

Construct Validity and Expert Benchmarking of the Haptic Virtual Reality Dental Simulator

Siriwan Suebnukarn, D.D.S., Ph.D.; Monthalee Chaisombat, D.D.S.;
Thanapohn Kongpunwijit, D.D.S.; Phattanapon Rhienmora, M.Eng., Ph.D.

Abstract: The aim of this study was to demonstrate construct validation of the haptic virtual reality (VR) dental simulator and to define expert benchmarking criteria for skills assessment. Thirty-four self-selected participants (fourteen novices, fourteen intermediates, and six experts in endodontics) at one dental school performed ten repetitions of three mode tasks of endodontic cavity preparation: easy (mandibular premolar with one canal), medium (maxillary premolar with two canals), and hard (mandibular molar with three canals). The virtual instrument's path length was registered by the simulator. The outcomes were assessed by an expert. The error scores in easy and medium modes accurately distinguished the experts from novices and intermediates at the onset of training, when there was a significant difference between groups (ANOVA, $p < 0.05$). The trend was consistent until trial 5. From trial 6 on, the three groups achieved similar scores. No significant difference was found between groups at the end of training. Error score analysis was not able to distinguish any group at the hard level of training. Instrument path length showed a difference in performance according to groups at the onset of training (ANOVA, $p < 0.05$). This study established construct validity for the haptic VR dental simulator by demonstrating its discriminant capabilities between that of experts and non-experts. The experts' error scores and path length were used to define benchmarking criteria for optimal performance.

Dr. Suebnukarn is Associate Professor, Faculty of Dentistry, Thammasat University, Pathumthaini, Thailand; Dr. Chaisombat is a dentist, Faculty of Dentistry, Thammasat University, Pathumthaini, Thailand; Dr. Kongpunwijit is a dentist, Faculty of Dentistry, Thammasat University, Pathumthaini, Thailand; and Dr. Rhienmora is a faculty member, School of Science and Technology, Bangkok University, Bangkok, Thailand. Direct correspondence and requests for reprints to Dr. Siriwan Suebnukarn, Faculty of Dentistry, Thammasat University, Pathumthaini, Thailand 12121; siriwan.suebnukarn@gmail.com.

Keywords: dental education, clinical education, endodontics, virtual reality, haptic simulation, Thailand

Submitted for publication 12/21/13; accepted 3/17/14

Clinical training is one of the most challenging areas for dentistry because the development of competence requires acquisition of problem-solving ability and clinical skills including technical skill in performing a procedure and patient communication. Currently, training towards clinical competence follows an apprenticeship approach, which consists of close expert supervision while practicing with preclinical simulations and interacting with patients.¹ This method of training may subject patients to discomfort, risk of complications, and prolonged procedure times.² At the same time, there may be limited access to more complex cases with corresponding difficulty in training in a time-effective manner.³

With recent advances in three-dimensional imaging and virtual reality (VR) technology, training in dentistry is moving towards an interactive environment.⁴ VR refers to a human-computer interface that facilitates interactive visualization and control of computer-generated three-dimensional cases and their related environment with sufficient detail and

speed so as to evoke a sensorial experience close to that of a real experience. There are three types of VR technologies: non-immersive systems use a standard high resolution monitor of the desktop system to display the virtual environment and interact with the user through keyboards or other 3D interaction devices; semi-immersive projection systems comprise a relatively high performance graphics computing system; and fully immersive head-mounted display systems provide the most direct experience of virtual environments.⁵ The haptic systems include tactile information, generally known as force feedback in medical applications.^{6,7} VR technology permits computed three-dimensional images obtained from medical imaging systems to be transformed into patient-specific anatomic models, with physical properties added, providing interaction with the manipulation of the realistic models.⁸⁻¹⁰

VR programs have been developed in dental applications such as a simulator for orthognathic surgery¹¹ and restorative dentistry.^{12,13} Several haptic VR dental simulators have been introduced into den-

tal education for clinical skill acquisition in several tasks: the Iowa Dental Surgical Simulator¹³ for teaching tactile skills in detecting dental caries; PerioSim¹⁴ for teaching tactile skills in detecting subgingival calculus; a haptic VR simulator with repetitive training developed by Yamaguchi et al.¹⁵ for teaching caries removal and periodontal pocket probing skills; and HapTEL¹⁶ for online dental training. The advantages of these simulators are that dental students and professionals are able to practice the procedures as many times as they want at no incremental cost and that the training can take place anywhere. The realism of these simulators has increased with the implementation of haptic devices that provide tactile sensations to students, utilizing indirect viewing as well as simulation of different tissue hardness. Haptic devices allow students to feel dental tools and oral tissues in a virtual environment and to perform clinical procedures such as pushing, pulling, and cutting tissue with realistic force feedback.

With increasing interest in simulation programs, one important step in evaluating new training tools is to assess validity. Validity is the property of being true, correct, and in conformity with reality. For the haptic VR dental simulator, this means: does it measure what it is designed to measure? Validation is comprised of a number of principles. To accomplish the different parts of validation, several expert benchmarking criteria have been developed to assess the validity of a testing system. These include content validity, construct validity, face validity, concurrent validity, discriminate validity, and predictive validity.^{17,18}

The aim of our study was twofold: to demonstrate construct validation of the haptic VR dental simulator and to analyze the feasibility of setting the skill level of experts on the simulator as benchmarking. Construct validity refers to the concept that the haptic VR dental simulator is able to measure the quality, ability, or trait it was designed to measure. Therefore, the parameters assessed must be related to the level of the performers' experience. This usually is accomplished by measuring the performance of groups that differ in precise skills being measured by the simulation. For example, practicing expert dentists should outperform inexperienced dental students. Establishing benchmarking criteria by experts is an important step in being able to measure the ability of the haptic VR dental simulator to train and improve the skills of the novice. Basic dental training must be formative and feasible, and it must encourage novice dentists to reach an optimal level of performance.

Methods

This study was approved by the Thammasat University Institutional Ethical Review Board. The participants were dental students and endodontists at the Thammasat University Faculty of Dentistry in Pathumthaini, Thailand. Those who agreed to participate and who met the inclusion criteria (who had experience performing endodontic access opening on 1, 2, 3, and 4 canal types from endodontic preclinical course) were enrolled. Any who had prior experience with the haptic VR dental simulation were not included in the study. All participants gave their informed consent.

The participants were divided into three groups according to their experience in endodontics: fourteen novices (fifth-year dental students who had performed endodontic access opening only in the laboratory), fourteen intermediates (sixth-year dental students who had performed access opening in at least five cases on patients), and six experts (endodontists). All participants were given reading and multimedia material, demonstration videos, and a demonstration of the simulator and were briefly instructed on the use of the system and the requirements of access opening. The participants received a verbal explanation about use of the system from the investigators and familiarized themselves with the system interface, but not with the task, for fifteen minutes. During this familiarization, each participant was allowed to ask questions and receive further verbal explanation and suggestions from the investigators.

The haptic VR dental simulator consists of a 2.8-GHz Pentium 4 PC, with 256 MB RAM and a thirteen-inch computer monitor, connected to two haptic devices (SensAble Inc., Woburn, MA, USA), each having five degrees of freedom.¹⁹ The haptic VR dental simulator displays the position and the movements of the virtual dental handpiece and mouth mirror in real time. Radiographic images of extracted human mandibular premolar, maxillary premolar, and mandibular molar, of which the original patients could not be identified, were acquired using 3D micro-computed tomography (RmCT, Rigaku Co., Tokyo, Japan) with a resolution of 50x50x50 μm . A 3D reconstruction of the tooth was performed using 600 of the 2D images processed by the volumetric rendering method.

All participants were asked to perform a total of thirty trials of endodontic access opening using the haptic VR dental simulator with different levels of

difficulty on each of three consecutive days (Figure 1). The trials were organized as follows: ten trials in the easy mode (mandibular second premolar with one root canal) on the first day; followed by ten trials in the medium mode (maxillary first premolar with two root canals) on the second day; and ten trials in the hard mode (mandibular first molar with four root canals) on the third day. There was no time limit for each trial. After each trial, a short break was mandatory. Based on the task completion times from a previous study,¹⁹ all participants were expected to finish ten trials of each task in no more than two hours. The task was designed to mimic real access opening and to require hand-eye coordination for quality performance. The number of tasks was chosen to provide a significant variety of exercises covering the basic endodontic skills. To ensure the same conditions for all participants, the position of the task displayed on the monitor and the haptic instruments were identical for each participant.

The parameters of performance explored in this study—error scores and instrument's path length—were chosen based on a previous study.¹² The error scores were randomly graded by one endodontist who did not participate in the data collection (Figure 2). Four-point scales¹⁹ were used to evaluate errors in buccal wall, lingual wall, mesial wall, distal wall, and pulpal floor: 0=minimally extended cavity affording unimpeded access to and visibility of the orifices of all canals present; 1=cavity permitting effective debridement of the canal system without prejudice to subsequent restoration; 2=incomplete removal of pulp chamber roof and/or inadequate retention form for the maintenance of an effective dressing; and 3=unidentified canals and/or perforation. The maximum total error score was 15. Figure 2 shows some examples of the outcome and error score: 0 for the top-left sample, and 6 for the bottom-right sample (3 from distal wall, 2 from mesial wall, and 1 from buccal wall). The haptic VR dental simulator registered motions of the instrument's path length (total path followed by the instrument on cutting in centimeters) (Figure 3).

A one-way analysis of variance was conducted on the first trial to identify initial differences among the three groups. Dynamic differences between trial series (before and after the plateau was reached) of each group were analyzed using analysis of variance for repeated measures with experience-level group as the between-subjects factor. Additional Tukey HSD post hoc tests were applied when significant

effects were encountered. All data were analyzed with the Statistical Package for the Social Sciences version 20.0 (SPSS, Chicago, IL, USA). Statistical significance was defined as a p-value less than 0.05.

Results

Enrolled participants were assigned to one of three groups according to their level of experience at the time of the study: novices (n=14), intermediates (n=14), and experts (n=6). All participants completed ten trials of the three procedural tasks (easy, medium, and hard).

Easy Mode: Mandibular Premolar

At trial 1, the three groups started their training at different error score and path length at the easy mode (Figure 4, upper left). The mean error score of experts (mean=2.02) was significantly lower than that of intermediates (mean=4.51; $p=0.011$), which was in turn lower than that of novices (mean=7.71; $p=0.013$). This trend was consistent until at least trial 6. From trial 7 on, the three groups achieved similar scores and could not be distinguished.

These observations were confirmed by comparison of two subsets of trials: on trial series 1 through 6, there was a statistically significant difference between groups ($F_2=12.3$; $p<0.05$). Contrast was observed on all paired comparisons (novices vs. intermediates $p=0.02$; intermediates vs. experts $p=0.01$; novices vs. experts $p<0.01$). On analysis of trial series 7 through 10, performance of all groups was comparable ($F_2=0.9$; $p=0.54$). No specific contrast was established (novices vs. intermediates $p=0.81$; intermediates vs. experts $p=0.73$; novices vs. experts $p=0.54$). By trial 10, all participants improved significantly: mean novice error score 3.2 (95 percent CI 2.8-3.4); mean intermediate error score 3.1 (95 percent CI 2.9-3.3); and mean expert error score 1.4 (95 percent CI 0.7-1.8). No significant difference was found between groups at the end of training ($F_2=0.2$; $p=0.71$).

Medium Mode: Maxillary Premolar

The mean error score of experts (2.13) was significantly lower than that of intermediates (4.01; $p=0.023$), which was in turn lower than that of novices (5.48; $p=0.036$) (Figure 4, middle left). This trend was consistent until at least trial 4. From trial 5 on, the three groups achieved similar scores and

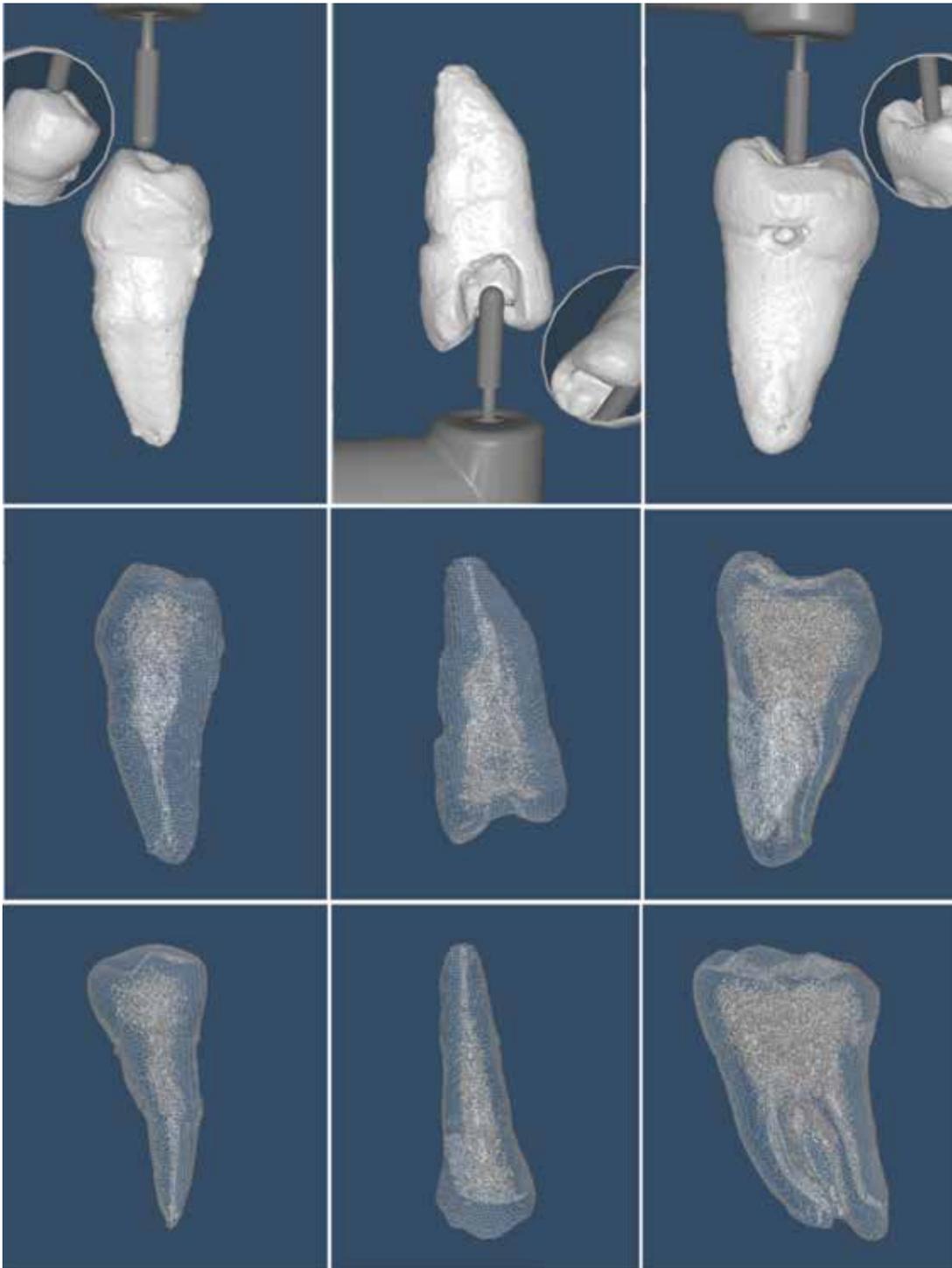


Figure 1. Virtual reality environment of easy (mandibular second premolar with one root canal), medium (maxillary first premolar with two root canals), and hard modes (mandibular first molar with four root canals) (left, middle, and right)

Note: The pulpal anatomy of each tooth in bucco-lingual and mesio-distal view is shown in the wireframe representation. The maxillary first premolar has a Class II preparation.

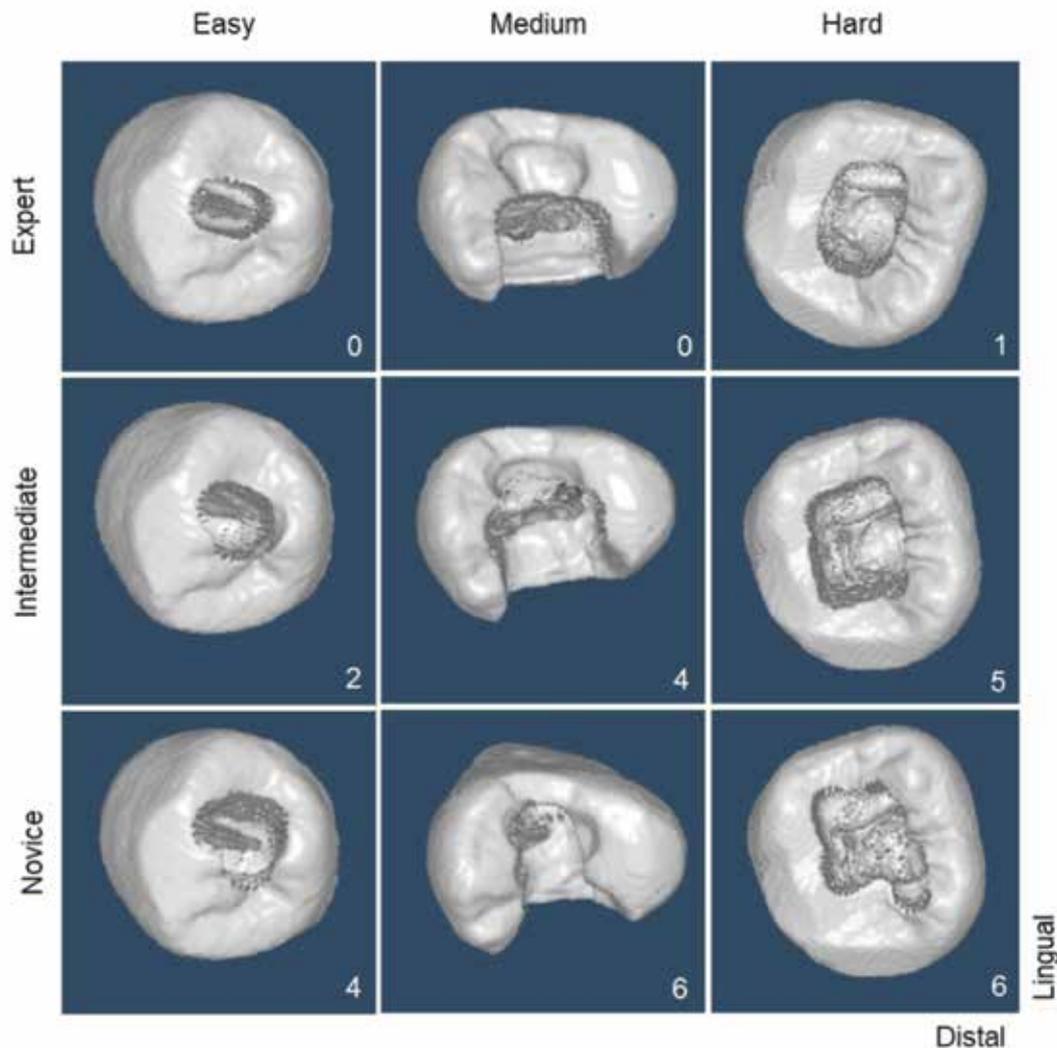


Figure 2. Examples of outcome performed by expert, intermediate, and novice participants (upper, middle, and lower rows) on easy, medium, and hard modes of training (left, middle, and right columns)

Note: All teeth are on the same orientation: buccal surface is on the left, lingual surface is on the right, mesial surface is on the top, and distal surface in on the bottom. The error scores are shown at the bottom right corner of each box.

could not be distinguished. These observations were confirmed by comparison of two subsets of trials: on trial series 1 through 4, there was a statistically significant difference between groups ($F_2=11.4$; $p<0.05$). Contrast was observed on all paired comparisons (novices vs. intermediates $p=0.03$; intermediates vs. experts $p=0.02$; novices vs. experts $p<0.01$). On analysis of trial series 5 through 10, performance of all groups was comparable ($F_2=0.8$; $p=0.61$). No

specific contrast was established (novices vs. intermediates $p=0.75$; intermediates vs. experts $p=0.62$; novices vs. experts $p=0.61$). By trial 10, all participants had improved significantly: mean novice error score 2.9 (95 percent CI 2.2-3.5); mean intermediate error score 2.7 (95 percent CI 2.3-3.2); and mean expert error score 1.6 (95 percent CI 1.1-1.9). No significant difference was found between groups at the end of training ($F_2=0.3$; $p=0.83$).

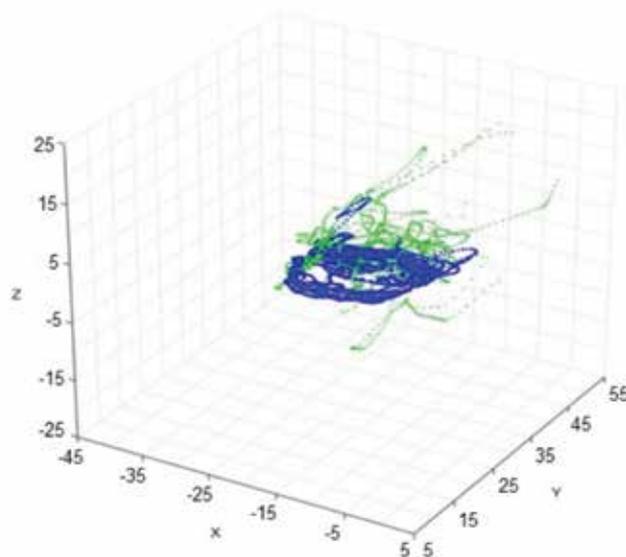


Figure 3. Example of tool path indicating position of tip of instrument during tooth preparation of a novice

Note: Blue line defines a cutting path. Green line defines a non-cutting path.

Hard Mode: Mandibular Molar

Error score analysis was not able to discriminate the difference among three groups at the hard level of training (Figure 4, lower left). However, the path length established a distinction between groups' training (Figure 4, lower right). At trial 1, the mean path length of experts (1.25 cm) was significantly shorter than that of intermediates (2.79 cm; $p=0.033$) and novices (3.16 cm; $p=0.025$). From trials 1 through 5, total path length showed a difference in performance according to groups ($F_2=6.2$; $p=0.01$). Specific contrasts showed that the observed difference was between experts and both other groups (experts vs. intermediates $p=0.04$; experts vs. novices $p=0.01$), whereas intermediates and novices achieved similar performance ($p=0.84$). From trials 6 on, the three groups achieved similar levels of performance ($F_2=1.7$; $p=0.13$). No specific contrast was noted on paired comparisons (novices vs. intermediates $p=0.25$; intermediates vs. experts $p=0.79$; novices vs. experts $p=0.52$).

Error score and path length obtained by experts were retrieved to establish expert benchmarking

criteria for clinical skill assessment. The results of this study have defined a model of training on this particular simulator, which can then become one of the components of a competency-based training curriculum for opening access practice. The resulting mean total error scores and total path length observed were as follows: for the easy level (mandibular premolar), error score=1.56, path length=0.44 cm; for the medium level (maxillary premolar), error score=1.79, path length=0.75 cm; and for the hard level (mandibular molar), error score=1.86, path length=1.02 cm.

Discussion

Haptic VR dental simulators have been introduced into dental curricula as training devices for clinical skill acquisition in several tasks. The important step in evaluating new training tools is to assess validity. The type of validity referred to the purpose of the concept of interest. A haptic VR dental simulator should pass multiple aspects of validity to become a reliable skills trainer and to predict the

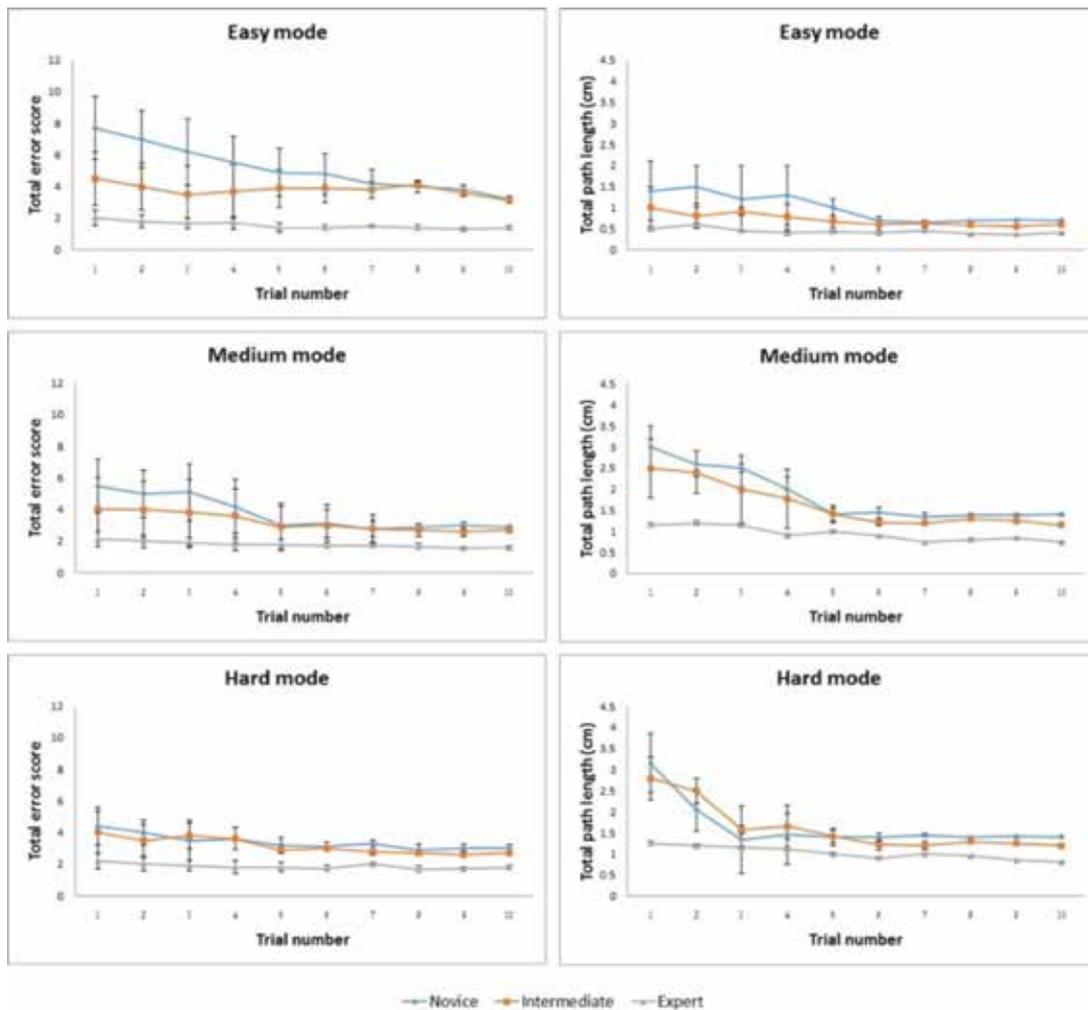


Figure 4. Learning curve showing performance progression from onset (trial 1) to end (trial 10) of training by groups in easy, medium, and hard modes

performance of the trainee.²⁰ This study focused on construct validity of the haptic VR dental simulator. We designed the study to use the task of access opening in endodontics. In access opening task, the operator should have knowledge of the outcome (e.g., identification of all canal orifices, complete roof removal) and knowledge of performance (e.g., movement of instrument). Our study thus focused on the total error score as an outcome measurement. The total path length was used as a process measurement.

In term of construct validity, a valid system should be able to differentiate between skill levels.²¹ Therefore, the metric or parameter assessment must

be related to the levels of personal experience. All participants in this study were asked to perform ten trials for each level of difficulty. The number of trials defined by this study was similar to a previous study on the haptic VR dental simulator for skill acquisition among dental students.¹⁹ Our results showed that the parameters combined in the exercises can be related to various levels of endodontic opening access experience. The results clearly demonstrated that the total error scores of easy and medium levels of difficulty were able to distinguish the three levels of experience. The learning curves for the novice and intermediate groups improved to the desired level

of performance as performed by the expert group within the defined number of trials, and the generated skill was consistent regardless of the initial level of experience. Error score analysis was not able to distinguish any group at the hard level of training, while the total path length parameter was able to discriminate levels of experience for a hard task. The experts therefore appeared to be more consistent in their performance than novices. The decrease in the power of discrimination as the trainer's difficulty level increased may be explained by the fact that each level provides adequate preparedness for the next. Our explanation of this observation is that the skill set taught in the easy level task is useful in completion of the medium level task, and the skill set taught on the medium level task in turn is useful in completion of the hard level task. This is verified by observing the differences among the total error scores at different levels. These results are in line with the construct validity studies conducted with other VR simulators.^{21,22}

Establishing benchmarking criteria by expert dentists is an important step in measuring the ability of the haptic VR dental simulator to train and improve the clinical skill of the novice. Benchmarking criteria for opening access must use valid tools that enable novices to practice on a series of standardized technical tasks. The major advantage of using goal-oriented training, such as performance standards, is the consistency of the final result since all novices are expected to reach the performance standard. For students with outstanding ability, minimal practice may be required. For those who require more practice, appropriate training may be scheduled until the predetermined level of performance is accomplished. In addition, some students are competitive and thrive from having a target to achieve.²³ The expert benchmarking criteria in our study were developed through a scientific validation process and incorporated a competency-based approach. Tasks were proven to be construct-valid, mainly on the basis of quantitative parameter. Learning curve analysis showed that the novices and intermediates improved their performance with repeated practice on the simulator and reached the experts' level of performance at the end of training. We plan to test this model on training using a haptic VR dental simulator in the clinical environment.

The strength of using a haptic VR dental simulator for clinical skill assessment is the ability to automatically record associated kinematic data on how experts or novices perform each procedure

step of the clinical task (e.g., position, angulations, force used), which is not available in conventional skill training environments. The ability of those variables to clearly distinguish between novice and expert skill performance is important for the development of objective scoring criteria. Such variables are needed to build algorithms for the new generation of improved clinical skill training systems that may allow for more effective training experience with real-time feedback on skill performance. The haptic VR dental simulator could be an invaluable tool for training in developing nations where there may be a dearth of teaching dental practitioners and may point the way in the development of VR training tools for other kinds of medical procedures.

Conclusion

This study established construct validity benchmarking criteria for the haptic VR dental simulator by demonstrating its discriminant capabilities between experts and non-experts. Expert error scores and path length were used to define benchmarking criteria for optimal performance. Learning curve analysis demonstrated that novices and intermediates improved their performance with repeat practice on the simulator.

Acknowledgments

This study was supported by the Thailand Office of Higher Education Commission, National Research Council of Thailand, the Thailand Research Fund, and the Center of Excellence in Biomedical Engineering of Thammasat University, Thailand.

REFERENCES

1. Horst JA, Clark MD, Lee AH. Observation, assisting, apprenticeship: cycles of visual and kinesthetic learning in dental education. *J Dent Educ* 2009;73(8):919-33.
2. Subramanian J, Anderson VR, Morgaine KC, Thomson WM. Effective and ineffective supervision in postgraduate dental education: a qualitative study. *Eur J Dent Educ* 2013;17(1):e142-50.
3. Spears R, Leite LP, Schnell RA, et al. AEGD programs: why now, why more? *J Dent Educ* 2013;77(1):17-23.
4. Rhiemora P, Haddawy P, Khanal P, et al. A virtual reality simulator for teaching and evaluating dental procedures. *Methods Inf Med* 2010;49(4):396-405.
5. Maran NJ, Glavin RJ. Low- to high-fidelity simulation: a continuum of medical education? *Med Educ* 2003;37 (Suppl 1):22-8.
6. Spicer MA, Apuzzo ML. Virtual reality surgery: neurosurgery and the contemporary landscape. *Neurosurgery* 2003;52(3):489-97.

7. Panait L, Akkary E, Bell RL, et al. The role of haptic feedback in laparoscopic simulation training. *J Surg Res* 2009;156(2):312-6.
8. De Boer IR, Wesselink PR, Vervoorn JM. The creation of virtual teeth with and without tooth pathology for a virtual learning environment in dental education. *Eur J Dent Educ* 2013;17(4):191-7.
9. Suebnukarn S, Rhiemora P, Haddawy P. The use of cone-beam computed tomography and virtual reality simulation for presurgical practice in endodontic microsurgery. *Int Endod J* 2012;45(7):627-32.
10. Yoshida Y, Yamaguchi S, Kawamoto Y, et al. Development of a multi-layered virtual tooth model for the haptic dental training system. *Dent Mater J* 2011;30(1):1-6.
11. Sohura T, Hojo H, Nakajima M, et al. Prototype of simulation of orthognathic surgery using a virtual reality haptic device. *Int J Oral Maxillofac Surg* 2004;33(8):740-50.
12. Suebnukarn S, Hataidechadusadee R, Suwannasri N, et al. Access cavity preparation training using haptic virtual reality and microcomputed tomography tooth models. *Int Endod J* 2011;44(11):983-9.
13. Thomas G, Johnson L, Dow S, Stanford C. The design and testing of a force feedback dental simulator. *Comput Methods Programs Biomed* 2001;64(1):53-64.
14. Steinberg AD, Bashook PG, Drummond J, et al. Assessment of faculty perception of content validity of PerioSim, a haptic-3D virtual reality dental training simulator. *J Dent Educ* 2007;71(12):1574-82.
15. Yamaguchi S, Yoshida Y, Noborio H, et al. The usefulness of a haptic virtual reality simulator with repetitive training to teach caries removal and periodontal pocket probing skills. *Dent Mater J* 2013;32(5):847-52.
16. Eaton KA, Reynolds PA, Grayden SK, Wilson NH. A vision of dental education in the third millennium. *Br Dent J* 2008;205(5):261-71.
17. Carter FJ, Schijven MP, Aggarwal R, et al. Consensus guidelines for validation of virtual reality surgical simulators. *Surg Endosc* 2005;19(12):1523-32.
18. Aggarwal R, Moorthy K, Darzi A. Laparoscopic skills training and assessment. *Br J Surg* 2004;91(12):1549-58.
19. Suebnukarn S, Haddawy P, Rhiemora P, et al. Augmented kinematic feedback from haptic virtual reality for dental skill acquisition. *J Dent Educ* 2010;74(12):1357-66.
20. Gallagher AG, Ritter EM, Satava RM. Fundamental principles of validation and reliability: rigorous science for the assessment of surgical education and training. *Surg Endosc* 2003;17(10):1525-9.
21. Van Dongen KW, Tournioij E, van der Zee DC, et al. Construct validity of the LapSim: can the LapSim virtual reality simulator distinguish between novices and experts? *Surg Endosc* 2007;21(8):1413-7.
22. Schreuder HW, van Dongen KW, Roeleveld SJ, et al. Face and construct validity of virtual reality simulation of laparoscopic gynecologic surgery. *Am J Obstet Gynecol* 2009;200(5):540.e1-8.
23. Korndorffer JR Jr, Scott DJ, Sierra R, et al. Developing and testing competency levels for laparoscopic skills training. *Arch Surg* 2005;140(1):80-4.