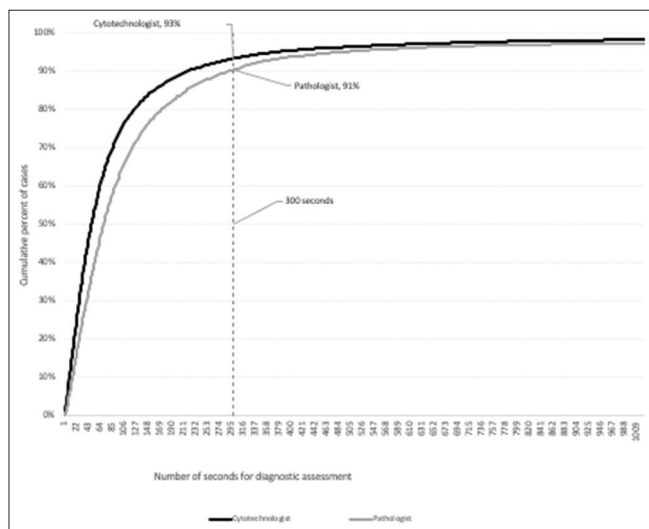


**Figure 5:** Cumulative percent of time (s) spent for online diagnostic assessment of digital Pap tests

### Table 1: Diagnostic summary of study cases

| | Reviewer diagnoses (%) | Reference diagnoses (%) |
|---|---|---|
| Unsatisfactory | 0.1 | 0.0 |
| NILM | 38.9 | 41.1 |
| ASC-US | 3.5 | 1.0 |
| LSIL | 22.3 | 22.3 |
| ASC-H | 2.1 | 0.2 |
| HSIL | 23.1 | 25.5 |
| AGC | 1.6 | 0.7 |
| AIS | 1.5 | 1.5 |
| Adenocarcinoma | 3.5 | 4.0 |
| Squamous cell carcinoma | 3.3 | 3.7 |
| Carcinoma, other | 0.1 | 0.0 |

NILM: Negative for intraepithelial lesion or malignancy, ASC-US: Atypical squamous cells of undetermined significance, LSIL: Low-grade squamous intraepithelial lesion, ASC-H: Atypical squamous cells cannot exclude high grade, AGC: Atypical glandular cells, AIS: Adenocarcinoma *in situ*, HSIL: High-grade squamous intraepithelial lesion

**Table 2: Cytotechnologist contingency table for diagnostic assessment**

| | Cytotechnologist reviewer diagnosis | | | | | | | | | | | Row total |
| | Unsatisfactory | NILM | ASC-US | LSIL | ASC-H | HSIL | AGC | AIS | Adenocarcinoma | SCC | Carcinoma other | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Reference diagnosis | | | | | | | | | | | | |
| NILM | 11 | 15,779 | 717 | 197 | 116 | 197 | 134 | 26 | 28 | 12 | 1 | 17,218 |
| ASC-US | 0 | 163 | 122 | 115 | 5 | 8 | 0 | 0 | 0 | 0 | 0 | 413 |
| LSIL | 1 | 87 | 362 | 8542 | 31 | 235 | 5 | 4 | 2 | 6 | 1 | 9276 |
| ASC-H | 0 | 6 | 10 | 3 | 20 | 48 | 2 | 0 | 0 | 1 | 0 | 90 |
| HSIL | 0 | 216 | 169 | 424 | 681 | 8574 | 139 | 87 | 108 | 254 | 3 | 10,655 |
| AGC | 1 | 40 | 5 | 3 | 22 | 86 | 65 | 12 | 66 | 7 | 2 | 309 |
| AIS | 9 | 76 | 3 | 4 | 1 | 48 | 125 | 315 | 58 | 7 | 0 | 646 |
| Adenocarcinoma | 6 | 50 | 2 | 2 | 10 | 131 | 169 | 119 | 1122 | 45 | 10 | 1666 |
| SCC | 6 | 10 | 6 | 3 | 9 | 302 | 34 | 39 | 128 | 992 | 7 | 1536 |
| Column total | 34 | 16,427 | 1396 | 9293 | 895 | 9629 | 673 | 602 | 1512 | 1324 | 24 | 41,809 |

The cells shaded in grey indicate an exact diagnostic match. NILM: Negative for intraepithelial lesion or malignancy, ASC-US: Atypical squamous cells of undetermined significance, LSIL: Low-grade squamous intraepithelial lesion, ASC-H: Atypical squamous cells cannot exclude high grade, AGC: Atypical glandular cells, AIS: Adenocarcinoma *in situ*, HSIL: High-grade squamous intraepithelial lesion, SCC: Squamous cell carcinoma

**Table 3: Pathologist contingency table for diagnostic assessment**

| | Pathologist diagnosis (hologic) | | | | | | | | | | | Row total |
| | Unsatisfactory | NILM | ASC-US | LSIL | ASC-H | HSIL | AGC | AIS | Adenocarcinoma | SCC | Carcinoma other | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Confirmed diagnosis | | | | | | | | | | | | |
| NILM | 3 | 2780 | 198 | 102 | 34 | 90 | 31 | 14 | 5 | 7 | 0 | 3,264 |
| ASC-US | 0 | 34 | 13 | 22 | 1 | 3 | 0 | 0 | 0 | 0 | 0 | 73 |
| LSIL | 1 | 65 | 70 | 1567 | 24 | 51 | 1 | 5 | 1 | 12 | 0 | 1,797 |
| ASC-H | 0 | 5 | 4 | 0 | 6 | 9 | 0 | 0 | 0 | 1 | 0 | 25 |
| HSIL | 0 | 62 | 42 | 96 | 78 | 1586 | 40 | 27 | 26 | 74 | 0 | 2,031 |
| AGC | 0 | 11 | 1 | 2 | 3 | 17 | 15 | 0 | 11 | 1 | 0 | 61 |
| AIS | 1 | 8 | 0 | 2 | 1 | 22 | 20 | 53 | 8 | 6 | 0 | 121 |
| Adenocarcinoma | 1 | 13 | 2 | 2 | 4 | 39 | 27 | 26 | 171 | 24 | 2 | 311 |
| SCC | 2 | 3 | 0 | 0 | 1 | 83 | 3 | 6 | 18 | 172 | 2 | 290 |
| Column total | 8 | 2,981 | 330 | 1,793 | 152 | 1,900 | 137 | 131 | 240 | 297 | 4 | 7,973 |

The cells shaded in grey indicate an exact diagnostic match. NILM: Negative for intraepithelial lesion or malignancy, ASC-US: Atypical squamous cells of undetermined significance, LSIL: Low-grade squamous intraepithelial lesion, ASC-H: Atypical squamous cells cannot exclude high grade, AGC: Atypical glandular cells, AIS: Adenocarcinoma *in situ*, HSIL: High-grade squamous intraepithelial lesion, SCC: Squamous cell carcinoma

**Table 4: Comparison of diagnostic matches for cytotechnologists and pathologists by diagnostic category**

| Diagnosis | Diagnostic category match | | Higher proportion | $\chi^2$ | P |
| | Cytotechnologist (%) | Pathologist (%) | | | |
|---|---|---|---|---|---|
| NILM | 91.6 | 85.2 | Cytotechnologist | 135.052 | <0.001 |
| ASC-US/LSIL | 94.3 | 89.4 | Cytotechnologist | 63.163 | <0.001 |
| ASC-H/HSIL/AGC/AIS/Ca | 92.9 | 90.9 | Cytotechnologist | 13.655 | <0.001 |
| Total | 92.7 | 88.2 | Cytotechnologist | 184.644 | <0.001 |

NILM: Negative for intraepithelial lesion or malignancy, ASC-US: Atypical squamous cells of undetermined significance, LSIL: Low-grade squamous intraepithelial lesion, ASC-H: Atypical squamous cells cannot exclude high grade, HSIL: High-grade squamous intraepithelial lesion, AGC: Atypical glandular cells, AIS: Adenocarcinoma *in situ,* Ca: Carcinoma

cases than did cytotechnologists (M = 4.1, SD = 16.01); $t(49780) = -14.841$, $P < 0.001$.

In 31,777 (62.0%) cases reviewers only examined the gallery and made no WSI clicks, of which 92.7% of these cases had the correct diagnostic category entered. In 5205 (10.1%) cases the reviewers made no OOI clicks, of which 96.4% were correctly diagnosed. When one or more OOI were clicked, reviewers' diagnostic accuracy decreased to 91.2% ($\chi^2 = 167.957$, $P < 0.001$). There was a negative, statistically significant correlation between a reviewer placing a case in the correct diagnostic category and the number of OOI clicks ($\tau_b = -0.089$, $P < 0.001$) as well as the number of WSI clicks ($\tau_b = -0.053$, $P < 0.001$).

**Table 5: Comparison of diagnostic matches for cytotechnologists and pathologists**

| Diagnosis | Exact diagnosis match | | Higher proportion | $\chi^2$ | P |
|---|---|---|---|---|---|
| | Cytotechnologist (%) | Pathologist (%) | | | |
| NILM | 91.6 | 85.2 | Cytotechnologist | 135.052 | <0.001 |
| ASC-US | 29.5 | 17.8 | Cytotechnologist | 4.256 | 0.039 |
| LSIL | 92.1 | 87.2 | Cytotechnologist | 45.220 | <0.001 |
| ASC-H | 22.2 | 24.0 | Pathologist | 0.035 | 0.851 |
| HSIL | 80.5 | 78.1 | Cytotechnologist | 6.057 | 0.014 |
| AGC | 21.0 | 24.6 | Pathologist | 0.380 | 0.538 |
| AIS | 48.8 | 43.8 | Cytotechnologist | 1.004 | 0.316 |
| Adenocarcinoma | 67.3 | 55.0 | Cytotechnologist | 17.703 | <0.001 |
| Squamous cell carcinoma | 64.6 | 59.3 | Cytotechnologist | 2.935 | 0.087 |
| Total | 85.0 | 79.8 | Cytotechnologist | 135.599 | <0.001 |

NILM: Negative for intraepithelial lesion or malignancy, ASC-US: Atypical squamous cells of undetermined significance, LSIL: Low-grade squamous intraepithelial lesion, ASC-H: Atypical squamous cells cannot exclude high grade, HSIL: High-grade squamous intraepithelial lesion, AGC: Atypical glandular cells, AIS: Adenocarcinoma *in situ*

Pathologists had significantly more OOI clicks (M = 16.8, SD = 12.70) than cytotechnologists (M = 13.0, SD = 12.59); t(49780) = −24.431, *P* < 0.001. ANOVA results identified a statistically significant difference in the average OOI clicks between the diagnoses (F (10,51278) = 206.736, *P* < 0.001); cases assessed as LSIL by the reviewer had the fewest OOI clicks (10.1), and ASC-H cases had the highest (21.6). Similarly, ANOVA results also identified a statistically significant difference in the average WSI clicks between the diagnoses groups (F (10,51278) = 18.045, *P* < 0.001); cases assessed as LSIL by the reviewer also had the fewest WSI clicks (3.5), whereas carcinoma cases had the highest (9.8).

Incorrect diagnoses were more often associated with diagnoses of lower severity (e.g., NILM, ASC-US), while more correct diagnoses were seen with diagnoses of higher severity scale (e.g., carcinoma). Overcalls were seen with NILM cases being called atypical or more severe, as well as cases in the ASC-US or LSIL diagnostic category being called ASC-H, HSIL, AGC, AIS, or carcinoma. Undercalls were seen with downgrading of ASC-US, LSIL, ASC-H, HSIL, AGC, AIS, and carcinoma cases. Pathologists had a higher proportion of overcalls (7.3%) and undercalls (4.5%) than cytotechnologists (4.1% and 3.1%, respectively).

Based on an assessment of the digital images provided in the study cases, many reviewers were able to provide an assessment about the presence or absence of the ECC. Most of the cases were identified as having the ECC present (54.5%), with 19.9% of cases showing the ECC absent. The ECC assessment was not documented by 25.6% of the reviews. Organisms were noted in 17.4% of the cases. The specific type of organisms that were identified included Candida (8.5%), *T. vaginalis* (5.8%), bacterial vaginosis (2.2%), actinomyces (0.6%), herpes virus (0.5%), and cytomegalovirus (0.04%). There was a positive, statistically significant correlation between a reviewer having a diagnostic category match and identifying the presence of organisms ($\tau_b$ = 0.028, *P* < 0.001). The results of the correlation analysis demonstrate that more time was spent on cases with organisms present ($\tau_b$ = 0.031, *P* < 0.001). In addition, reviewers made significantly more OOI clicks on cases with organisms present (M = 14.4, SD = 12.60) than with no organisms present (M = 13.5, SD = 12.71); t(51287) = −6.147, *P* < 0.001. They also clicked more on the WSI for cases with organisms present (M = 5.7, SD = 21.43) than with no organisms present (M = 4.5, SD = 18.22); t(51287) = −5.375, *P* < 0.001.

## CONCLUSIONS

The Hologic education website has been popular with reviewers. Over the 13 month study period, over 50,000 digital Pap test cases were studied by 918 reviewers. The participation of this many reviewers provided abundant data to explore how end users such as cytotechnologists and pathologists interact with a digital platform when reviewing Pap tests. It is important to point out that the intent of this study was to evaluate a digital education website and not the accuracy of the digital cytology system in a clinical setting. While most cytologists today do not routinely screen and/or sign out Pap tests digitally, many of them have become familiar with online digital content. For example, many cytologists were exposed to web-based image atlases such as those provided by the Bethesda Interobserver Reproducibility Study.[11] WSI is much more interactive than static snapshots and as such this modality has been leveraged for a variety of educational purposes.[12-14] Samulski *et al*. have shown that adaptive, self-paced eLearning modules similar to the Hologic education website can serve as an engaging and effective teaching method in cervical cytopathology.[15] Limitations of our study are that many more cytotechnologists than pathologists participated, data about participants (e.g., years in practice, practice setting) were not recorded, users were not tracked over time to document continued improvement, and the variety of displays used was not recorded. Users were also global and may not use the Bethesda system and may not all be proficient in English.

Despite this, there was high diagnostic agreement when comparing reviewers' interpretations to the reference diagnoses. This success can be attributed to a combination of the following factors: (i) high image quality due to volumetric scanning without being hampered by poor focus and artifacts, (ii) diagnostically relevant objects displayed in the gallery, and (iii) user access to freely and immediately navigate the WSI of the case to enable interpretation within the context of the entire slide. This work stems from early artificial neural network systems applied to cytology.[16,17] For example, the PAPNET cytologic screening system for quality control of cervical smears was an automated interactive instrument that employed an image classifier coupled with a trained neural network whose output displayed cells on a high-resolution video screen to be assessed by human observers.[18] Displaying images in the gallery that easily fit onto a computer screen is optimal. Showing end users more cytology images in a gallery has been proven to be detrimental as users have to scroll down on the screen and they have a tendency to call many of these cases "atypical" even if they are negative.[19] Cases reviewed by pathologists in this study were not first screened by a cytotechnologist, which differs from the typical workflow in many cytology practices and may add bias as primary cytopathologist Pap interpretation has been shown to be less accurate. Most participants were able to complete their reviews within under 2 min/case. In fact, in almost two-thirds of the cases these reviewers appeared to only have viewed the gallery. This is considerably less time than it takes to manually screen a glass slide using a conventional light microscope. However, it is plausible that this may also be due to screening educational cases which have no clinical or legal consequences. Furthermore, despite not clicking on the gallery images or reviewing the WSI in 92.7% of these instances the reviewers still obtained the correct diagnosis. These findings overall suggest that digital Pap test review when comprised of a composite image gallery and access to the adjacent corresponding WSI is quick, easy and accurate.

When analyzing the time spent by reviewers the LSIL cases were the quickest to review, taking on average only 52.1 s to diagnose. This is probably because koilocytes are easy to recognize in the image gallery, and in such cases there is little need for reviewers to further examine the WSI for verification. Longer times were associated with the subset of cases containing microorganisms. One of the complaints from cytologists about the current generation of computer-assisted screening systems for Pap tests is that the algorithms employed to analyze digital images concentrate on detecting and classifying epithelial cells and not organisms. As a result, cytologists have expressed concerns regarding missing microorganisms if these are not present within the field of views the imaging system directs them to examine microscopically. Including microorganisms within the gallery of diagnostic images therefore allows the additive capability to screen for specific organisms. Further, unlike prior digital pathology systems that have had major limitations with the diagnosis of microorganisms,[20] images acquired with the Hologic scanner were of satisfactory quality to permit the identification of several organisms.

Cytotechnologists and pathologists interacted differently with digital cases. Pathologists as a group made significantly more clicks on the images in the gallery and they also interacted much more with the WSI than did cytotechnologists. This may reflect end user distraction, serve as an indicator that the reader was struggling with the case, and/or that they were just more cautious in ensuring that they reached the correct diagnosis. When tracking how pathology users read virtual slides, Mello-Thoms *et al*. identified two types of slide exploration strategy: (i) focused, more efficient searches that typically result in a correct final diagnosis, versus (ii) a more dispersed, time-consuming search strategy that is associated with an incorrect final diagnosis.[21] Hence, not only do clicks add extra time to review a case but they may also serve as a distraction and/or an indicator that the reader may be struggling with their interpretation. Given that the reference diagnosis for abnormal Pap tests was made by a pathologist from the donating laboratory we feel that this should limit potential bias when comparing these original diagnoses to those of pathologists who participated in this study. A limitation of this study is that the size and resolution of the display, web browser and Internet connection were not controlled. Further research is needed to determine how other parameters (e.g. color, computer display, perception, ambient lighting, etc.) impact cytologist workflow, accuracy and mindset when reviewing digital Pap tests.

In summary, the educational website offered by Hologic to deliver high-quality volumetric scanned images indicates that reviewers were able to diagnose digital Pap test cases with impressive accuracy. The digital format in which these cases were presented, comprised of a gallery with 24 diagnostically relevant tiles and corresponding WSI, was also quick and easy for reviewers to use. Despite not being able to control for computer equipment used and not training end users on the digital platform or digital morphology, reviewers still from around the world exhibited very high agreement with the reference diagnosis when interpreting digital Pap test cases on Hologic's Digital Education website. Further research will be needed to determine whether similar results will be observed once a digital cytology platform for Pap test review is available for clinical use.

## Conflicts of interest
Liron Pantanowitz is a consultant for Hamamatsu and on the medical advisory board for Ibex. Sarah Harrington is an employee of Hologic.

## REFERENCES

1. Wright AM, Smith D, Dhurandhar B, Fairley T, Scheiber-Pacht M, Chakraborty S, *et al*. Digital slide imaging in cervicovaginal cytology: A pilot study. Arch Pathol Lab Med 2013;137:618-24.

2. El-Garby EA, Pantanowitz L. Whole slide imaging: Widening the scope of cytopathology. Diagn Histopathol 2014;20:456-61.

3. Capitanio A, Dina RE, Treanor D. Digital cytology: A short review of technical and methodological approaches and applications. Cytopathology 2018;29:317-25.

4. Mori I, Ozaki T, Taniguchi E, Kakudo K. Study of parameters in focus simulation functions of virtual slide. Diagn Pathol 2011;6 Suppl 1:S24.

5. Evered A, Dudding N. Accuracy and perceptions of virtual microscopy compared with glass slide microscopy in cervical cytology. Cytopathology 2011;22:82-7.

6. Donnelly AD, Mukherjee MS, Lyden ER, Bridge JA, Lele SM, Wright N, *et al*. Optimal z-axis scanning parameters for gynecologic cytology specimens. J Pathol Inform 2013;4:38.

7. Van Es SL, Greaves J, Gay S, Ross J, Holzhauser D, Badrick T. Constant Quest for Quality: Digital Cytopathology. J Pathol Inform 2018;9:13.

8. Fan Y, Bradley AP. A method for quantitative analysis of clump thickness in cervical cytology slides. Micron 2016;80:73-82.

9. Khalbuss WE, Cuda J, Cucoranu IC. Screening and dotting virtual slides: A new challenge for cytotechnologists. Cytojournal 2013;10:22.

10. Li Z, Quick M, Pantanowitz L. Whole-slide imaging in cytopathology. In: Bui MM, Pantanowitz L, editors. Modern Techniques in Cytopathology. Karger, Basel, Switzerland: Monogr Clin Cytol.; 2020. p. 84-90.

11. Kurtycz DFI, Staats PN, Chute DJ, Russell D, Pavelec D, Monaco SE, *et al*. Bethesda Interobserver Reproducibility Study-2 (BIRST-2):

12. Pantanowitz L, Szymas J, Yagi Y, Wilbur D. Whole slide imaging for educational purposes. J Pathol Inform 2012;3:46.

13. Pantanowitz L, Parwani AV. Education. In: Kaplan KJ, Rao LK, editors. Digital Pathology: Historical Perspectives, Current Concepts & Future Applications. New York, NY: Springer; 2015. p. 71-8.

14. Saco A, Bombi JA, Garcia A, Ramírez J, Ordi J. Current status of whole-slide imaging in education. Pathobiology 2016;83:79-88.

15. Samulski TD, Taylor LA, La T, Mehr CR, McGrath CM, Wu RI. The utility of adaptive eLearning in cervical cytopathology education. Cancer Cytopathol 2018;126:129-35.

16. Dytch HE, Wied GL. Artificial neural networks and their use in quantitative pathology. Anal Quant Cytol Histol 1990;12:379-93.

17. Pantanowitz L, Bui MM. Computer-assisted pap test screening. In: Bui MM, Pantanowitz L, editors. Modern Techniques in Cytopathology. Karger, Basel, Switzerland: Monogr Clin Cytol.; 2020. p. 67-74.

18. Koss LG, Lin E, Schreiber K, Elgert P, Mango L. Evaluation of the PAPNET™ cytologic screening system for quality control of cervical smears. Am J Clin Pathol 1994;101:220-9.

19. Sanghavi A, Callenberg K, Allen E, Dietz R, Pantanowitz L. Analysis of cell galleries as an interface for reviewing urine cytology cases. J Pathol Inform 2019;28:S32.

20. Rhoads DD, Habib-Bein NF, Hariri RS, Hartman DJ, Monaco SE, Lesniak A, *et al*. Comparison of the diagnostic utility of digital pathology systems for telemicrobiology. J Pathol Inform 2016;7:10.

21. Mello-Thoms C, Mello CA, Medvedeva O, Castine M, Legowski E, Gardner G, *et al*. Perceptual analysis of the reading of dermatopathology virtual slides by pathology residents. Arch Pathol Lab Med 2012;136:551-62.

Bethesda System 2014. J Am Soc Cytopathol 2017;6:131-44.