Comparing the effectiveness of two diagnostic approaches for the interpretation of oral radiographic lesions by dental students

Charmaine Ling Wei Kho¹ · Dian Yi Chow¹ · Jun Ming Wong¹ · Jin Wei Loh¹ · Yu Fan Sim¹ · Mark Joo Seng Gan² · Kelvin Weng Chiong Foong³ · Li Zhen Lim⁴

Received: 29 September 2021 / Accepted: 10 July 2022 / Published online: 29 July 2022 © The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract
Interpreting radiographic lesions on dental radiographs is a challenging process especially for novice learners, and there is a lack of tools available to support this diagnostic process. This study introduced dental students to two diagnostic aids with contrasting reasoning approaches—ORAD DDx, which uses an analytic, forward reasoning approach, and a Radiographic Atlas, which emphasizes a non-analytic, backward reasoning approach. We compared the effectiveness of ORAD DDx and the Atlas in improving students’ diagnostic accuracy and their ability to recall features of radiographic lesions. Participants (99 third-year dental students) were assigned to ORAD DDx, Atlas and Control groups. In the pre-test and post-test, participants provided their diagnosis for eight types of radiographic lesions. All groups also completed a Cued Recall Test. Feedback about ORAD DDx and the Atlas was collected. Results indicated that the Atlas was more effective than ORAD DDx in improving diagnostic accuracy (Estimated marginal mean difference = 1.88 (95% CI 0.30–3.46), \( p = 0.014 \), Cohen’s \( d = 0.714 \)). Participants in the Atlas group also outperformed the Control group in the recall of the lesions’ radiographic features (Estimated marginal mean difference = 3.42 (95% CI 0.85–5.99), \( p = 0.005 \), Cohen’s \( d = 0.793 \)). Students reported that both ORAD DDx and Atlas increased their confidence and decreased the mental effort required to develop differential diagnosis (\( p \leq 0.001 \)). This study demonstrates the effectiveness of a non-analytic approach in interpreting dental radiographs among novice learners through the novel use of diagnostic aids.

Keywords Diagnostic reasoning · Oral radiology · Dental education · Image interpretation · Diagnostic aids

*Li Zhen Lim
limlizhen@gmail.com

1 Faculty of Dentistry, National University of Singapore, Singapore
2 Centre for Development of Teaching and Learning, National University of Singapore, Singapore
3 Discipline of Orthodontics and Paediatric Dentistry, Faculty of Dentistry, National University of Singapore, Singapore
4 Discipline of Oral and Maxillofacial Surgery, Faculty of Dentistry, National University of Singapore, 9 Lower Kent Ridge Road, Singapore 119085
Introduction

Various intraosseous lesions may be detected on dental radiographs, ranging from inflammatory lesions to cysts, benign tumours, and even malignancies. Within the spectrum of these lesions, dentists may have difficulty with diagnosing them accurately due to insufficient exposure to such conditions (Stheeman et al., 1996; White, 1989). This difficulty is even greater for dental students who lack sufficient knowledge and experience in developing differential diagnoses (Kratz et al., 2018). The importance of training dental students in developing accurate differential diagnoses, which impacts treatment decisions and patient outcomes, cannot be overstated (Rohlin et al., 1995).

The considerations involved in the diagnostic process can be overwhelming to novice learners when encountering an unknown lesion. Oral radiology textbooks are often structured by disease classifications and categories, rather than by radiographic appearance. One existing resource that may help to bridge this gap is the Oral Radiographic Differential Diagnosis (ORAD) program (https://www.orad.org/). ORAD is based on probabilistic/Bayesian calculations using conditional probabilities—the likelihood that a specific pathology would have a specific radiographic feature and the prevalence of the pathology in the target population. However, inaccuracies of these calculations can arise from the lack of prevalence data, which results in the use of estimates and variations of disease prevalence among different populations (White, 1989).

Another promising development that has garnered significant interest is the field of artificial intelligence (AI). However, while there have been advances in AI in diagnosing dental caries, root fractures, apical lesions, salivary gland diseases, maxillofacial cysts and tumours (Heo et al., 2020; Putra et al., 2022), these tools are still currently not widely accessible.

As such, there is a need for alternative materials that students can refer to for a group of lesions to consider, as the first step in the process of developing their differential diagnoses. Knowledge of the existing current frameworks surrounding diagnostic thinking is essential to the design of these materials. There are two main approaches that guide clinical reasoning (Eva, 2005; Kok et al., 2017). Analytic reasoning is based on the careful, stepwise correlation of signs, symptoms and diagnoses. In contrast, non-analytic reasoning is based on the comparison of the current case with past cases, allowing pattern recognition to be the dominant mode of diagnosis (Eva, 2005; Norman & Brooks, 1997).

A related framework is the dual process theory, which posits that there are two different modes of processing—System 1 (non-analytic) and System 2 (analytic) processes. While System 1 is known to be faster and uses less effort to arrive at a clinical decision, it is also thought to be more prone to intrinsic biases that may lead to diagnostic error. On the other hand, System 2 engages rule-based, forward reasoning. While this may take more effort in deduction, it follows the rule of logic and is hence thought to be “error-free”. However, this also implies that System 2 errors may present to be more consequential as it confers greater level of confidence in the diagnosis arrived (Croskerry, 2009; Evans, 2008).

ORAD DDx (https://www.dentistry.nus.edu.sg/orad-ddx/) was developed at our institution, using information about lesion features from a textbook, “Oral Radiology: Principles and Interpretation” (White & Pharoah, 2014). It is a logical/deductive system based on an analytic/System 2 approach that produces a list of possible differentials based on inputs/filters (radiographic features) that users select. This manual input of lesion descriptors would require students to go through a checklist of lesion features as part of the diagnostic process, and would impose a features-based, forward reasoning framework on the users.
A demonstrated advantage of checklists is the ability to confirm preliminary diagnoses or suggest alternatives, resulting in improved diagnostic accuracy (Abimanyi-Ochom et al., 2019). An instantaneously updated output would demonstrate how the list of differentials is affected as the filters are toggled.

A contrasting approach for a diagnostic aid would be built upon a backward reasoning framework. The findings published by Baghdady et al. have demonstrated how non-analytic strategies now challenge the traditional analytic approaches to learning oral radiology (Baghdady et al., 2014). In their study, students who were instructed to employ a non-analytic/System 1 approach had higher diagnostic accuracy than those who employed the analytic/System 2 approach. For our study, we considered the use of an atlas. An image-based resource would support a non-analytic/System 1 approach by providing reference images which a student could visually compare with their unknown lesion to help them to derive their differential diagnosis. Apart from being a resource to aid in pattern recognition, an atlas may also serve as a substitute for a “knowledge bank” that experts would otherwise gain through experience over time. An example of such a resource would be the “Atlas of Oral and Maxillofacial Radiology” (Koong, 2017).

While atlases have not been previously studied in oral radiology, they have been reported in other visual fields such as dermatology and medical radiology. A photographic atlas of severe bullous skin disorders helped non-experts reach 75% agreement with experts in their diagnosis, although comparisons with unaided decision-making were not made (Bastuji-Garin et al., 1993). Use of dermatology atlases have also been reported (Eysenbach et al., 1998; Kamel Boulos, 2006). In medical radiology, a pictorial atlas of nuclear imaging findings increased diagnostic accuracy and decreased interobserver variability among observers in differentiating parkinsonian and non-parkinsonian patients (Goethals et al., 2009). A study among medical students also showed that visual comparison techniques were useful for learning interpretative skills when reading chest radiographs (Kok et al., 2015).

The goal of our study was to compare the effectiveness of two resources that utilized opposing strategies, in helping dental students to improve their accuracy in radiographic diagnosis and recall of lesion features. In addition, we wanted to gather insights as to how each resource aided in the student’s diagnostic process, and whether further improvements could be made. Evidence of the effectiveness of either method could encourage educators to incorporate the respective tools and approaches as part of their teaching armamentarium. This study aimed to answer the research questions outlined below.

**Research Questions**

1. What are the differences in the effectiveness of two diagnostic aids (ORAD DDx and Radiographic Atlas) based on two contrasting approaches, in improving the accuracy of students’ differential diagnosis of oral radiographic lesions and their ability to recall lesion features?
2. How did each resource help students with deriving differential diagnosis? How can either one be improved?

Based on the existing literature (Baghdady et al., 2014), we hypothesize that the Radiographic Atlas (backward reasoning approach) would outperform ORAD DDx (forward reasoning approach).
Participants

This study received an exemption approval from the National University of Singapore’s Institutional Review Board (NUS-IRB S-19-010E).

Sample size calculation was performed based on the primary objective of comparing the improvement in test scores between the study groups. Using Cohen’s large effect size ($d = 0.8$), a total of 102 participants (34 per group) was required for 5% significance level and 80% statistical power. The calculated sample size was inflated by the Bonferroni correction method to ensure an appropriate number of subjects for post-hoc comparisons. 100 third-year dental students were recruited over two cohorts through convenience sampling. With Cohen’s large effect size ($d = 0.8$), the statistical power was slightly reduced to 78%. One participant’s results was excluded due to technical difficulties with using the ORAD DDx tool, leaving a total of 99 participants (Fig. 1). All participants were undergoing an Oral Radiology module at the time of the study.

Materials and procedures

Selection of oral radiographic lesions

Eight oral radiographic lesions with similar or potentially confusing presentations were chosen. The lesions selected were:

- **Radiolucent** Apical rarefying osteitis, Simple bone cavity

![Study design and recruitment of third-year dental students over two cohorts](image)

Fig. 1  Study design and recruitment of third-year dental students over two cohorts
Comparing the effectiveness of two diagnostic approaches for...

- **Radiopaque** Dense bone island, Condensing osteitis, Odontoma
- **Radiolucent, mixed or radiopaque** Osteomyelitis, Periapical osseous dysplasia and Florid osseous dysplasia.

### Preparation of ORAD DDx

ORAD DDx was modified to display only the eight selected lesions, and the number of filters was reduced to 6 (https://www.dentistry.nus.edu.sg/orad-ddx-urop/). The chosen filters and options were:

- Internal density (Radiolucent/Mixed/Radiopaque)
- Border definition (Well-defined/Ill-defined)
- Border cortication (Yes/No)
- Encapsulation within soft tissue border or PDL space (Yes/No)
- Association with a single tooth periapex (Yes/No)
- Number (Single/Multiple).

ORAD DDx was also validated using the radiographic images that were selected for the pre- and post-tests to ensure that the correct diagnosis would be included as one of the possible options when the appropriate filters were applied for each case.

### Selection of radiographic images for Radiographic Atlas, and pre- and post-tests

Radiographic images were sourced from (1) Teaching files, (2) Case reports from open access publication portals, and (3) Online educational portals, such as MedEdPortal (Velez et al., 2014), Radiopaedia (https://radiopaedia.org/) and Eurorad (https://www.eurorad.org/). As far as possible, radiographic images of lesions with confirmed histopathological diagnoses were used. If no biopsy was performed, two oral radiologists agreed on the final diagnosis of a case based on the radiographs, and this was used as a “silver standard”.

For the Radiographic Atlas, two radiographs were selected for each diagnosis. In the pre-test (without aids) and post-tests (with aids for intervention groups), a total of 15 radiographs were used (one of condensing osteitis, and two for each of the remaining seven diagnoses). Different radiographic images were used in the pre-test, post-test and Atlas, and were not repeated.

### Study design

For the pre-test, the 15 radiographic images were loaded onto an online survey platform, Qualtrics (https://www.qualtrics.com/, Provo, UT). For each image shown, participants selected their top differential diagnosis from a list of eight drop-down options. The sequence of images was randomized during testing.

The pre-test scores were recorded and sorted from lowest to highest. The “alternate ranks” design assignment method (Dalton & Overall, 1977) was used to assign the students into three groups: Control, ORAD DDx and Atlas. This ensured that the distribution of pre-test scores would be similar between the groups.

Following the pre-test, participants completed the interventions in their own time. Control group participants were only provided with the correct diagnosis to the 15 images used
in the pre-test. Participants in the intervention groups were asked to review the same 15 images again, but using their respective diagnostic aids.

The type of feedback provided differed for each intervention group. For ORAD DDx, the correct features were listed along with the correct diagnosis/answer in each case. If students did not provide the correct diagnosis/answer, they were instructed to correct any features they incorrectly identified on ORAD DDx and to observe how the list of possible diagnoses changed (Fig. 2). For the Atlas group, the correct diagnosis/answer was provided in each case. If participants did not provide the correct diagnosis/answer, they were instructed to visually compare the image in the question with the correct diagnosis’ reference images on the Radiographic Atlas (Fig. 3).

The post-test was carried out between one to three weeks after the intervention. A new set of 15 radiographic images were selected. The post-test was conducted in the same manner as the pre-test, except that participants in the intervention groups (ORAD DDx and Atlas) were allowed to refer to their respective diagnostic aids. Control group participants did not use any aids.

All participants also completed a Cued Recall Test (CRT), which consisted of 24 True/False questions about the radiographic features of the eight lesions. The questions that were developed for the CRT were based on information from a textbook (White & Pharoah, 2014), and were reviewed for clarity and accuracy by two oral radiologists. Participants were not allowed to refer to the diagnostic aids during this test. A negative marking system was used. One point was awarded or deducted for each correct or incorrect answer. If the "I don’t know" option was selected, no marks were awarded or deducted.

Lastly, participants in the ORAD DDx and Atlas groups provided feedback on their respective interventions through an online survey. Feedback questions assessed the usability of the aid using a System Usability Scale (SUS), and participants’ perception of whether the aids influenced their confidence in their diagnosis and mental effort used. Qualitative feedback was also gathered from the survey. Open-ended questions asked what participants liked and did not like about the tools, challenges faced and suggestions for improvement.
Comparing the effectiveness of two diagnostic approaches for...

Analyses

Statistical analysis was performed using STATA 15 (StataCorp. 2017. Stata Statistical Software: Release 15. College Station, TX). Descriptive statistics was used to summarize the data. Pre-test scores for the 3 study groups were compared using one-way ANOVA test. Assumptions for normality and homogeneity of variance were checked using Shapiro–Wilk test and Levene’s test respectively. Linear regression models were built separately for improvement in test score and CRT to study the intervention effect while adjusting for the cohort effect. Estimated marginal mean with 95% confidence interval was reported for improvement in diagnostic accuracy. To compare the difference between study groups, pairwise comparison with Bonferroni correction was conducted. Interaction effect between cohort and study groups was also tested using Wald test. To assess the confidence and mental effort with and without the use of aids, Pearson’s chi-square test was performed. Statistical significance was set at $p$-value $<0.05$.

Qualitative analysis was conducted using thematic analysis of the student responses to the four open-ended questions to gain a deeper understanding of the students’ perspectives on the use of the two resources. Two of the team members read the participants’ responses to identify common themes, which was then used to code the data when the responses were re-read iteratively. These categories were verified by a senior researcher, who also counted the frequency at which they appeared in the responses. Any disagreement in coding was resolved by further discussion among all team members.
Results

Comparing the improvement in diagnostic accuracy scores within and across groups

Table 1 shows the pre-test and post-test scores for both cohorts. The post-test scores were significantly higher than the pre-test scores within all groups ($p < 0.001$). Within both cohorts, there was no significant difference in pre-test scores across groups ($p = 0.875$ and $p = 0.973$).

Analysis of the ORAD DDx participants’ performance on the post-test revealed that among the 152 incorrect diagnoses out of 480 responses, 80 of these misdiagnoses resulted from participants incorrectly selecting a radiographic feature filter that would have excluded the correct diagnosis by the ORAD DDx tool. In the other 72 misdiagnoses, participants were presented with the correct diagnosis as one of the possible differentials, but the participant decided on a different (and incorrect) diagnosis.

Table 2 provides the improvement scores from a combined-cohort analysis. There was a significant difference in the performance of the two classes. The class of 2021 had higher improvement scores ($p = 0.022$, Cohen’s $d = 0.469$) than the class of 2020. Adjusting for the difference between the two cohorts, the average improvement in the pre-test to post-test scores were 1.24 (95% CI 0.33–2.15), 1.99 (95% CI 1.06–2.91), 3.86 (95% CI 2.97–4.76) for the Control, ORAD DDx and Atlas groups respectively. Using the estimated marginal mean differences, the Atlas group’s improvement was significantly higher compared to both the Control group at 2.62 (95% CI 1.06–4.19), $p < 0.001$, Cohen’s $d = 0.997$, and ORAD DDx group at 1.88 (95% CI 0.30–3.46), $p = 0.014$, Cohen’s $d = 0.714$. There was no significant interaction effect between the two cohorts and study groups ($p = 0.408$).

Comparing cued recall test (CRT) scores across groups

Table 3 reports the CRT scores. There were statistically significant differences across the study groups ($p = 0.006$). After adjusting for differences between the two cohorts, the Atlas group outperformed the Control group at 3.42 (95% CI 0.85–5.99), $p = 0.005$, Cohen’s $d = 0.793$. There were no significant differences across the remaining comparisons. There was no significant interaction effect between the two cohorts and study groups ($p = 0.279$).

Comparing feedback for ORAD DDx and Atlas

System usability scale (SUS)

The usability of the ORAD DDx and Atlas tools was compared using the System Usability Scale (Brooke, 1996). Estimated marginal mean SUS scores for ORAD DDx and Atlas were 75.15 (95% CI 70.95–79.36) and 73.90 (95% CI 69.82–77.98) respectively, both of which are interpreted as having “Good” usability (Bangor et al., 2009). There was no significant difference in SUS scores between both interventions ($p = 0.670$).
### Table 1 Pre-test and post-test diagnostic accuracy scores for Control, ORAD DDx and Atlas groups across both cohorts

| Class of 2020 | Control | ORAD DDx | Atlas | F-statistics | $\eta^2_p$ | $p$-value$^a$ | Class of 2021 | Control | ORAD DDx | Atlas | F-statistics | $\eta^2_p$ | $p$-value$^a$ |
|---------------|---------|----------|-------|--------------|------------|--------------|---------------|---------|----------|-------|--------------|------------|--------------|
| Pre-test      | 8.63 (2.94) | 8.38 (2.73) | 8.13 (2.50) | F (2,45) = 0.13 | 0.006 | 0.875 | 8.00 (2.42) | 8.19 (2.71) | 8.00 (2.85) | F (2,48) = 0.03 | 0.001 | 0.973 |
| Post-test     | 8.94 (3.49) | 9.50 (1.93) | 11.88 (2.53) | **F (2,45) = 5.23** | **0.189** | **0.009*** | 10.12 (3.00) | 11.00 (1.90) | 12.00 (1.85) | F (2,48) = 2.91 | 0.108 | 0.064 |
| $p$-value$^b$ | < **0.001*** | < **0.001*** | < **0.001*** | | | | < **0.001*** | < **0.001*** | < **0.001*** | |

Mean and Standard Deviation (in brackets) values are reported

$^a$Comparing test score across groups

$^b$Comparing pre-test and post-test score within groups

*p-value < 0.05 (bolded)
Table 2  Mean Improvement between pre-test and post-test diagnostic accuracy scores, and pairwise comparisons across groups

| Group | Estimated marginal mean (95% CI) | η²p |
|-------|----------------------------------|-----|
| Control | 1.24 (0.33–2.15) | 0.157 |
| ORAD DDx | 1.99 (1.06–2.91) | |
| Atlas | 3.86 (2.97–4.76) | |
| Class | 0.054 |
| Class of 2020 | 1.75 (0.99–2.50) | |
| Class of 2021 | 2.98 (2.25–3.71) | |

*(a) Estimated marginal mean of improvement in test scores from linear model*

| Group     | Estimated marginal mean difference (95% CI) | p-value | Cohen's d |
|-----------|--------------------------------------------|---------|-----------|
| ORAD DDx vs Control | 0.75 (−0.85–2.34) | 0.769 | 0.283 |
| Atlas vs Control | 2.62 (1.06–4.19) | <0.001* | 0.997 |
| Atlas vs ORAD DDx | 1.88 (0.30–3.46) | 0.014* | 0.714 |
| 2021 vs 2020 | 1.23 (0.18–2.28) | 0.022* | 0.469 |

*(b) Pairwise comparison of improvement in test scores*

CI Confidence Intervals
*p-value < 0.05 (bolded)*

Table 3  Mean Cued Recall Test scores and pairwise comparisons across groups

| Group     | Estimated marginal mean (95% CI) | η²p |
|-----------|----------------------------------|-----|
| Control | 9.36 (7.87–10.86) | 0.100 |
| ORAD DDx | 11.29 (9.77–12.80) | |
| Atlas | 12.79 (11.32–14.26) | |
| Class | 0.003 |
| Class of 2020 | 10.91 (9.67–12.15) | |
| Class of 2021 | 11.40 (10.20–12.60) | |

*(a) Estimated marginal mean of test scores from linear model*

| Group     | Estimated marginal mean difference (95% CI) | p-value | Cohen's d |
|-----------|--------------------------------------------|---------|-----------|
| ORAD DDx vs Control | 1.92 (−0.68–4.53) | 0.226 | 0.446 |
| Atlas vs Control | 3.42 (0.85–5.99) | 0.005* | 0.793 |
| Atlas vs ORAD DDx | 1.50 (−1.09–4.09) | 0.486 | 0.347 |
| 2021 vs 2020 | 0.49 (−1.24–2.21) | 0.577 | 0.113 |

*(b) Pairwise comparison of test scores*

CI Confidence Intervals
*p-value < 0.05 (bolded)*

 позволит отобразить текст документа в чистом виде, как если бы вы читали его естественным образом.
Confidence levels and mental effort

There were statistically significant differences when comparing confidence levels and mental effort required to develop differential diagnoses with and without the diagnostic aids ($p \leq 0.001$) (Fig. 4). Students from Atlas and ORAD DDx groups both reported increased confidence and decreased mental effort when using their respective tools. However, there was no significant difference between ORAD DDx and Atlas on participants’ confidence and mental effort ($p = 0.701$ and $p = 0.496$ respectively).

Qualitative feedback

As both cohorts provided similar comments about the diagnostic aids, the feedback from both classes are presented together in Table 4. Major themes that emerged from the responses included the ease and difficulty of use; knowledge pre-requisites prior to using the tools; and suggestions for improvements.

Ease and difficulty of use

The first common theme was the ease of use. Both tools were reported to be simple to use and facilitated the formation of the list of differential diagnosis in an efficient manner. ORAD DDx users found it straightforward to select filters and receive suggested
Table 4 Qualitative feedback about ORAD DDx and Radiographic Atlas organised by themes

| Ease of use | ORAD DDx | Atlas |
|-------------|----------|-------|
| Simplicity  | 20 out of 32 participants commented that the tool was simple or easy to use. They felt that selecting filters and receiving suggested differentials was a straightforward process | 19 out of 34 participants commented that the Atlas was easy or simple to use. They were able to come up with an answer efficiently and quickly by comparing images. Another 5 participants mentioned the visual nature of the tool |
| Ease of diagnosis | 4 participants also liked that the tool helped them to narrow down their choices when making differential diagnoses | |
| Difficulty of use | | |
| Laborious | 1 participant found that the test cases were straightforward, and the use of the tool actually slowed down the diagnostic process (when the participant appeared to have a preferred diagnosis) | 2 participants commented that the atlas would be tedious and difficult to use if more lesions were added in the future, while another felt that even with only 8 lesions, scrolling was already inconvenient |
| Applicability | 2 participants found it difficult to select the appropriate filters for lesions with mixed features (i.e. partially well-defined and partially ill-defined) Another commented that they identified the features wrongly, resulting in a wrong differential diagnosis | 6 participants found that the atlas was not useful when faced with atypical presentations of lesions, since the atlas did not include the entire spectrum of possible appearances |
| Knowledge pre-requisite | | |
| Prior knowledge required | 2 participants stated that (even with the narrowed list of differentials) it still required some background knowledge to pick their top differential | 1 participant pointed out that pre-existing knowledge about possible variations was still required to avoid falling into the trap of [visually] comparing images without thinking [critically] |
| Conflict with pre-existing knowledge | 4 participants commented that the tool did not give them the diagnosis that they wanted, or that “didn’t seem right”, or that viable diagnoses were left out. This could be due to selecting the wrong features that excluded the correct diagnosis | |
| Suggestions for improvements | | |
| Preference for descriptors | 3 participants suggested that the tool should provide descriptions of each lesion (i.e. commonly found locations in the jaw, association with crown of tooth, etc.) | 17 participants stated that descriptions of the features or characteristics of the lesions (especially salient ones) would be useful in their diagnosis |
| Preference for more radiographic images | 6 participants felt that adding radiographic images as visual aids would be helpful for reference | 18 participants suggested providing more images for each lesion, which could also accommodate for variations in presentations |
| Other supplements                                      | ORAD DDx                                                                 | Atlas                                                                                                                                 |
|------------------------------------------------------|--------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|
|                                                      |                                                                          | 1 participant suggested including differential diagnoses associated with each lesion                                               |
|                                                      |                                                                          | Another participant felt that textbooks could provide more information about aetiology and pathophysiology, which correlates |
|                                                      |                                                                          | with radiographic appearances, and would aid in memory recall                                                                   |
differentials, and they liked that the tool helped them to narrow down their answer choices. Other positive comments included “systematic”, “convenient”, “comprehensive”, “quick” and “useful”. Atlas users felt that they could come up with answers quickly by comparing images.

The second theme that emerged from the feedback, though less prevalent than the former, was the difficulty of use. A few participants felt that the tools could be laborious—having to select individual filters on ORAD DDx may slow down the diagnostic process, while having more lesions in the Atlas in the future may result in it becoming more tedious to use. Variations in lesion appearances that could not be accounted for was also a concern for the applicability of both tools. In the case of the ORAD DDx, participants were uncertain about the correct choice of filters when the lesion had mixed presentations (i.e. both partially well-defined and ill-defined borders), or felt that the wrong identification of features resulted in a wrong differential diagnoses. For the Atlas, participants felt that the resource was inadequate for lesions with atypical presentations as only two examples of each lesion was included.

Knowledge pre-requisites

The third theme revolved around knowledge pre-requisites for using the tools. Participants commented that prior knowledge may still be required in order to pick the top differential from the suggested list on ORAD DDx. This could be due to the lack of supplemental information or images provided on ORAD DDx. A few participants also mentioned that conflicts with their pre-existing knowledge arose when the suggested diagnosis did not match their initial perception of what the lesion was. With the Atlas, one participant remarked that it was necessary to have knowledge of possible variations in lesion presentations as visual comparisons alone may be misleading.

Suggestions for improvement

Participants from both groups reported that they would prefer to have more descriptions of the features or characteristics of the lesions. Participants from the ORAD DDx group expressed that including radiographic images would be helpful for reference, while participants from the Atlas group also suggested having more images for each lesion to accommodate for variations in presentations. One participant from the Atlas group proposed including the differential diagnoses associated with each lesion. Although the next point was answered under “what I did not like [about the tool]”, its incorporation could also enhance the learning value of the Atlas. This insight was that information about aetiology and pathophysiology, which likely can be correlated with radiographic appearances, could aid with memory recall.

Discussion

To our knowledge, this is the first study comparing two different diagnostic approaches with the use of diagnostic aids in oral radiology.
Comparison of diagnostic approaches

It should be noted that in the second run of this study, the COVID-19 pandemic outbreak occurred in between the pre-test and post-test study phases. This is one possible explanation for the class of 2021’s superior performance over the previous cohort (Table 2). As all clinical activities were cancelled and students were required to stay home, increased time spent preparing for upcoming examinations could have resulted in higher post-test scores across all groups. Nevertheless, the statistical analysis performed in this study adjusted for the differences between the two cohorts, such that the data interpretation could focus on the different diagnostic approaches.

The use of the Atlas tool encouraged students to form a diagnosis based on their first impression, as would be the case in a non-analytic diagnostic strategy. In our study, the participants in the Atlas group outperformed both the ORAD DDx and control groups in diagnostic accuracy, and the control group in memory recall. This result is consistent with previous studies, where participants using a backward reasoning/diagnosis-directed approach outperformed those who used a forward reasoning/feature-based approach (Baghdady et al., 2014; Norman et al., 1999).

Norman and colleagues found that among psychology students with no prior knowledge in electrocardiogram (ECG) interpretation, the group that used a non-analytic approach demonstrated improved diagnostic accuracy compared to the group that used an analytic approach. (Norman et al., 1999). Similarly, in a study conducted amongst dental students and dental hygiene students, the authors varied diagnostic strategies (analytic versus non-analytic) and instructional methods (basic science method versus structured algorithm). They found that students who were explicitly instructed to employ the non-analytic approach had improved diagnostic accuracy compared to those who were instructed to use the analytic approach (Baghdady et al., 2014).

The success of the Atlas tool in this study also complements previous work on comparison learning in complex visual tasks. Kok et al., studied the effect of comparison on diagnostic performance when interpreting chest radiographs under four conditions: (1) cases of the same disease, (2) cases of different diseases, (3) cases of disease against normal images and (4) identical images (control group) (Kok et al., 2015). While the overall scores were not significantly different across groups, the authors found that the second group (where participants were trained with pairs of images of different diseases) had the best diagnostic performance when efficiency was taken into consideration. They proposed that placing images of different conditions next to each other was helpful for students in identifying the discriminating features of each disease (Kok et al., 2015).

The Atlas group also outperformed the Control group in the CRT, designed to test memory recall. One theoretical framework that could explain this result is the chunking theory, first studied in chess, which describes memory as a form of pattern recognition (Chase & Simon, 1973). This was later extended to the template theory, which additionally accommodated for “slots” where additional information could be added and is thought to contribute to superior expert memory (Gobet, 1998; Gobet & Simon, 1996). We propose that the use of the Atlas provides opportunities for increased visual registration and processing of lesion appearances, allowing the formation of a mental representation in their “mind’s eye” of how a lesion should appear (Wood, 1999). When applied to oral radiology, the template theory could also allow for subtle variations in one or more features of the lesion.

In contrast, there was no significant difference in diagnostic accuracy between the ORAD DDx and control groups. In the ORAD DDx group’s post-test results, 80 of the
152 misdiagnoses resulted from participants incorrectly selecting a radiographic feature filter that would have excluded the correct diagnosis by the ORAD DDx tool. The binary (or ternary) nature of the filters on ORAD DDx may have further compounded this. Students offered in their feedback that there was no room for ambiguity in feature identification. However, they may have been unaware of the ability to toggle between options to assess the impact of selecting different filters on the list of differentials, or leaving filters blank. The latter option would have generated a longer list of differentials for them to consider.

It is also important to note that ORAD DDx is not intended to provide the “best” diagnosis, but rather, suggestions of differential diagnoses to consider. Users must still be able to justify their final diagnosis based on previously gained knowledge. Correspondingly, among the remaining 72 misdiagnoses, ORAD DDx had provided the correct diagnosis as one of the possible options. The participants’ final decision contributing to an incorrect diagnosis was therefore unrelated to the use of the ORAD DDx tool, apart from the fact that the tool did not narrow their options down to a single diagnosis.

Furthermore, the underlying cognitive mechanism of image interpretation is more complex than individual feature recognition alone. Polanyi, in his work on tacit knowledge, states “We know a person’s face, and can recognize it among a thousand, indeed a million. Yet we usually cannot tell how we recognize a face we know…” (Polanyi, 1966). This analogy has been applied in the field of pathology—a pathologist does not focus on a cell’s nuclei or cytoplasm, but recognizes cancerous tissue in its totality (Heiberg Engel, 2008). Similarly, developing differential diagnoses from radiographs requires processing of the entire image rather than individual features.

Conversely, analysing an entity feature by feature may adversely impact diagnostic accuracy by disrupting the coherent mental representation of the disease (Baghdady et al., 2014). In addition, there is a risk that novices using the analytic approach may find themselves overwhelmed by features that are irrelevant to the correct diagnosis due to their inability to impose appropriate frameworks on the data gathered (Ark et al., 2006; Norman et al., 1999).

Feedback on ORAD DDx and Atlas

Both the ORAD DDx and Atlas tools scored higher than 70 on the SUS scale, corresponding to ‘Good’ on an adjective rating scale (Bangor et al., 2009). These scores may be helpful in later studies as benchmarks for comparison for successive iterations of each tool. The scores may also be useful when evaluating the usability of radiographic aids with similar functions.

The finding that ORAD DDx and Atlas helped students to increase their confidence and decrease mental effort in developing differential diagnosis is unsurprising. This data can be triangulated with the open-ended feedback that was collected. The use of ORAD DDx allowed participants to narrow down their list of differential diagnoses quickly, while the Atlas provided an efficient means for them to decide on the correct diagnosis through visual comparisons. Both of these aids would help decrease the cognitive load that otherwise would be required in generating differential diagnoses independently (Gunderman et al., 2001). Additionally, any confirmation of participants’ preliminary diagnosis by either tool might also have contributed to an increase in confidence in their final diagnosis.

We might not be able to extrapolate the decrease in mental effort reported in this study if there was an increase in the scope of lesions included. Using the Atlas would naturally
become more cumbersome with greater numbers of lesions if no further categorization system was introduced. Similarly, for ORAD DDx, if additional filters were implemented to make the tool more specific, the effort of filling in the filters may exceed the savings of mental effort from using the tool. However, the benefit of ORAD DDx may be more significant with the introduction of more lesions, and it could help participants to weigh the possibility of more uncommon lesions that they might otherwise not have considered.

Interestingly, in the open-ended feedback collected, comments from ORAD DDx participants such as “I didn’t get the diagnosis I wanted” or that “[the tool] slows down the diagnosis [diagnostic] process” alluded to the fact that even novices were able to form their initial impression of an image by gestalt. This speaks to the dominance of the diagnosis-based approach in image interpretation, which is consistent with the main findings of this study.

Limitations

This study is not without limitations. In both cohorts, we conducted this study after the students had completed their Oral Radiology didactics, but before their final examinations. There is a possible confounding effect of participants studying relevant course material in their own time. This could explain why even the control groups demonstrated significantly better post-test scores compared to the pre-test. As described earlier, this could also explain the difference in performance between the two cohorts. We attempted to control for this confounding effect by limiting the time span between the intervention and post-test to a maximum of 3 weeks, and through the use of a control group.

While some may argue that the pre-test introduces bias in the study design since the participants become familiarized with the testing format (Cook & Beckman, 2010), pre-testing served various purposes in this study. Firstly, the interventions were integrated with the pre-test materials. Secondly, pre-test scores were used for group allocation so that the baseline mean scores across all groups were similar. Lastly, comparing the improvement scores (which is also evaluating the difference-in-difference), allowed us to compare the effect of each intervention, while accounting for the possible confounding effect described above and for changes due to factors other than the interventions.

Another limitation is that the study only used eight types of oral radiographic lesions, and the findings may be different if the number of lesions was increased. Finally, clinical information about each case (such as patient demographics and presenting signs and symptoms) were withheld from the participants. While this helped to focus the study on image interpretation, it may not be reflective of the diagnostic reasoning process that would occur in real life.

Implications and future directions

An important implication from this study is the potential for enhancing radiology education among dental students by introducing adjunctive tools. While the evaluation of both diagnostic aids yielded insightful quantitative and qualitative results, it appears that an emphasis on a backward reasoning approach through the use of a radiographic atlas is especially promising.
Future directions may include integrating the two aids into a single system. In an initial study using ECGs, Ark et al. found that novices who were directed to use only the analytic approach performed no better than those using the non-analytic approach exclusively (Ark et al., 2006). However, those instructed to use elements of both in a combined approach outperformed those using only a single approach, suggesting that both approaches were complementary rather than mutually exclusive (Ark et al., 2006, 2007). This suggests the possibility of tapping the synergies of the analytic and non-analytic approaches by fusing the ORAD DDx and Atlas tools.

A proposed application of this combined tool could be the broad use of the filters on ORAD DDx on distinctive or discriminating features (i.e. Radiolucent/Mixed/Radiopaque, or Well-defined/Ill-defined lesion borders) to narrow down the list of lesions to consider. Subsequently, the remaining lesions could be displayed as an atlas, allowing users to apply a comparative approach between the lesion in question with the atlas images displayed. This may improve the efficiency of searching for appropriate images for comparison and to serve as a substitute for exemplars that novices have not yet encountered in their own experiences (Norman et al., 2007). The combined ORAD DDx/Atlas tool, in addition to serving as a diagnostic aid before students arrive at a diagnosis, could also be retrospectively applied on some of the radiographs to encourage students to reconsider their diagnoses and strengthen metacognition.

**Conclusion**

The Radiographic Atlas (utilizing a non-analytic, backward reasoning approach) was more effective than ORAD DDx (using an analytic, forward reasoning approach) in improving the diagnostic accuracy of eight types of oral radiographic lesions among third-year dental students. Participants in the Atlas group also outperformed the Control group in recalling the lesions’ radiographic features. Students reported that both ORAD DDx and Atlas increased their confidence and decreased the mental effort required to develop differential diagnosis. We encourage educators to explore and use diagnostic aids to complement cognitive reasoning approaches to enhance students’ learning in oral radiographic diagnosis.

**Acknowledgements** The authors would like to sincerely thank the participants from Faculty of Dentistry, National University of Singapore, Classes of 2020 and 2021 for their time and participation in this project.

**Author contributions** All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by CLWK, DYC, JMW, JWL, YFS and LZL. The first draft of the manuscript was written by CLWK, DYC, JMW and JWL. The manuscript was edited by LZL. All authors read and approved the final manuscript.

**Funding** ORAD DDx was developed by the corresponding author, and was funded by the Technology in Dental Education Grant from the Faculty of Dentistry, National University of Singapore (FOD, NUS).

**Declarations**

**Conflict of interest** The authors have no financial interests to declare.

**Ethical approval** This study received an exemption approval from the National University of Singapore’s Institutional Review Board (NUS-IRB S-19-010E).
Other declarations  This research project was presented at Undergraduate Research Opportunities Programme Day at FOD, NUS in 2019, and was awarded first place. CLW Kho, DY Chow, JM Wong and JW Loh also received the Outstanding Undergraduate Researcher Prize from NUS in 2020 for their work on this study.

References

Abimanyi-Ochom, J., Bohingamu Mudiyanselage, S., Catchpool, M., Firipis, M., Wanni Arachchige Dona, S., & Watts, J. J. (2019). Strategies to reduce diagnostic errors: A systematic review. *BMC Medical Informatics and Decision Making, 19*(1), 7–11. https://doi.org/10.1186/s12911-019-0901-1

Ark, T. K., Brooks, L. R., & Eva, K. W. (2006). Giving learners the best of both worlds: Do clinical teachers need to guard against teaching pattern recognition to novices? *Academic Medicine, 81*(4), 405–409. https://doi.org/10.1097/00001888-200604000-00017

Ark, T. K., Brooks, L. R., & Eva, K. W. (2007). The benefits of flexibility: The pedagogical value of instructions to adopt multifaceted diagnostic reasoning strategies. *Medical Education, 41*(3), 281–287. https://doi.org/10.1111/j.1365-2929.2007.02688.x

Baghdady, M. T., Carmanah, H., Lam, E. W. N., & Woods, N. N. (2014). Dental and dental hygiene students’ diagnostic accuracy in oral radiology: Effect of diagnostic strategy and instructional method. *Journal of Dental Education, 78*(9), 1279–1285.

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies, 4*(3), 114–123.

Bastuji-Garin, S., Rzany, B., Stern, R. S., Shear, N. H., Naldi, L., & Roujeau, J. (1993). Clinical classification of cases of toxic epidermal necrolysis, Stevens-Johnson syndrome, and erythema multiforme. *Archives of Dermatology, 129*(1), 92–96.

Brooke, J. (1996). SUS: A “Quick and Dirty” usability scale. In *Usability evaluation in industry* (pp. 207–212). CRC Press. https://doi.org/10.1201/9781498710411-35

Chase, W. G., & Simon, H. A. (1973). The mind’s eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215–281). Elsevier. https://doi.org/10.1016/B978-0-12-170150-5.50011-1

Cook, D. A., & Beckman, T. J. (2010). Reflections on experimental research in medical education. *Advances in Health Sciences Education, 15*(3), 455–464. https://doi.org/10.1007/s10459-008-9117-3

Croskerry, P. (2009). Clinical cognition and diagnostic error: Applications of a dual process model of reasoning. *Advances in Health Sciences Education, 14*(1 SUPPL), 27–35. https://doi.org/10.1007/s10459-009-9182-2

Dalton, S., & Overall, J. E. (1977). Nonrandom assignment in ANCOVA. *The Journal of Experimental Education, 46*(1), 58–62. https://doi.org/10.1080/00220973.1977.11011611

Eva, K. W. (2005). What every teacher needs to know about clinical reasoning. *Medical Education, 39*(1), 98–106. https://doi.org/10.1111/j.1365-2929.2004.01972.x

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*, 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Eysenbach, G., Bauer, J., Sager, A., Bittorf, A., Simon, M., & Diepgen, T. (1998). An international dermatological image atlas on the WWW: Practical use for undergraduate and continuing medical education, patient education and epidemiological research. *Studies in Health Technology and Informatics, 52*(Pt 2), 788–792.

Gobet, F. (1998). Expert memory: A comparison of four theories. *Cognition, 66*(2), 115–152. https://doi.org/10.1016/S0010-0277(98)00020-1

Gobet, F., & Simon, H. A. (1996). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology, 31*(1), 1–40. https://doi.org/10.1006/cogp.1996.0011

Goethals, I., Ham, H., Dobbeleir, A., Santens, P., & D’Asseler, Y. (2009). The potential value of a pictorial atlas for aid in the visual diagnosis of 123I FP-CIT SPECT scans TT - Potenzieller Nutzen eines Bildatlas zur Unter stützung der visuellen Diagnostik von 123I FP-CIT-SPECT-Scans. *Nuklearmedizin, 48*(04), 173–178.

Gunderman, R., Williamson, K., Fraley, R., & Steele, J. (2001). Expertise: Implications for radiological education. *Academic Radiology. https://doi.org/10.1016/S1076-6332(03)80708-0*

Heiberg Engel, P. J. (2008). Tacit knowledge and visual expertise in medical diagnostic reasoning: Implications for medical education. *Medical Teacher. https://doi.org/10.1080/01421590802144260*
Heo, M. S., Kim, J. E., Hwang, J. J., Han, S. S., Kim, J. S., Yi, W. J., & Park, I. W. (2020). Dmfr 50th anniversary: Review article artificial intelligence in oral and maxillofacial radiology: What is currently possible? Dentomaxillofacial Radiology. https://doi.org/10.1259/dmfr.20200375

Kamel Boulos, M. N. (2006). Map of dermatology: “first-impression” user feedback and agenda for further development. Health Information and Libraries Journal, 23(3), 203–213. https://doi.org/10.1111/j.1471-1842.2006.00660.x

Kok, E. M., de Bruin, A. B. H., Leppink, J., van Merriënboer, J. J. G., & Robben, S. G. F. (2015). Case comparisons: An efficient way of learning radiology. Academic Radiology, 22(10), 1226–1235. https://doi.org/10.1016/j.acra.2015.04.012

Kok, E. M., van Geel, K., van Merriënboer, J. J. G., & Robben, S. G. F. (2017). What we do and do not know about teaching medical image interpretation. Frontiers in Psychology, 8, 309. https://doi.org/10.3389/fpsyg.2017.00309

Koong, B. (2017). Atlas of oral and maxillofacial radiology. Wiley. https://doi.org/10.1002/9781118939604

Kratz, R. J., Nguyen, C. T., Walton, J. N., & MacDonald, D. (2018). Dental students’ interpretations of digital panoramic radiographs on completely edentate patients. Journal of Dental Education, 82(3), 313–321. https://doi.org/10.1018/jde.018.033

Norman, G. R., & Brooks, L. R. (1997). The non-analytical basis of clinical reasoning. Advances in Health Sciences Education. https://doi.org/10.1023/A:1009784330364

Norman, G. R., Brooks, L. R., Colle, C. L., & Hatala, R. M. (1999). The benefit of diagnostic hypotheses in clinical reasoning: experimental study of an instructional intervention for forward and backward reasoning. Cognition and Instruction, 17(4), 433–448. https://doi.org/10.1207/S1532690XC1704_3

Norman, G., Young, M., & Brooks, L. (2007). Non-analytical models of clinical reasoning: The role of experience. Medical Education, 41(12), 1140–1145. https://doi.org/10.1111/j.1365-2923.2007.02914.x

Polanyi, M. (1966). The Tacit Dimension (p. 4). Chicago and London: The University of Chicago Press, Revised ed. edition (May 1, 2009).

Putra, R. H., Doi, C., Yoda, N., Astuti, E. R., & Sasaki, K. (2022). Current applications and development of artificial intelligence for digital dental radiography. Dentomaxillofacial Radiology. https://doi.org/10.1259/dmfr.20210197

Rohlin, M., Di, O., Hirschmann, P. N., & Matteson, S. (1995). Global trends in oral and maxillofacial radiology education. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology and, 80(5), 517–526. https://doi.org/10.1016/S1079-2104(95)80151-9

Stheeman, S. E., Mileman, P. A., van’t Hof, M., & van der Stelt, P. F. (1996). Room for improvement? The accuracy of dental practitioners who diagnose bony pathoses with radiographs. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontics, 81(2), 251–254. https://doi.org/10.1016/s1079-2104(96)80425-2

Velez, I., Tamara, L., Hogge, M., & Gonzalez, T. (2014). Guide to the diagnosis of jaw lesions according to radiographic manifestations. MedEdPORTAL, 10(1), mep_2374-8265.9731. https://doi.org/10.15766/mep_2374-8265.9731

White, S. C. (1989). Computer-aided differential diagnosis of oral radiographic lesions. Dentomaxillofacial Radiology, 18(2), 53–59. https://doi.org/10.1259/dmfr.18.2.2699592

White, S. C., & Pharoah, M. J. (Eds.). (2014). Oral radiology: Principles and interpretation (7th ed.). Mosby Inc.

Wood, B. F. (1999). Visual expertise. Radiology, 211(1), 1–3. https://doi.org/10.1148/radiology.211.1.99ap431

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.