CrossMark

DEVELOPMENT ARTICLE

Advancing virtual patient simulations through design research and inter*PLAY*: part II—integration and field test

Atsusi Hirumi[1,8] (iD) **· Teresa Johnson[2] · Ramsamooj Javier Reyes[1] ·
Benjamin Lok[3] · Kyle Johnsen[4] · Diego J. Rivera-Gutierrez[3] ·
Kenneth Bogert[4] · Stacey Kubovec[5] · Michael Eakins[6] ·
Andrea Kleinsmith[7] · Michael Bellew[5] · Juan Cendan[5]**

Abstract In Part I of this two-part series, we examined the design and development of NERVE: A virtual patient simulation created to give medical students standardized experiences in interviewing, examining, and diagnosing virtual patients with cranial nerve disorders. We illustrated key design features and discussed how design-based research studies improved the total learning experience, including the virtual patient (VP) simulations and the instructional features incorporated with the simulations. In Part II, we examine the efficacy of NERVE and the strategy used to integrate the system into the

---

✉ Atsusi Hirumi
Atsusi.Hirumi@ucf.edu

Teresa Johnson
teresa.johnson@jhmi.edu

Ramsamooj Javier Reyes
javierreyes@knights.ucf.edu

Benjamin Lok
lok@cise.ufl.edu

Kyle Johnsen
kjohnsen@uga.edu

Diego J. Rivera-Gutierrez
djrg@cise.ufl.edu

Kenneth Bogert
kbogert@uga.edu

Stacey Kubovec
s.a.kubovec@Knights.ucf.edu

Michael Eakins
meakins@ist.ucf.edu

Andrea Kleinsmith
andreak@umbc.edu

🖉 Springer

medical school curriculum by field-testing it with 119 s-year medical students, and measuring students' use, reactions, learning, and transfer. We report findings and reflect on lessons learned from the field-test to posit recommendations for improvement and guide the future research and development of virtual patient simulations.

Keywords Virtual patient simulations · Instructional design · Design-based research · Integration · Implementation · Field-test

A variety of simulations may be used to facilitate medical education, ranging from: (a) low-tech simulators (e.g., models or mannequins used to practice simple physical procedures); (b) simulated, standardized patients (i.e., actors trained to role-play as patients); (c) screen-based computer simulators; (d) complex-task trainers (e.g., computer-driven physical models of body parts); and (e) realistic patient simulators (e.g., computer-driven, full-length mannequins that simulate anatomy, physiology, clinical reasoning and decision making) (Maran and Glavin 2003).

Research supports the efficacy of virtual patient simulations (VPs)—a form of screen-based computer simulators that are particularly well suited for fostering the development of clinical reasoning skills (Botezatu et al. 2010; Cook and Triola 2009). Synthesis research concludes that VPs yield consistently higher learning outcomes than conventional educational methods. For instance, based on a meta-analysis of 14 studies published between 1990 and 2010 (including 6 randomized trials, 3 cohort studies, 1 case-controlled study, and 4 pre-post baseline studies), McGaghie et al. (2011) calculated an overall effect size of 0.71 (95 % confidence interval [CI], 0.65–0.76; $P < 0.001$) and concluded that simulation-based medical education, including VPs with deliberate practice, is superior to clinical education that was characterized by the traditional "see one, do one, teach one" instructional approach.

Similarly, a meta-analysis of twelve randomized controlled studies by Consorti et al. (2012) showed a clear positive pooled overall effect for VPs compared to other educational methods, such as lectures, handouts, textbooks, and standardized patients. The overall pooled effect expressed as Odds Ratio (OR) was 2.39 in favor of VPs for all 12 studies

Michael Bellew
Michael.Bellew@ucf.edu

Juan Cendan
juan.cendan@ucf.edu

[1]   College of Education & Human Performance, University of Central Florida, Orlando, USA

[2]   School of Medicine, Johns Hopkins University, Baltimore, USA

[3]   College of Engineering, University of Florida, Gainesville, USA

[4]   College of Engineering, University of Georgia, Athens, USA

[5]   College of Medicine, University of Central Florida, Orlando, USA

[6]   Institute for Simulation and Training, University of Central Florida, Orlando, USA

[7]   College of Engineering and Information Technology, UMBC, Baltimore, USA

[8]   Instructional Design & Technology Program, Department of Education and Human Sciences, University of Central Florida, Orlando, FL 32816, USA

assessing 25 outcomes (95 % CI 1.48–3.84; $p < 0.001$), while the effect size (ES) in favor of VPs as an additional resource was 2.55 in 5 studies assessing 15 outcomes (95 % CI 1.36–4.79; $p = 0.003$) and 2.19 in favor of VPs as an alternative method to more traditional instructional methods for 7 studies assessing 10 outcomes (95 % CI 1.06–4.52; $p = 0.034$). The ES expressed as OR represents the odds of a positive effect for students exposed to VPs and for students not exposed to VPs. An ES > 1 is said to indicate an effect of VPs on the achievement of the learning outcome higher than the control group. However, a meta-analysis of quantitative VPs studies showed that the positive effects were not always large.

In a systematic review and meta-analysis of four qualitative studies, 18 no-intervention control studies, and 21 non-computer comparative studies, Cook et al. (2010) found that VPs are associated with large positive effects compared with no intervention. The pooled effect size (95 % CI; number of studies) was 0.94 (0.69 to 1.19; $N = 11$) for knowledge outcomes, 0.80 (0.52 to 1.08; $N = 5$) for clinical reasoning, and 0.90 (0.61 to 1.19; $N = 9$) for other skills. But in comparison to non-computer instruction, the effects were relatively small: -0.17 (-0.57 to 0.24; $N = 8$) for satisfaction, 0.06 (-0.14 to 0.25; $N = 5$) for knowledge, -0.004 (-0.30 to 0.29; $N = 10$) for reasoning, and 0.10 (-0.21 to 0.42; $N = 11$) for other skills.

Yet, in a subsequent cumulative meta-analysis of 51 studies published between 1978 and 2011, Cook (2014) found a positive effect size of 0.37 (95 % CI 0.23–0.51; $p < 0.001$) for technology-enhanced simulation in comparison with non-simulation training, and concluded that, "some replication is necessary to obtain stable estimates of effect and to explore different contexts, but the number of studies of SBE [simulation-based education] often exceeds the minimum number of replications required" (p. 750), and "At some point, further replication is no longer needed…" and "we should: (i) put into practice what we know, and (ii) perform different studies to investigate different questions (or more nuanced aspects of the phenomenon)" (p. 758).

The purported benefits for using VPs are well documented. VPs can give students extensive opportunities to practice clinical skills and receive feedback in safe environments, and enable educators to give students standardized experiences, particularly with cases that are rare or otherwise difficult to replicate, as well as present variations and promote mastery within and across institutions (Cendan and Lok 2012; Cook and Triola 2009). Nevertheless, it cannot be assumed that VPs will always facilitate student learning. Students demonstrate low usage and report low satisfaction when VPs are not well integrated with, or are offered as an add-on to their curriculum (Haag et al. 2007; Fischer et al. 2007).

In Part I of this two-part article, we described the design and development of NERVE – a web-based, VPs made accessible online to give medical students a standardized experience in interviewing, examining, and diagnosing patients with cranial nerve disorders. We summarized four years of research and development (R&D), and detailed methods used to systematically improve the Alpha, and create the Beta prototype of the system (Hirumi et al. 2016). In Part II, we examine the integration of NERVE into medical school curriculum. Based on existing research and literature, and interactions with the instructor, we formulated a strategy to integrate NERVE and field-tested the Beta version with 119 second-year medical students. We then measured students' use, reactions, learning, and transfer to examine the efficacy of the system, and the strategy used to integrate the system into the curriculum. Here, we report the results of the field-test, describe lessons learned, and discuss how the results are being used to generate the next iteration of NERVE and improve the integration strategy.

VPs integration strategies

Based on a review of 109 peer reviewed journal articles published between 1969 and 2003, Issenberg et al. (2005) concluded that curriculum integration was one of the top three weighted characteristics of high-fidelity medical simulations that lead to effective learning. VPs that are simply "add-ons" result in poor integration and suboptimal learning outcomes (Haag et al. 2007). Explicit and deliberate strategies for integrating VPs are essential for students' acceptance and learning (*ibid*).

Studies examining the integration of VPs highlight the importance of addressing key contextual factors that affect student learning in medical school. Factors such as when and how the VPs are introduced into the curriculum, the perceived relevance of the content and cases, and the nature and scope of both related and competing curriculum resources all affect the use of VPs and the impact they have on student learning. For example, Berman et al. (2009) validated a 15-question survey for measuring the students' perceptions of the effectiveness of VP integration strategies and found that, "elimination of other teaching methodologies was directly associated with perceived effectiveness of the integration strategies" and "students' positive perceptions of integration directly affected their satisfaction and their perceptions of improved knowledge and skills" (p. 942). Based on survey results, Berman et al. (2009) posited an effective strategy to ease VP integration to include: (a) providing effective orientation; (b) integrating Computer Assisted Instruction (CAI) into existing didactics; (c) fostering faculty development to build from students' learning from the cases at the bedside; and (d) eliminating redundant reading.

Huwendiek and de Leng (2010) developed two instruments for facilitating the integration of VPs, including (a) a checklist enabling reviewers to characterize the curricular integration of VPs, and (b) a questionnaire assessing students' experiences with VPs curricular integration in relation to the development of clinical reasoning skills. Based on input received from a sample of 116 medical students completing their pediatric rotation, Huwendiek et al. (2013) found the preferred sequence of VPs and related educational activities was: (a) lecture, (b) interactions with 1–2 VPs, (c) tutor-led small group discussion, and (d) practice with real patients. Edelbring et al. (2012) also found that "more intense follow-up seminars AARs [After Action Reviews] pay off in terms of the benefit perceived by students" (p. 417).

When integrating new technology, such as VPs, into educational contexts, we know that their use and surrounding practices need to be considered (Edelbring, 2010). However, the majority of VPs research examine the relative effectiveness of VPs versus other forms of instruction. In most cases, the integration of VPs in medical education has been performed pragmatically, with limited theoretical or empirical foundations (Edelbring et al. 2011).

Our goals were to improve NERVE and the strategy used to integrate it into the curriculum. To achieve our goals, we field-tested NERVE with 119 s-year medical students and sought to answer four basic questions: How did students use the system? What were learners' reactions to the system? What did students learn from their interactions with the system? And, to what degree were students able to transfer the skills and knowledge derived from the system?

Method

Study context and participants

One hundred seventeen of 119 students (98.3 %) enrolled in a 2nd year neurology course consented to the use of their data for analysis and reporting. We did not include scores from one consenting student in the data analyses because she contributed extensively to the design, development, and testing of NERVE. Study participants included 58 females and 58 males, median age of 24 years (IQR 23–25; range 22–38), representing one or more diverse ethnic groups—34 Asian (29.1 %), 5 Black or African-American (4.3 %), 7 Hispanic or Latino (6.0 %), and 79 White or Caucasian (67.5 %).

NERVE: fundamental components and pedagogical foundations

NERVE is made up of three fundamental components (i.e., a series of introductory frames, a Learning Center, and an Exam Room). The introductory frames consist of an initial splash screen, a list of objectives, and an overview of the system that includes short video tutorials on the contents and use of the Learning Center and Exam Room.

The NERVE Learning Center consists of an interactive area where students can practice using physical examination tools associated with each of the twelve cranial nerves (CNs); information about CN anatomy and physiology, symptoms and pathology; and case studies of each CN. Twelve, ten-item quizzes in the Learning Center also enable students to test and monitor their knowledge of each CN. Figure 1 depicts the tools available in the interactive area for examining CN #2 – the optic nerve.

The Exam Room consists of six virtual patient cases that are accessible using two interfaces: A Selection NERVE and a Chat NERVE interface. Figure 2 compares both interfaces.

The Selection NERVE interface is designed for novice medical students, allowing users to select questions for interviewing the patient from a drop down menu. In comparison, the Chat NERVE interface is designed for more experienced students. It enables users to formulate, organize, and input their own questions.

The Inter*PLAY* instructional theory served as the pedagogical foundation for NERVE (Hirumi et al. 2015) and guided the design of NERVE's three basic components. As an instructional theory, Inter*PLAY* prescribes methods for facilitating experiential learning (Stapleton and Hirumi 2014, 2011). It is based on two central principles posited by Dewey in the 30′s; continuity (the idea that students learn from their experiences), and interaction (the notion that experiences are derived from their interactions with the environment and other individuals) (Dewey 1938). Inter*PLAY* is further grounded in the belief that children and adults learn best when presented with authentic challenges, and when skill development and the learning of facts, concepts, and principles occur in context of how they are used (Barrows, 1985; Schank et al. 1999).

The merits of experiential learning are evidenced by the number of related strategies that have been published over the past 40 years, including those posited by Pfeiffer and Jones (1975), Kolb (1984), Schank et al. (1999) and Clark (2004). With the plethora of interpretations, Lindsey and Berger (2009) distilled three universal principles of experiential learning, including:

Principle 1 - Framing the Experience. Communicate the instructional objectives, assessment criteria, expected behaviors, and social structure (with peers, instructors and
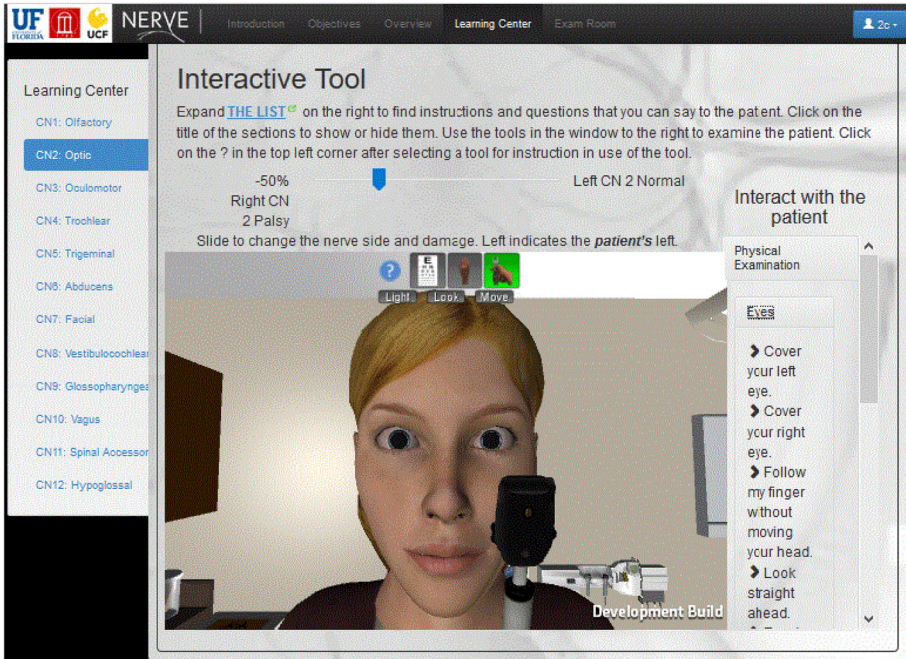
Fig. 1 Screen shot of interactive tools provided for CN2 within NERVE Learning Center
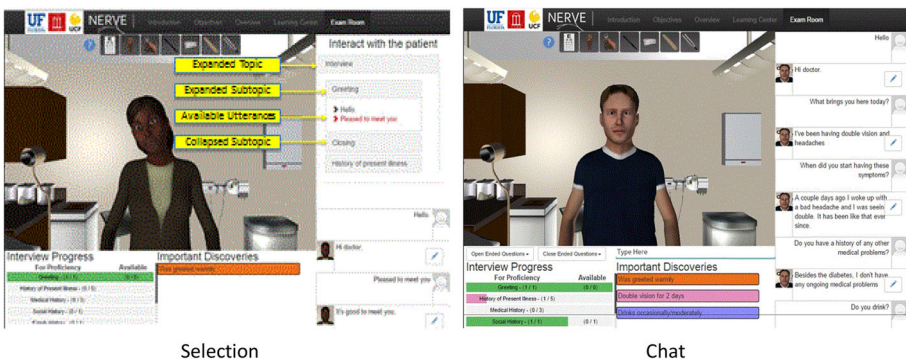

Fig. 2 Screen shots comparing the Selection and Chat NERVE user interfaces

the environment). Variable methods, such as didactic instruction, may frame the experience by providing foundational knowledge required to interpret the experience. Principle 2 - Activating the Experience. Initiate new and prior experiences. Multiple methods may activate experience ranging from laboratory practice to simulations. Key characteristics include (a) providing an authentic experience to facilitate transfer, (b) making decisions with authentic outcomes, (c) orienting students to see the relevance of prescribed learning activities, and (d) presenting optimal difficulty to challenge students with reasonable expectations for success.

Principle 3 - Reflecting on Experience. Experience must be analyzed to learn from it. Reflection should involve students answering the questions, "What happened?" "Why did it happen?" "What did I learn?" and "How would I apply this knowledge to future experiences?" Specific methods for stimulating reflection include teacher facilitation and community building.

The Introduction in NERVE defined the instructional objectives and the prescribed social structure as posited by Principle 1 – Framing the Experience. Other aspects of Principle 1 (i.e., communicating assessment criteria, and expected behaviors) were addressed by the instructor during the initial demonstration of the system. The NERVE Learning Center and Exam Room contained the contents, activities, and simulations that activated the learners' experience as specified by Principle 2. To address Principle 3, the instructor facilitated a two hour AAR that asked students to reflect on their experience after interacting with the system. Inter*PLAY* seeks to advance experiential learning by integrating story, play and game with the universal experiential learning principles to enhance student engagement and learning (Fig. 3).

Application of story, play, and game facilitates six instructional events that form the Inter*PLAY* instructional strategy. Figure 4 illustrates how learners may navigate through the strategy, and the relationship between the instructional events and concepts of story, play, and game posited by the theory.

One reason we adopted Inter*PLAY* is the separation of game and play clearly distinguished the role of the VPs in the Exam Room (that enable students to practice and refine their diagnostic skills) and CN content information provided in the Learning Center (that learners may want or need to inform the diagnostic process). During the first four years of R&D, prerequisite knowledge of CN anatomy, physiology, and pathology was presented by medical school faculty in conventional lecture style format before students were given access to NERVE. Inter*PLAY* illustrated how the addition of content information to the system could enable NERVE to become an independent learning platform that medical schools and students could use to cover both the acquisition and application of relevant CN skills and knowledge.

Fig. 3 Diagram depicting key conceptual elements of the InterPLAY instructional theory

Fig. 4 Diagram illustrating how learners may navigate through key instructional events and components of the InterPLAY instructional strategy
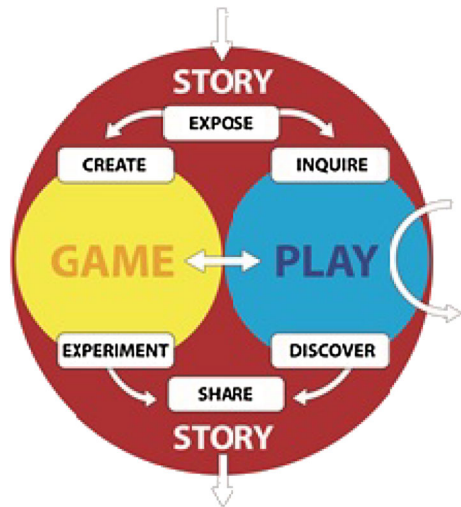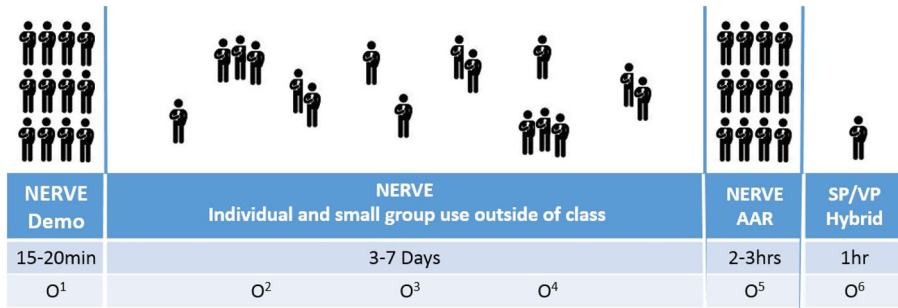
With Inter*PLAY*, we added the Learning Center to NERVE that gave students the opportunity to: (a) learn how and when to use relevant physical examination tools, (b) review relevant information about CN anatomy, physiology, symptoms and pathology, (c) explore published case studies about CN disorders, and (d) take multiple-choice quizzes to monitor their own knowledge acquisition. A link allowed students to access the Learning Center before, during and/or after interacting with the VP cases in the Exam Room.

The addition of the Learning Center also made NERVE complete from a student's perspective. After making a mistake, a student may find it frustrating to hunt for a reliable source of information to correct the misunderstanding. This is especially true in medicine where misinformation is easily available and reliable texts on specific topics can be too in-depth or difficult to understand. The Learning Center serves as an easily accessible resource that can be used as a quick reference to access information and reinforce or correct errors in knowledge.

Further details regarding the design and development of NERVE was presented in Part I of this two-part article (Hirumi, et al. 2016). An in depth study of how the Inter*PLAY* instructional theory informed the design of NERVE, and how the design of NERVE advanced theory was also presented by Hirumi et al. (2015). Here, we highlighted key components of the system and summarized the pedagogical foundations that guided the design of NERVE to give a sense of what students experienced both in class and online, and to help frame the NERVE integration strategy, and interpret field-test results.

NERVE integration strategy

Patterned initially after Huwendiek et al. (2013) preferred sequencing of VPs and educational activities, we used existing VPs integration research, input received from the instructor, and the universal principles of experiential learning to hone our strategy for integrating NERVE into the medical school curriculum. The strategy included (a) a lecture on neurology, (b) a demonstration of NERVE with explicit expectations and requirements, (c) VPs interactions within NERVE, (d) an instructor-led AAR with the entire class, and (e) a standardized patient/virtual patient (SP/VP) hybrid encounter, as depicted in Fig. 5.

NERVE Demo	NERVE Individual and small group use outside of class			NERVE AAR	SP/VP Hybrid
15-20min	3-7 Days			2-3hrs	1hr
O^1	O^2	O^3	O^4	O^5	O^6

Metrics
[1] Demographic
[2] Level 0: Use (computer log)
[3] Level 2: Knowledge Acquisition (quiz)
[4] Level 2: Skill Application (patient assessments)
[5] Level 1: Reactions/Reflections (VP Design Evaluation)
[6] Level 3: Transfer (patient assessments/SP checklists)

Fig. 5 Diagram depicting duration and measurements taken during main components NERVE integration strategy and field-test

Before the term began, the principal investigator (PI) from the collaborating medical school contacted the instructor of the second year neurology class to integrate NERVE and participate in the study. The instructor had heard about the development of NERVE through informal discussion at the school and agreed to discuss the opportunity. During subsequent meetings, the PI and the co-investigator (Co-I) and lead instructional designer demonstrated the latest version of the system and discussed system integration, data collection, and future R&D. The instructor agreed to participate and over two follow-up meetings, worked with the Co-I to refine the strategy and prepare materials to integrate NERVE.

The beta prototype of NERVE was then field-tested with 119 medical students enrolled in a second year neurology course as illustrated in Fig. 5. To activate the experience, the instructor initially demonstrated the use of NERVE during the last 20 min of his regularly scheduled neurology class using pre-prepared MS PowerPoint^{TM} slides, and two video tutorials contained in the system. The short videos demonstrated key components of NERVE, including the contents of the Learning Center and simulated interactions with patients in the Exam Room. The PPT slides communicated requirements and expectations. Students were then briefed about the design study, completed consent forms, and encouraged to use the system over the next week either individually, in pairs, or teams of three-to-four.

During the week, students interacted with NERVE on their own time. Several students commented that it was a particularly busy week with 20–22 h of neuroscience content to digest and NERVE added to the load. On Day 5, the instructor sent a message to all students reminding them of the AAR that was to be completed during the next class session. Students' access and use of NERVE were recorded by the system throughout the week.

For the AAR, the instructor asked students to reflect on their experience with NERVE, and answer three basic questions (i.e., What did they learn? What did they like and why? What should be improved and how?). Students' responses were recorded using Qualtrics. As the students answered the first question, the instructor reviewed their comments, and led a discussion on CN anatomy, physiology, and pathology. At the end, the instructor

Fig. 6 Picture illustrating how medical students interacted with both a standardized patient and virtual patient in clinical setting

reviewed expectations for the SP/VP hybrid encounter that students were to complete the following day.

Due to limited space, half of the students completed the SP/VP interaction one day after the AAR and the other half completed SP/VP encounter two days after the AAR. For the interaction, groups of twelve students were brought into simulated exam rooms (one student/room) and asked to interview a SP while using a nearby computer to interact with a similar VP to complete relevant physical examinations as illustrated in Fig. 6.

Each SP/VP encounter took approximately 20 min. The SP was trained to respond to questions as if they had a particular CN disorder. Students were presented with the same NERVE tools and user interface to interact with the VP that was also programmed to exhibit signs that were consistent with the SP's specified disorder. At the end of the SP/VP encounter, students were asked to submit a differential diagnosis as well as a recommended plan of action.

The field-test represented the final design study completed during the last year of the five-year R&D grant. After the data were compiled, team members were asked to review the field-test results and reflect on their experience throughout the entire last year to help answer the research questions, and formulate lessons learned and recommended improvements.

Instruments

Different instruments were used to measure students' use, reactions, learning and transfer. An AAR was also completed to gain further insights on students' reaction and learning, and recommendations for improving the design and integration of NERVE.

Use

To track students' use, a preliminary login was added to NERVE during the last year of R&D that required students to enter a unique user name, along with the user names of classmates if they were accessing the system in groups. For every session, NERVE recorded the time spent on each accessed page by keeping a persistent connection between the NERVE server and the client side application, and saving the information to a database in the server. Additionally, NERVE recorded students' responses to quizzes, their interactions with the VPs, and their diagnoses of virtual patients in a similar manner.

Reactions

Keller's (1987) Attention-Relevance-Confidence-Satisfaction (ARCS) model of motivational design was applied to measure students' reactions to NERVE using 24-items that were adapted from the Instructional Materials Motivation Survey (IMMS) (Keller 2010). Permission to adapt the IMMS for the field-test was granted by the author. All items were measured on a 5-point Likert-type scale, where 1 = Strongly Disagree and 5 = Strongly Agree. Each of the four ARCS sub-scales contained six items, such that each of the sub-scale scores may range from 6 to 30, and total scale scores may range from 24 to 120. Internal consistency, as measured by Cronbach's coefficient alpha, has been strong for the total scale (0.96) and for each of the sub-scales (0.81–0.92) (Keller, 2010). The survey was distributed to students through Qualtrics following their use of NERVE for the one-week period.

Learning

Three measures were used to assess student learning: (a) scores on multiple-choice quizzes in the NERVE Learning Center, (b) diagnoses of VP cases in the NERVE Exam Room, and (c) students' explanation of what they learned described later under AAR.

The purpose of the multiple-choice quizzes was to test the student's recall of key verbal information, concepts and rules covered in the NERVE Learning Center. Separate ten question multiple-choice quizzes were created by members of the R&D team for each one of the twelve CNs covered in the Center. The questions were later vetted by one of the R&D team members.

Students were given unlimited attempts to complete each quiz. Each attempt was recorded and scored separately. After completing a quiz, the students were provided with one-two sentence feedback on which questions they answered correctly and incorrectly. In case of wrong answers, the system did not provide students with the correct answer, simply an explanation for why their answer was incorrect.

The purposes of the VP cases were to simulate experiences that students may encounter in their medical practice, and test their ability to apply the skills and knowledge covered in the Learning Center. The R&D team worked with physicians throughout the project to generate the six cases that were made available for the field-test. NERVE kept track of the interactions each student had with the VPs, including the questions they asked and examinations they performed on each VP. After students completed their interview and examination, the system asked to complete a Patient Encounter Note which included their diagnosis of the patient and localization of the patient's condition.

Transfer

Two measures were used as indicators of students' ability to transfer clinical reasoning skills and knowledge from NERVE one to two days after the AAR: (a) diagnosis of a SP/VP hybrid case, and (b) a performance checklist assessing students' interviewing skills.

Following each SP/VP hybrid encounter, students exited the exam room to complete a post-encounter note through Qualtrics. Students were asked to describe the localization of the problem by selecting from among 12 CNs and three options representing side (i.e., left, right, bilateral), and to indicate the primary diagnosis by selecting from among 20 options. Localization of the problem was scored as correct when both the selected CN and affected side were accurate. Open-ended text boxes requested students to also provide a differential diagnosis, evaluation plan, and management plan, which were subsequently evaluated for accuracy and appropriateness by the medical school faculty member on the NERVE design and development team.

The SP assessed students' performance using a 15-item checklist that they completed following each SP/VP hybrid encounter. Items assessed history-taking, interpersonal, and communication skills, and were scored as dichotomies (observed/not observed) to reflect percent of 15 total behaviors observed. The checklist items were developed by medical school faculty over 6 years of use and reflect items that are assessed on national board examinations.

After Action Review

The instructor facilitated an AAR during class one week after students were given access to NERVE to gain further insights on students' reactions and learning. Systematic reviews of research and guidelines for facilitating debriefings support the assumption that AARs optimize learning from simulation-based training (Issenberg et at., 2005; Levett-Jones and Lapkin 2014; McGaghie et al. 2010; Paige et al. 2015). The AAR was based on a "Guide to the After Action Review" published by the Veteran's Affair Office as a resource for improving work (Salem-Schatz et al. 2010). The AAR consisted of three questions: What did you learn from your interactions with NERVE? What did you like about NERVE and why? What should be changed to improve NERVE and how? Approximately 110–115 of the 119 registered for the second year neurology courses attended the session. A few were noted as missing by the instructor and fellow students but attendance was not taken.

Throughout their first 2 years of medical school, the students were divided into teams of 10–15 members to complete class assignments and accommodate logistical issues at the college with no predetermination of their pre-medical degrees, prior experiences, etc. For the AAR, students were first asked to sit together with their teammates to facilitate discussion and input. The questions were projected onto the main screen at the front of the auditorium. Students were asked to bring their laptops to class to enter responses to each question that were captured in Qualtrics. The instructor and the researchers at the session could see students' responses to the questions as they were entered but the responses were not projected onto the main screen, prohibiting students to read what others wrote. The entire session lasted a little over 2 hours with approximately 1 hour being spent to obtain, refine and discuss responses to the first question (what did you learn?), and approximately 30 min to gather students' input on what they liked and would they would recommend changing.

Statistical analysis

Categorical variables are presented as frequency and percentage, ordinal and non-normal continuous variables are reported as median and interquartile (IQR) range and/or minimum–maximum (range), and normal continuous variables are displayed as mean ± standard deviation (SD) with 95 % confidence interval (CI) of the mean. Instrument reliability was assessed as internal consistency using Cronbach's coefficient alpha. Bivariate correlations were calculated using Pearson's correlation coefficient (r). Statistical analyses were conducted using SPSS 22.0 (IBM; Armonk, NY).

Results

Field-test results are organized according to students' use of NERVE and three out of the four levels of training evaluation posited by Kirkpatrick (1994), including Level I Student Reactions, Level II Student Learning, and Level III Student Behavior (Transfer).

Students' use of NERVE

Table 1 depicts students' use of NERVE, including the number of times logged into the system, the amount of time spent interacting with the system, the number of students who interacted with system individually or in teams of two-four, the number of quizzes and VP cases completed, and the number of CNs explored in the Learning Center.

On average, students logged into NERVE six times and spent 0–8.5 h interacting with the system with a median of 1.9 h. The majority of students interacted with the system individually (53 %) or in pairs (24 %). The majority of students (101/116, 87 %) also completed all five quizzes as prescribed by the instructor, and a few opted to complete

Table 1 Summary of students' use of NERVE ($n = 116$)

# Login Sessions	Median[2] (IQR)	6.0 (3.5–9.0)
Expectation: none specified	Range	0–23
Time	Median[2] (IQR)	1.9 (1.4–2.9)
Expectation: none specified	Range	0–8.5 h
Primary access mode	Individual	61 (53 %)
Expectation: individually or in teams of 2–4	Pair	28 (24 %)
	Team of 3–4	14 (12 %)
	Did not use	3 (1 %)
Quizzes completed	0	6 (5 %)
Expectation: complete minimum of 5 quizzes	1–4	9 (8 %)
	5	96 (83 %)
	>5	5 (4 %)
VP cases completed	0	16 (14 %)
Expectation: complete minimum of 3 cases	1–2	21 (18 %)
	3	71 (61 %)
	>3	8 (7 %)
12 CN content exploration	No CN	29 (25 %)
Expectation: whatever necessary to prepare	1–6 CN	54 (47 %)
for SP/VP and AAR	7–11 CN	33 (28 %)
	All 12 CN	0 (0 %)

additional quizzes (05/116 students, 4 %). Seventy-nine students (68 %) completed three or more VP cases as directed. Fifty-four (47 %) students examined the interactive tools and contents about 1–6 CNs and 33 (28 %) explored tools and content about 7–11 CNs in the Learning Center.

Figure 7 shows the average amount of time students spent interacting with NERVE during the prescribed week of use. It clearly illustrates that students' use of the system peaked the day before the scheduled AAR and SP/VP hybrid patient encounter.

Level I student reactions

Students' reactions to NERVE were studied using Keller's IMMS and two questions asked during the AAR.

Perceived attention, relevance, confidence and satisfaction

One hundred ten of 116 students (94.0 %) completed the IMMS survey (Table 2). Internal consistency was strong for the ARCS total scale (0.89, $k = 24$), and moderately strong for each of the four 6-item sub-scales—attention (0.68), relevance (0.62), confidence (0.61), and satisfaction (0.80). Mean score and SD for the ARCS total scale was 75.6 ± 11.2 (95 % CI 73.5–77.8) and for each of the sub-scales, as follows: attention, 17.9 ± 3.5 (95 % CI 17.3–18.6); relevance, 20.2 ± 3.1 (95 % CI 19.6–20.8); confidence, 18.9 ± 2.4 (95 % CI 18.5–19.4); and satisfaction, 18.6 ± 4.1 (95 % CI 17.8–19.3). Student responses and summary data by individual item are provided in Table 2.

After action review (student reactions)

Two questions asked during the AAR examined students' reactions to NERVE: What did you like and why? What would you improve and how? Twenty-four responses were received from individual and groups during the 30-minute period given to students to respond to the first question. Most responses include 3–5 different statements. Tables 3 presents a sample of student responses to the question on what they liked.
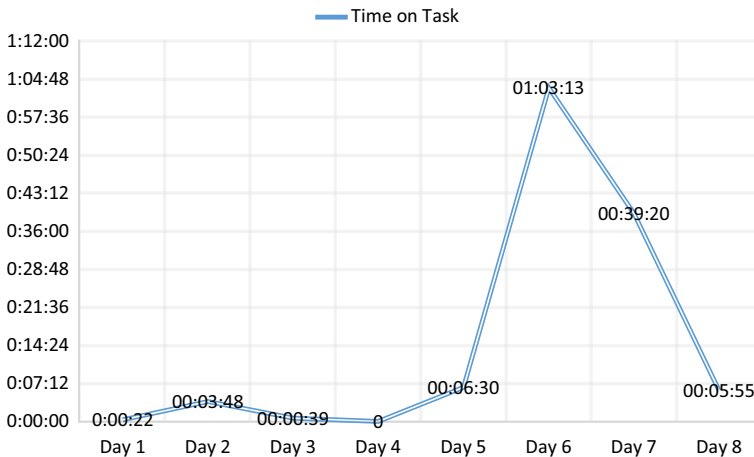

Fig. 7 Average amount of time students spent interacting with NERVE during the week of prescribed use

Table 2 Perceived levels of Attention, Relevance, Confidence and Satisfaction as a measure of Students' reactions to NERVE

Scale	IC	Mean	Avg. (scale 1–5)
Overall ARCS	0.89	75.6 ± 11.2	
Attention	0.68	17.9 ± 03.5	3.6
Relevance	0.62	20.2 ± 03.1	4.1
Confidence	0.61	19.9 ± 02.4	3.7
Satisfaction	0.80	18.6 ± 04.1	3.1

Content analysis of the responses revealed that statements centered around five general themes, including the objectives, visualization, accessibility, the NERVE Learning Center and the NERVE Exam Room. Comments about the Learning Center were also further directed toward the Learning Tools; content information provided about CN anatomy, physiology, pathology, and symptoms; and the quizzes. Approximately 110 statements were distilled from the twenty-four responses received regarding what students liked about NERVE. Table 3 consists of a small sample of the student comments to illustrate key points and response patterns.

As illustrated by the sample responses, positive comments were received for each of the seven learning objectives specified by NERVE. Apparently, students felt that some aspect of NERVE facilitated achievement of each of the objectives. The most recurrent positive comments were made about the interactive tools, content information, and quizzes provided in the NERVE Learning Center, supporting the teams' decision to add the Learning Center to the system during the final year of R&D, and students' recommendation to include quizzes received during one-to-one and small group formative evaluations. However, no one commented about the additional case studies that were posted about each CN.

The second most frequent positive comments were made about the VP cases and interactions in the NERVE Exam Room. Students liked different aspects of interviewing, examining and diagnosing patients as well as progress monitoring, and the selection and chat interfaces. Apparently, students also thought the visuals elements included in NERVE were particularly useful, such as labeled diagrams, simulated manifestations of cranial nerve palsies, and interactive animations of eye movements. Not all comments, however, were positive.

Twenty-four responses were received from individuals and groups during the 30-min period given to students to respond to the question, what would improve about NERVE and how? Again, three responded that they did not interact with NERVE, and most responses included 3–5 different statements. Approximately 145 specific statements expressing concern and/or recommendations were distilled from the twenty-four responses regarding improvements to the system that centered around seven basic issues: integration, technical, content, quiz, diagnostic, interviewing and examining, and interface issues. Table 4 presents a sample of student responses to the question, illustrating key points and response patterns.

Most recommendations for improving the integration of the system revolved around timing; when NERVE should be used within the students' program of study. The primary issues with timing were the development of pre-requisite skills. NERVE was integrated in the 2nd week of a 6-week neuro module. Prior to the use of NERVE, students noted that they had yet to receive much instruction on diagnosing patients, which also explains the relatively low scores on performance assessments reported later in Table 6. Additional

Table 3 Sample of student reactions when asked what they liked about NERVE and why

Theme	Student comments
Objectives	Objective 1—recognize the anatomy and physiology of cranial nerves. "The anatomy was very complete during the reading sections for the cranial nerves." Objective 2—recognize the pathology and symptoms of cranial nerve dysfunctions. "We think that the simulator did a good job of demonstrating the symptoms of cranial nerve dysfunction." Objective 3—distinguish cranial nerve dysfunctions. "The eye ball diagram with the muscles and nerves did a good job of demonstrating the different cranial nerve palsies." Objective 4—perform appropriate physical exams. "Many of us have never seen a patient with these lesions, so it was worthwhile to perform a physical exam on a patient with classic presentation of these lesions." Objective 5—select or formulate organized interview questions. "The interview questions were extremely thorough." Objective 6—identify damaged cranial nerves. "This program did a good job of demonstrating where damaged cranial nerves were." Objective 7—formulate hypotheses about the cause of identified cranial nerve palsies. "The program helped us identify what a certain cranial nerve palsy will manifest as"
Learning cente: interactive tools	"The CN examination tools and maneuvers presented helped us narrow in on what physical exam components were relevant to the presenting complaint. For example in the patient with blurry vision attributed to CN 3 and increased ICP the expected maneuvers (pupillary reflex, visual acuity, visual fields, sensation in the areas of the facial nerve) helped us understand additional pathology that may be associated with this complaint." "The [interactive] tools did a good job of demonstrating how the different palsies manifest. It was very helpful to be able to switch the palsy between the left and right side and see how their symptoms would change. The charts were helpful at the beginning of each section for learning where the nerves originate and where the go. It helped us relearn how patients will have a contra-lateral head tilt for cranial nerve 4."
Learning Center: CN anatomy, physiology, pathology and symptoms	"We enjoyed reading through the 12 CN in Learning Center and having key information presented in a high yield, concise and consistent manner for all CN nerves." "The learning center was also very helpful because all of the information regarding anatomy, physiology, symptoms, and pathology was in one place." "Anatomy descriptions were very specific and correlated well with the pictures included right next to the information." "Physiology - Explained very well. Helped understand normal functioning of extra-ocular muscles, and the cranial nerves involved in their function." "The symptoms and pathology component provided information that allowed us to consider mechanisms of injury that were not directly associated with the head."

Table 3 continued

Theme	Student comments
Learning center: Quizzes	"I learned the most form the quizzes"
	"The quizzes provided a good overview and helped identify gaps in knowledge that could then be discovered within the symptoms and pathology section or the anatomy and physiology section"
	"We liked how the quiz allowed you to push the button to go back to the reading to explain the answers to the quiz questions"
	"The quizzes were engaging and challenging The feedback provides was useful as well"
Exam room: interviewing patients	"The CN examination allowed us to learn what the appropriate questions are to ask patients presenting with suspected CN palsies"
	"In addition, the virtual patient encounter had multiple reminders about empathizing with the patient, which was useful to reiterate its importance in patient communication"
	"Additionally, we thought that it was useful that the system then pointed out when significant findings were made because it helped us understand what information was relevant to the case. (for example in the patient with the trauma it was relevant to discover when the injury had occurred, how long the double vision had been going on, associated symptoms, things that improved/worsened the double vision, medications, social history involving smoking and alcohol use among others)"
Exam room: examining patients	"The interactive portion of the examination was very helpful. It was realistic and allowed a unique learning experience"
	"Furthermore, the fact that we were able to perform the tests multiple times was helpful because it allowed us to review the pathology that was present (whereas this would probably not be possible with a real patient)"
	"The convergence tool was also useful, and the ability that it had to make this a three dimensional experience. It was also beneficial to use the ophthalmoscope to see a close up view of the optic nerve, and to make sure the patient did not have papilledema"
	"The fact that you could act out the hand motions of the physical exams was helpful. It was helpful to change the amount of fingers and touch the patient, very realistic"
	"All discoveries and 'important discoveries' sections were really helpful- as we did the exams it was nice that the program kept track of our findings and made them concise"
Exam room: user interfaces	"As a group, we felt that [the menu driven] VP cases were very helpful especially in solidifying the order in which to ask questions and complete testing procedures"
	"During the guided physical exam, having all of the questions there was helpful in how to word questions when typing in questions in the other mode"
	"The interview portion went really well, especially having the selection option in addition to the chat option. The selection option already has the questions which makes it easier for people doing a neuro interview exam for the first time"
	"NERVE's chat mode allowed us to build on this and apply critical thinking to our virtual patient interviews"

Table 3 continued

Theme	Student comments
Visualization	"The physical manifestations of cranial nerve palsies were well represented"
	"The interactive patient was very helpful to demonstrate pathologies involving extra-ocular muscles. Seeing the dilation and constriction of pupils showed up nicely on the virtual patient. Like it was mentioned, some of these signs can be hard to visualize on a standardized patient"
	"The labeled diagrams in the symptoms and pathology section helped us learn how the different palsies will manifest"
	"Charts of where the CNs came into and out of the skull"
	"The tables and the succinct cranial nerve summaries describing the functions and lesions of the cranial nerves, along with the images given, were the most useful"

instruction on how to complete a primary and differential diagnosis is needed either prior to using or incorporated into NERVE.

Several students also recommended clarifying time expectations. Expectations for what had to be completed were delineated during the introductory lecture but several students thought further clarity on how much time each student was expected to spend interacting with the system would help optimize their use of the system. Another unexpected finding related to integration was students' preference for learning resources. Several students commented that they would like NERVE as a supplemental, rather than a mandatory learning tool, citing preferences for other resources that they have grown accustomed to over the past 2 years.

The majority of student recommended improvements to the system centered on fixing technical issues. Evidently, students had problems logging into the system and with the patient encounters found in both the Learning Center and Exam Room. Grammatical and spelling errors were also found in the quizzes. Technical issues were identified during expert reviews, one-to-one and small group evaluations, and repeated tests by all members of the R&D team. We also implemented a code freeze, conducted a load test, and focused on debugging the system a week prior to the field-test, but evidently, such efforts were not sufficient. Thankfully, the majority was still able to complete the assigned tasks but technical issues may be one reason why most students only met minimum requirements.

Level II student learning

Four sources of information were used as indicators of student learning, including (a) students' scores on the five required quizzes; (b) students' performance on three required VP cases embedded in NERVE, (c) students' explanation of what they learned provided during the AAR; and (d) the instructor's assessment of students' explanations.

Quiz scores

Table 5 reports students' scores on the five required quizzes specified by the instructor in the NERVE Learning Center, including the number of students who took each quiz, and the mean, range, and mode for each quiz.

Table 4 Sample of student recommendations for improving NERVE

Theme	Student comments
Integration	"We haven't been taught much on diagnoses so it was hard to give a differential diagnoses"
	"We also felt that the time period to learn about both the NERVE system and the actual cranial nerves themselves was not adequate; we suggest about a week to familiarize ourselves with the NERVE system and about a week to master the cranial nerves since there was no accompanying lecture"
	"There was also SO much information last week and with studying for discretionary IRATs constantly, I felt that I barely had enough time to synthesize the information and felt burdened by having a deadline for using the simulation"
	"Implement this system during first year when students are learning CN for the first time because it is a really helpful resource and it is nice to have all of the CN information in one place"
	"Clearer expectations about how much time we should be spending with the program (30 min or 5 h?)"
Technical Issues	"Our entire team had some technical difficulties, whether during the log-in process or during the patient interviews themselves and felt that our learning was somewhat compromised by this"
	"All of our group members encountered some technical difficulties, whether during the log-in process or in the virtual patient encounter (and the associated software download), and we felt that this could be greatly improved"
	"When the tools worked, they were very intuitive (but sometimes the tools didn't work ex. tongue depressor)"
	"Sometimes credit is not given for questions that seem to be relevant to case. Sometimes the empathy comments did not seem appropriate to the situation"
	"Sometimes it seemed like the completion status was given arbitrarily. For example, I did the "H" extraocular eye movement test the same way on three patients but I only received a completion notice for the third patient"
Learning center: CN anatomy, physiology, pathology and symptoms	"Many of us have alternative and varied resources that we are used to using for the past 2 years, so we did not like being forced to use a new resource that we felt we didn't learn as effectively from. We believe that this should be a supplementary tool, not a mandatory part of the curriculum because we learn different ways"
	"Anatomy, Physiology—we did not learn anything new. Even if there were things from last year we didn't remember, we used our other outside resources or notes from last year rather than reading it from NERVE
	"Symptoms and Pathology—there was new information, but I did not learn it effectively from this resource. This was an introduction to some of the diseases we could encounter with CN deficits, but since we hadn't learned them in lecture or elsewhere, it wasn't super helpful. Some of us would prefer to use other resources we are already accustomed to learn this information"

Table 4 continued

Theme	Student comments
Learning center: Quizzes	"Grammar and spelling should be reviewed. Quiz questions were sometimes unclear and sometimes had multiple correct answers. The order of true and false in standard examinations is always true first and false second"
	"Some of the questions did not provide clear feedback and it was sometimes difficult to discern what the right answer was"
	"Not all the questions on the quizzes had explanations that explained the correct answer. It was not possible to see which quizzes had been completed. Showing the correct answer along with the incorrect selection would help increase understanding"
Exam room: diagnosing patients	"There could be further additions regarding the differential diagnosis for cranial nerve lesions along with the assessment and plan. There should be more information regarding the plan for cranial nerve pathologies"
	"The diagnosis and the causes are not well explained. There should be a section that emphasizes how to construct a differential diagnosis. There should be a section that emphasizes DIFFERENCES between different palsies (CN 4 will not have symptoms X, Y, and Z that are present in CN 3 palsies etc.)"
	"We learned differential diagnoses for different CN lesions; however, felt that the causes of the lesion could be covered in better detail during the wrap up following the cases"
	"It was sometimes difficult to understand the entire presentation of the patient from each individual exam finding that we may or may not have gotten…if we could have a short video at the end for summary, that would be great"
Exam room: interviewing and examining patients	"We also thought that maybe the history could have been listed out in a summary rather than having us click on every single questions and have a response (a time saver that we believe would not interfere with the learning process)"
	"Streamline the actions of the physical exam to prevent needing to do certain parts of the physical exam. For example, click "Perform EOM" and watch the motion occur, as opposed to what currently is needed which is clicking the "Hand" then clicking the "EOM" button then telling the patient to "Follow my finger" then moving the hand in the EOM motion using the mouse"
Exam room: user interfaces	"In the case simulations we thought that in selection mode the question options could be presented in a better order. The sequence was sometimes illogical (i.e. the Chief Complaint information was presented below the History of Present Illness, the empathy options were presented at the very end instead of below the greeting and History of Present Illness)"
	"The selection mode in the examination room sometimes did not account for knowledge that had already been obtained from the patient. For instance if the patient had been asked about past medical history that was counted as a significant finding but if that information included that the patient had diabetes in order for that system to count the finding of diabetes that information had to be asked as a specific question
	"In the chat format, it was frustrating that the patient couldn't respond to most of the things we tried to ask (maybe this was because we were all using the server?)"

Table 5 Students' scores on 5 required quizzes in NERVE Learning Center

Scale	n	Mean	Range	Mode
CN 3	101	82.4	40–100	80
CN 4	102	77.6	30–100	90
CN 5	97	85.8	40–100	90
CN 7	97	76.4	20–100	80
CN 10	99	74.9	10–100	100

In comparison to quizzes taken in class, mean scores were somewhat low and ranges were high. Typically, mean scores on in-class quizzes fall between 80 and 90 %, and students rarely score below 60–70 %. Further examination of the NERVE quiz scores indicate that a relatively small number of students scored lower than usual. For example, from the 102 students who took the CN 4 quiz, 11 (10.7 %) scored between 30 and 50, and 14 (13.7 %) students scored a 60. Some students are thought to score lower on NERVE quizzes for two possible reasons: (a) the quizzes did not count toward the students' course grade; and/or (b) the quizzes were taken prior to reviewing related content to assess prior knowledge. In other words, students may have used the quizzes as a diagnostic tool to determine which aspects of each CN (e.g., anatomy, physiology, symptoms and/or pathology) to concentrate on when they did review the content information to optimize their time interacting with the system. During the AAR, students often noted they liked the quiz features and the ability to test their own learning. Student use data also revealed that a few repeated quizzes to earn a higher score; an option typically not available in class.

Virtual patient (VP) assessments

To diagnose the VPs in the Exam Room, students had to identify the CN and the side of the nerve that was damaged. Table 6 depicts the results of students' diagnoses, including the name of the VP, the CN that was damaged, the number of students who completed each case, and the number and percentage of correctly diagnosed cases in terms of the CN and side of injury.

Student performance on the VP assessments varied by task. When identifying the damaged nerve, students often correctly identified the injury for VP Cathy (88 %). When presented with the additional challenge of identifying the laterality of the damage, performance was best for VP Molly (87 %) as well as VP Cathy (83 %). In this complex interactive evaluation model, student accuracy for identification of the injured nerve or laterality were more often in the 55–79 % range.

Table 6 Students' scores for diagnosing virtual patients in NERVE Exam Room

VP	CN	N	Correct	Side	Correct
Bill	6	57	34 (60 %)	Left	37 (65 %)
Jennifer	4	47	35 (74 %)	Left	32 (69 %)
David	None	35	19 (55 %)	None	17 (49 %)
Molly	3	30	25 (67 %)	Left	26 (87 %)
Monica	3	24	18 (75 %)	Left	19 (79 %)
Cathy	6	41	36 (88 %)	Left	34 (83 %)

What appears to be a disparity between students' assessment of David (who presented with Myasthenia Gravis which mimics but is not classified as a CN palsy), and students' AAR statements about what they learned is worth noting. Of the six VPs in the Exam Room, David was correctly diagnosed the lowest percentage of times (55 %). In contrast, students often cited Myasthenia Gravis as a key learning point during the AAR. While scores on the VP assessment measures were relatively low, it appears that a number of students learned from misdiagnosing the case and receiving feedback from the system. Reflecting on this issue, the medical student on the project commented, "thinking about why you were wrong tends to result in better retention of the information, possibly because the explanations are examined more closely to identify gaps in knowledge to avoid repeating the same mistake."

Such findings support the results of systematic reviews of VPs research that indicate that providing feedback may be one of the most important features of SBE (Issenberg et al. 2005). Assessment scores by themselves may not accurately reflect student learning, but deliberate practice coupled with informative feedback may increase clinical reasoning skills in controlled settings (McGaghie et al. 2011). Evidently, formative and summative feedback received from erroneous interactions with the VPs may reveal misconceptions about CN pathology and push students to engage in ways that are not necessarily measured by embedded performance tests.

After action review (what did students learn?)

Thirty-one responses were received from individuals and groups during the 60-min period that was taken to obtain and discuss the responses to the question, what did you learn from your interactions with NERVE? Again, three responded that they did not interact with NERVE and did not contribute any further comments.

Initially, students' responses strayed to what they liked and didn't like about NERVE. Apparently, students wanted to first note their frustration with as well as express what they felt were useful aspects of the system. After 5–10 min, the researchers stopped the session, noted how students were inputting responses that were more appropriate for other AAR questions, projected a few statements that focused on what they learned, and then asked students to continue providing input. Ensuing responses focused on what they learned. After 30 more minutes, the instructor stopped the session and began discussing the input, first noting particularly insightful student comments, and then elaborating on key topics, verbal information, concepts, and rules that he felt were valuable but were either missed or addressed in an insufficient manner. Table 7 provides a sample of students' responses.

Analysis of students' responses indicate that students spent most time examining content information related to the five required quizzes; all but two sets of comments focused on the anatomy, physiology, symptoms and pathology associated with CN 3, 4, 5, 7 and 10. One additional set noted what students learned about examination and interviewing skills in general. The other set stated what students learned about CN 6, indicating that some students went beyond minimum expectations. The instructor also noted during the AAR, a number of students associated elevated pressure with CN6 palsy. He thought it was great that students made this association because such pathology is not often/necessarily learned in class. In other words, interactions with NERVE resulted in acquisition of key verbal information and concepts that may otherwise not be acquired.

Table 7 Sample student responses to when asked what learned about CNs from their interactions with NERVE

Theme	Student comments
CN 3: Oculomotor	The difference between myasthenia gravis and CN 3 palsy, is the symmetric nature of MG vs. the asymmetric nature of CN3 palsy
	Also, CN3 does not affect the superior tarsal muscle, so the ptosis will be more severe with CN3 palsy
	CN3 palsy: "Down and out" + dilated pupil + ptosis. Loss of direct and consensual reflex in the eye on the affected side
	The adductors have the most abnormal function in CN 3 lesions
CN 4: Trochlear	Trochlear is the only CN that exits the posterior brainstem, and it is also the longest
	The most common cause of CN 4 palsy in children in congenital. CN 4 originates from the posterior midbrain and has the longest intercranial tract. It innervates the contralateral eye, which is unique
	I learned that otitis media was the most common cause of trochlear nerve palsy, and that that Trochlear nerves are associated with ipsilateral pain above the eyebrow
	We also learned that trochlear nerve palsy affects the position of the optic disc from the macula.—Head tilting to the contralateral for trochlear palsy— Affected eye on oculomotor palsy cannot fully follow the 'H' movement with the finger
CN 5: Trigeminal	I learned things such as the Trigeminal nerve supplies the anterior 2/3 of GENERAL sensation.
	Different trigeminal branches exit out of different foramen. V1– superior orbital fissure, V2- Foramen Rotundam, V3—Foramen Ovale
	Learned that jaw claudication is associated with trigeminal nerve lesion and with giant cell arteritis. Learned that sudden intense pain is likely due to an aneurysm. The pain associated with ruptured aneurysms can diminish with time
	Trigeminal palsy—loss of sensation and chewing difficulty. Sickle Cell is involved in trigeminal nerve pathology
CN 6: Abducens	We learned that when a patient has a CN6 lesion, binocular, vertical diplopia worsens when the gaze is directed downward (such as when going down stairs)
	I learned the exits of the CNs I learned how the left vagus and the right vagus travel across the esophagus
	I learned that Myasthenia gravis can mimic any cranial nerve deficiency
	As a team, we also learned an increased ICP can cause CN 6 damage—but the reasoning behind this pathology was not within the literature
CN 7: Facial	The chordi tympani is from the facial nerve and does taste for anterior tongue
	The cranial nerve 7 had five branches was a reminder but I learned a great mnemonic for remembering the names of the branches: To Zanzibar By Motor Car
	Facial nerve supplies the anterior 2/3 of TASTE sensation
	Recurrent laryngeal nerve loops under the aortic arch
CN 10: Vagus	Many of us had forgotten that the nodose ganglion was the home for the visceral afferents of CN10
	I learned the exits of the CNs I learned how the left vagus and the right vagus travel across the esophagus
	Vagus—symptoms would be when saying "Ahh" the uvula moves toward the intact side
	Most cases of vagus nerve damage is caused by surgery. For children, always suspect congenital disorders

Table 7 continued

Theme	Student comments
Examination and interviewing skills	How to perform an appropriate exam relevant to the various cranial nerve disorders
	CN Examination Tools: Most specific ways to assess CN palsies, correlating exact PE skills (following finger with H pattern) to specific palsies (CN 3, 4, 6)
	We also learned the pertinent questions to ask a patient
	I learned how to respond to patient's questions such as "Will I be okay?"

Instructor's assessment of students' explanations

To further evaluate student learning, the instructor was asked to assess the depth, breadth, and accuracy of students' explanations of what they learned from NERVE reported during the AAR. Overall, the instructor felt that the vast majority of students' comments were accurate, and the depth of learning was good but the breadth of what they learned could be improved per unit time invested in interacting with the system.

Based on what students reported, the instructor felt that NERVE helped students better understand how and why a given cranial nerve injury produces specific clinical symptoms and signs. He noted that the VPs seemed particularly useful for helping students visualize the action of the superior oblique muscle, and better understanding the reason a patient tilts his or her head in a certain way as a response to impaired nerve function. The simulation appeared to develop visual recognition skills for CN lesions which is particularly important since it is generally not reasonable for a standardized patient to simulate a CN deficit. The instructor was also surprised that the VPs appeared to help students better define new as well as commonly used healthcare phrases such as "a pupil-sparing third nerve palsy," and "monocular versus binocular diplopia."

In terms of breadth, the instructor noted that students' explanation of what they learned from NERVE focused on the peripheral nervous system (PNS). Evidently, medical schools typically cover the PNS (matters outside the brain and brainstem) in general anatomy courses during the first year, and the central nervous system (CNS) (matters inside the brain and brainstem) in neuroanatomy courses during the second year. He felt that NERVE could increase the breadth of what students learn as well as increase the utility of the system by adding information content information to the Learning Center, and cases to the Exam Room that focused more on the CNS. He also predicted that higher efficiency would lead to greater usage and recommended optimizing text by further distinguishing information about CN anatomy, physiology, and pathology inside versus outside the brainstem, and eliminating verbiage.

Level III student behavior/transfer

Two indicators of student behavior/transfer were recorded after students interacted with NERVE, including students' ability to diagnose a virtual patient-standardized (VP-SP) hybrid case and students' ability to interview the standardized patient.

Table 8 Frequency and percentage of students providing congruent v. non-congruent diagnoses for the clinical SP case and VP examination hybrid encounter

		Frequency	Percent	Cumulative percent
Not congruent with clinical case and VP examination	Hemorrhagic Stroke	1	0.9	0.9
	Transient ischemic attack	1	0.9	1.8
	Optic neuritis	3	2.6	4.4
	Ruptured aneurysm	3	2.6	7.0
	Other (specify in differential diagnosis)	7	6.0	13.0
Congruent with Clinical Case and VP examination	Compressive palsy	13	11.1	24.1
	Raised intracranial pressure	19	16.2	40.3
	Ischemic stroke	28	23.8	64.1
	Cranial neuropathy (unspecified)	42	35.9	100.0
Total		117	100.0	

Diagnosis of SP/VP case

The majority of students correctly identified the injured CN (CN6; 108/117, 92.3 %) and side (left; 115/117, 98.3 %) for the SP/VP case—108 students (92.3 %) correctly identified both the affected CN and side. Students were also tasked with identifying the possible underlying pathology and providing a differential diagnosis for the case. Table 8 reports that frequency, percent, and cumulative percentage of students who identified pathologies that were or were not congruent with the SP/VP hybrid examination.

Data from the combined SP/VP encounter reveal that students correctly identified a possible underlying pathology in 87 % of patients. The injury could have been any of the 4 more often given answers (i.e., compressive, raised pressure, ischemic stroke, or neuropathy). The four least commonly given answers, and "others" were, indeed, unlikely in the given clinical presentation. CN injury was likely NOT an intracranial hemorrhage, hemorrhagic stroke or neuritis. Apparently, distinguishing a CN 6 palsy from other CN palsies on physical exam was not that difficult, but determining why the nerve was damaged and being able to tell whether the eye movement issue was due to the nerve damage or to something else, such as an infection in the muscle, was more challenging, hence the variation in primary diagnosis.

Standardized patient checklists

Standardized patient (SP) checklist scores were available for all 116 consenting students (Table 9). Percent correct scores on the 15-item checklist ranged from 66.7 % (10 items) to 100.0 % (median = 93.3 % or 14 items; IQR 93.3–100.0). Student performance by individual item is presented in Table 9.

The checklist scores indicate a high percentage of students properly interviewed and examined the SP. However, as a measure of transfer, the results must be interpreted with caution. The SP checklist items were related to overall interpersonal and communication skills, and general interviewing and history-taking techniques. There were no CN-specific

Table 9 Student performance on SP checklist ($N = 116$)[a]

Item	Completed n (%)	Not completed n (%)
Asked if the patient has any other questions or concerns prior to leaving room	108 (93.1)	8 (6.9)
Communicated clearly, avoided medical jargon or explained terms when used	115 (99.1)	1 (0.9)
Demonstrated genuineness, care, concern, empathy	112 (96.6)	4 (3.4)
Encouraged patient to develop full and accurate understanding of key messages	86 (74.1)	30 (25.9)
Explained purpose of encounter within the first 1–2 min	109 (94.0)	7 (6.0)
Explored how health issues have affected the patient	100 (86.2)	16 (13.8)
Explored the patient's worries/fears about cause(s)/implications	100 (86.2)	16 (13.8)
Expressed interest in the patient as a person	108 (93.1)	8 (6.9)
Greeted patient warmly and verified patient's identity	115 (99.1)	1 (0.9)
Introduced him/herself to the patient (first and last name, full title)	115 (99.1)	1 (0.9)
Listened attentively	115 (99.1)	1 (0.9)
Provided information related to the working diagnosis and/or next steps	109 (94.0)	7 (6.0)
Treated the patient with respect	116 (100.0)	0 (0.0)
Used open-ended techniques that encouraged the patient to tell his/her story	116 (100.0)	0 (0.0)
Washed hands before patient contact and maintained clean technique	111 (95.7)	5 (4.3)

[a] Median percent correct score = 93.3 % or 14/15 items; interquartile range = 93.3–100.0; minimum–maximum = 66.7–100.0

checklist items, and many factors contribute to the development of students' interviewing skills throughout the curriculum. As stated explicitly in its objectives, NERVE is designed to give students practice in "performing appropriate physical exams," and "selecting or formulating organized interview questions." Statements made by students during the AAR, such as "[we learned] how to perform appropriate exams relevant to various cranial nerves," "we learned the pertinent questions to ask a patient," and "we learned how to respond to patient's questions" suggest that NERVE enhanced students' ability to interview and examine patients, but we cannot directly attribute students' scores on the SP checklist to the use of NERVE.

Correlations between levels

Bi-variate correlations were calculated using Pearson's correlation coefficient (r) to determine if there were any relationships between students' use, reactions, learning, and transfer. The analyses revealed that the total number of hours spent using the NERVE system during the 1-week period was significantly correlated with the ARCS confidence sub-scale score ($r = 0.20$; $p = 0.04$), ARCS satisfaction sub-scale score ($r = 0.26$; $p = 0.01$), and ARCS total score ($r = 0.22$; $p = 0.02$). Students who spent more time interacting with NERVE reported greater confidence, satisfaction, and overall motivation with the use of NERVE than students who spent less time interacting with NERVE. Significant correlations were also found between the total number of hours spent using

NERVE during the 1-week period and students' quiz scores achieved on the first attempt for CN 4 ($r = 0.23$; $p = 0.02$) and CN 5 ($r = 0.32$; $p = 0.002$). Correlations found between students' use, reactions, and learning reinforce the concept that time-on-task matters, and accentuate the potential value of exploiting design and integration strategies that promote students' use and engagement. However, it is also important to keep in mind that these correlations, though significant, do not imply causation; they neither indicate the direction of the relationship, nor rule out other factors that may be affecting the outcomes. They suggest a relationship may exist that, in turn, may be worth further study. No other significant correlations were identified between field-test levels of use, reactions, learning and transfer.

Lessons learned & recommended improvements

The instructor and the R&D team members were asked to review the field-test results, and reflect on the data and their overall experience to identify lessons learned, and forward recommendations for improving NERVE and the strategy used to integrate NERVE into the medical school curriculum. Responses were compared to prior research to guide future R&D.

Lessons learned

We learned six fundamental lessons about the design and integration of NERVE from the field-test and the last year of R&D.

Overall, the NERVE integration strategy was effective in motivating students to meet minimum expectations

Patterned after Huwendiek et al.'s (2013) preferred sequencing of VPs and educational activities, the NERVE integration strategy included (a) a lecture on neurology, (b) VPs interactions within NERVE, (c) an instructor-led AAR with the entire class (rather than a tutor-led, small group discussion as recommended by Huwendiek et al.), and (d) a SP/VP hybrid patient encounter. Student use data indicate that the majority of students completed the prescribed tasks (i.e., 5 quizzes, 3 virtual patient cases), providing support for Huwendiek et al.'s (2013) basic strategy, along with Edelbring et al.'s (2012) finding that "more intense follow-up seminars (AARs) pay off in terms of the benefit perceived by students" (p. 417). Significant correlations were also found between student use (time on task) and perceived levels of confidence, satisfaction, and overall motivation.

Integrating a VP simulation into medical school curriculum is a non-trivial task that is influenced by more than the design of the simulation and nature of activities presented before, during and after the simulation

Student use data indicate that a majority of students completed prescribed tasks. However, students spent varying amounts of time interacting with the system (0–8 h), the majority accessed the system the day before the AAR and SP/VP encounter, and few went beyond specified minimum requirements. Reflections by team members and the instructor as well as student responses during the AAR indicate that competing curriculum requirements, positioning within students' overall program of study, the nature of educational activities

completed prior to integration, and the familiarity and preferred use of alternative learning resources also had a significant influence on student use and learning. The timing of when students are exposed to differential diagnoses needs to be optimized; it appears that students were asked to consider diagnoses prior to learning about them "in class." Despite our efforts to set up a section of the platform as a learning center, it appears that some students still see this as a backup to classroom teaching. The instructor also suggested that better distinguishing the content information and quiz items contained in the Learning Center, and the cases in the Exam Room according to 1st and 2nd year medical school curriculum requirements may facilitate integration, and enhance both students' and instructors' use of the system.

A code freeze and focused effort by team members to test the system, and identify and fix technical problems were not sufficient for ensuring bug free experiences

Technical issues can have a significant adverse effect on students' use and perceptions of the system. Throughout the five-year project, students reported technical issues with the system. During one experiment, the entire system crashed and we could not gather valid data. We learned that it is very difficult to eliminate bugs from such a complex system that was constantly evolving through experimentation and formative feedback. Based on such lessons, we ran load tests, implemented a code freeze, and tasked all team members to test the system and report errors to the programmers 1 week prior to the field-test.

Evidently, our strategy for tracking changes, and for testing and debugging the system were not sufficient. Although technical issues did not prevent students from completing prescribed tasks, students did report a number of technical problems logging into the system and interacting with the VPs. These technical problems were mainly caused by the high number of students accessing the system just prior to the AAR and SP/VP encounter. While the system was ready for loads of up to 50 % of the students accessing it at once, as much as 70 % logged in and performed VP interviews during that last 2 days which lead to decreases in performance.

Authenticity of simulated interactions should be based on desired learning outcomes

Reviews of VP research and VP design studies accentuate the importance of authenticity (of the user interface and student tasks as well as in the presentation of content, language, clinical data, clinical context). Researchers and practitioners often conclude that authenticity is essential to effective learning and VP simulation design (e.g., Botezatu et al. 2010; Huwendiek et al. 2009; Issenberg et al. 2005). Our experience suggests that such deductions must be further qualified. Not all VP interactions need to be highly authentic; rather, the authenticity of VP interactions should depend on the desired learning outcome.

One-to-one and small group evaluations during the final year of development indicated that attempts to make the physical exams (e.g., using an ophthalmoscope, a tongue depressor, tuning fork, or hands) as realistic as possible often required students to take an insubordinate amount of time learning how to manipulate the instrument which frustrated them. For example, maneuvering the virtual ophthalmoscope in 3 dimensions with a mouse and keyboard requires a 3D control scheme. It was difficult to judge distance on a 2D monitor. Learning to properly manipulate medical instruments to complete physical exams, however, was not a specified outcome. We rationalized that the manipulation of such tools was best taught in clinical rather than simulated environments. Simplifying the simulated interactions necessary to complete the physical exams to a few mouse clicks decreased the

authenticity of the interaction, but reduced frustration and increased satisfaction with the user interface.

Adding content information on anatomy, physiology, symptoms and pathology, along with related quizzes to the VP simulations enhances integration

Berman et al. (2009) found that the elimination of redundant readings and other teaching methodologies was directly associated with perceived effectiveness of VP integration strategies. By adding content on CN anatomy, physiology, symptoms and pathology to the system during the final year of development, we allowed the instructor to eliminate redundant reading assignments. Adding content also reduced the need to lecture and review basic concepts about CN anatomy, physiology, symptoms and pathology, allowing the instructor to spend valuable in-class time addressing other curriculum requirements. Positive perceptions of the Learning Center reported during the AAR reinforce the decision to add content and quizzes to the system.

It was useful to give students examples of desired responses to AAR questions

Students' preliminary responses to the first AAR question varied, initially straying to feelings about the system. Even though students were asked, "What specifically did you learn about CN anatomy, physiology, symptoms, pathology, examination tools and interview questions?" and were prompted to be specific, several started by commenting on what went wrong and how to improve the system, rather than focusing on what they learned about CNs.

The AAR guidelines we followed did provide sample ground rules for facilitating the AAR, along with prompts to encourage specific input from participants (Salem-Schatz et al. 2010), but they neither suggested providing sample responses, nor did they point out that respondents may want to start by discussing issues with the system and recommending improvements. The AAR guidelines do point out the importance of starting with what transpired before determining what was good or bad about the system. The guidelines also recommend building on best practice by asking what went well and why before addressing problems.

After providing several examples of desired responses, and noting that they would get an opportunity to vent their frustrations and recommend improvements to NERVE later in the AAR, students focused their responses to reflect what they learned from the system. This lesson learned supports findings and recommended best practices for facilitating debriefings that note the importance of managing and otherwise directing students' reactions and emotions in a positive manner during AARs (Ahmed et al. 2012; Paige et al. 2015).

Improving NERVE

Based on insights gained from the field-test and final year of R&D, recommendations for improvement by R&D team members include tactics for facilitating the integration of NERVE and refining the system to enhance student engagement, reactions, learning, and transfer.

Address additional curricular factors

Lessons learned from the integration and field-testing of NERVE identify factors beyond those addressed by the planned integration strategy that affect student use and learning. In addition to the tactics delineated in Fig. 3, we should consider (a) if related CN diagnostic tools and techniques are taught prior to the use of NERVE, (b) additional curriculum requirements imposed on students during the planned use of the system, (c) refining expectations and requirements based on how the VP stimulation relates to students' program of study (e.g., first year vs. second year curriculum), and (d) adding instructor testimonials and use during initial demonstration of the system that show students what they perceive as the value of the system. We should also ensure that the instructor sends reminders throughout the week, and requirements are also posted on the student information system.

Consider adding story and game mechanics

The Inter*PLAY* instructional theory posits the integration of story, play, and game elements with experiential learning principles to enhance learner engagement (Hirumi et al. 2015). Findings from experiments conducted during the initial years of R&D suggest that game mechanics (such as leaderboards and customizable avatars) may also increase students' engagement with the system (Halan et al. 2010). Limited time and resources during the final year of R&D prohibited much development of story and gameplay in the beta version of NERVE that was field-tested in this study. Recommendations for improving NERVE and increasing students' use and engagement include: (a) adding spoken narrative by virtual patients during initial student interactions in the Exam Room to develop further interest and empathy with each case, (b) adding a story to increase engagement and expectations across cases in Exam Room, (c) adding game mechanics, such as leaderboards, scoring counters, customizable student avatars, and badges (for properly diagnosing patients).

Well-crafted narrative provides the opportunity to form an empathetic connection with the student. By establishing rapport and avoiding medical jargon or superficial conversation with the VPs, the characters become more interesting and engaging, and the students' understanding can go beyond information presentation and make the experience seem more natural.

Introducing leaderboards and scoring may be particularly beneficial for competitive personalities. By adding such additional levels of feedback, students may have a better understanding of how they are performing compared to peers, extending beyond the feedback on personal performance. However, in simulated environments, game mechanics have the potential to introduce the risk of students focusing on "gaming" the system and should be carefully designed to ensure key learning objectives are not circumvented in an attempt to "win." An example would be a speed-based score. Introducing a ranking system for speed could have an adverse effect by students prioritizing speed of completion rather than conducting a thorough interview. The addition of game mechanics as well as story should be considered through formative testing and a series of design studies.

Consider adding cumulative feedback across cases

In addition to adding story and game mechanics, we should consider providing cumulative feedback to students on how their diagnostic skills are developing across cases to

encourage repeated use and practice with the system. Currently, the system provides feedback on students' examination and interviewing skills by: (a) indicating the number of discoveries made, along with the total number of discoveries available, and the number of discoveries that an expert deemed as sufficient; and (b) providing a progress report immediately after students completed each case that contained the students' primary diagnosis, a list of discoveries, and a transcript of the students' interaction with the VPs (Hirumi, et al. 2016). Compiling data from the progress reports to provide cumulative feedback on students' progress and learning across cases, along with indicators of overall performance may encourage students repeated use of NERVE over time. Like other recommended enhancements, such additions should be created and considered through formative testing and a series of design studies.

Better distinguish and add control over content, quizzes, and cases

As indicated as one of the key lessons learned after field-test, the instructor thought we could increase students' and instructors' use of NERVE by better distinguishing content information and quiz items presented in the Learning Center, and the virtual patient cases presented in the Exam Room based on first and second year curriculum requirements. By better distinguishing PNS and CNS content information and quiz items in the Learning Center, and virtual patient cases in the Exam Room, and adding control features that enable students and the instructor to select their interactions, we may optimize students' time-on-task and offer what the instructor referred to as "high-yield" educational experiences that are valued by both medical students and instructors.

Offer different levels of challenge in SP/VP encounter and VP cases

In video games, "difficulty levels" are a common feature. Certain elements are varied to make gameplay more or less challenging from one level to the next. For example, the ferocity of an enemy attack may range from easy, medium to hard. Games present different levels to enable those more or less skilled to enjoy gameplay and to encourage "re-playability" as players become more proficient. Our SP/VP hybrid case had a static difficulty level, which may have been too easy given students' high scores. We neither designed the VP cases with varying levels of difficulty, nor did we indicate the difficulty of each case on the case selection page.

There are several ways to vary the difficulty level of the SP/VP encounter and the VP cases in the Exam Room. First, our simulation may allow for adjustments to the patient's physical presentation, such that it would be more difficult to detect an abnormality (e.g. ocular range of motion could be reduced by only 25 % instead of a full 100 %). Second, we may introduce comorbidities (multiple disorders presenting simultaneously) that require several diagnoses and may overlap in presentation. There are also physical findings, such as variation in pupil size (anisocoria) that may not represent an abnormality, but require further physical examination to be sure. Finally, we may reduce the information provided by patients during interviews. This would require greater depth of questioning and require students to remember a larger set of responses from patients. These features could enable instructors to tailor the experience to the perceived skill level of the learners. They could also be made adjustable by students directly, or be "unlocked" through case completion.

Implement alternative tactics for tracking changes and debugging system

Improved code reliability could be realized through applying two concepts throughout software development. First, writing unit tests around the code should decrease errors in the code. Unit tests would help determine if individual units of source code are fit for use, and if the computer program modules work properly together with associated control data. Second, load testing to ensure our systems would support large groups of simultaneous users should be done throughout the development process, not just after the code freeze. These load tests should anticipate 100 % of the students accessing the system at one single point in time; as opposed to assuming that 50 % may access the same features at the same time. We should also consider implementing a project management system, such as Trello, to track team members' progress and changes made to the system, and to give greater visibility of activities across the team.

Conclusion

The field-test represented the practical delivery of what drove us to develop NERVE 5 years ago. The ability to present an unusual and important physical abnormality to a student couched in the context of a standardized patient delivering the history. Before NERVE, there was really no way to bring those two components together – the realistic patient-physician interactions that are possible with standardized patients, and the correct interactive anatomic representations of CN palsies that could not exist without the use of simulations.

Field-test results indicate that the strategy used to integrate NERVE and the current design of the system facilitated achievement of specified learning objectives and completion of minimum requirements for using the system. Nonetheless, the results also highlight the influence medical school curriculum and other contextual factors may have on student achievement and performance. Application of sound instructional design (ID) principles and processes that address the analysis, design, and development of instruction is important but not sufficient. To impact learning and ensure the success of VPs, educators and instructional designers must look beyond the application of instructional theories, and narrowly defined ID practices that center on design, rather than the integration of instructional inventions in real-world school settings. Consistent with prior VPs studies, we found that the alignment of content and VP cases with 1st and 2nd year medical school curriculum, and what educators do to (a) orient students to the VPs, (b) eliminate redundant activities and resources, (c) set and communicate expectations, and (d) integrate VPs use with existing instructional practices may also have a substantive effect on students' use, reactions, learning and transfer.

Our goal is to inspire both instructor and student use beyond the specified minimum expectations. Reflecting on field-test results revealed six primary lessons learned along with six main recommendations for improvement. Insights derived from the field-test and experience gained over the last year of R&D also demonstrate the value of conducting an iterative series of design studies to improve the system, along with the theory, tools and tactics used to create the system. Design-based investigations are again recommended for continued research and development of NERVE and virtual patient simulations.

Compliance with Ethical Standards

References

Ahmed, M., Sevdalis, N., Paige, J., Paragi-Gururaja, R., Nestel, D., & Arora, S. (2012). Association for Surgical Education: Identifying best practice guidelines for debriefing in surgery: a tri-continental study. *The American Journal of Surgery, 203*, 523–529.

Barrows, H. S. (1985). *How to design a problem based curriculum for the preclinical years*. New York: Springer.

Berman, N., Fall, L., Smith, S., Levine, D. A., Maloney, C. G., Potts, M., et al. (2009). Integration strategies for using virtual patients in clinical clerkships. *Academic Medicine, 84*(7), 943–949.

Botezatu, M., Hult, H., Kassaye Tessma, M., & Fors, U. G. (2010). As time goes by: Stakeholder opinions on the implementation and use of a virtual patient simulation system. *Medical Teacher, 32*(11), e509–e516.

Cendan, J., & Lok, B. (2012). The use of virtual patients in medical school curricula. *Advances in Physiology Education, 36*(1), 48–53.

Clark, R. E. (2004). *Design Document for a Guided Experiential Learning Course*. Submitted to satisfy contract DAAD 19-99-D-0046-0004 from TRADOC to the Institute for Creative Technologies and the Rossier School of Education, University of Southern California.

Consorti, F., Mancuso, R., Nocioni, M., & Piccolo, A. (2012). Efficacy of virtual patients in medical education: A meta-analysis of randomized studies. *Computers & Education, 59*(3), 1001–1008.

Cook, D. A. (2014). How much evidence does it take? A cumulative meta-analysis of outcomes of simulation-based education. *Medical Education, 48*(8), 750–760.

Cook, D. A., Erwin, P. J., & Triola, M. M. (2010). Computerized virtual patients in health professions education: A systematic review and meta-analysis. *Academic Medicine, 85*, 1589–1602.

Cook, D. A., & Triola, M. M. (2009). Virtual patients: a critical literature review and proposed next steps. *Medical Education, 43*(4), 303–311.

Dewey, J. (1938). *Logic: The theory of inquiry*. New York: Holt, Rinehart and Winston.

Edelbring, S. (2010). A three-fold framework for relating to innovations and technology in education: learning from, with and about technology. In A. Bromage, L. Clouder, J. Thistlethwaite, & F. Gordon (Eds.), *Interprofessional E-learning and collaborative work: Practices and technologies* (pp. 23–33). Hershey: IGI Global.

Edelbring, S., Broström, O., Henriksson, P., Vassiliou, D., Spaak, J., Dahlgren, L. O., & Zary, N. (2012). Integrating virtual patients into courses: follow-up seminars and perceived benefit. *Medical Education, 46*(4), 417–425.

Edelbring, S., Dastmalchi, M., Hult, H., Lundberg, I. E., & Dahlgren, L. O. (2011). Experiencing virtual patients in clinical learning: a phenomenological study. *Advances in Health Sciences Education, 16*(3), 331–345.

Fischer, M., Hege, I., Hörnlein, A., Puppe, F., Tönshoff, B., & Huwendiek, S. (2007). Virtual patients in medical education: a comparison of various strategies for curricular integration. *Zeitschrift fur Evidenz, Fortbildung und Qualitat im Gesundheitswesen, 102*(10), 648–653.

Haag, M., Singer, R., Bauch, M., Heid, J., Hess, F., & Leven, F. (2007). Challenges and perspectives of computer-assisted instruction in medical education. Lessons learned from seven years of experience with the CAMPUS system. *Methods of Information in Medicine, 46*(1), 67.

Halan, S., Rossen, B., Cendan, J., & Lok, B. (2010). High score! - motivation strategies for user participation in virtual human development. *Intelligent Virtual, 9783642158919*, 482. doi:10.1007/978-3-642-15892-6_52.

Hirumi, A., Johnson, T. & Reyes, R. J. (2015). *Field-testing Strategies to Improve the Integration of Virtual Patient Simulations into Medical School Curriculum*. Concurrent session presented the annual Association for Educational Communication and Technology conference, Indianapolis, IN. November 4–6.

Hirumi, A., Kleinsmith, A., Johnsen, K., Kubovec, S., Eakins, M., Bogert, K., et al. (2016). Advancing virtual patient simulations through design research and interplay: Part I: Design and development. *Educational Technology Research and Development,*. doi:10.1007/s11423-016-9429-6.

Huwendiek, S., & De Leng, B. A. (2010). Virtual patient design and curricular integration evaluation toolkit. *Medical Education, 44*(5), 519.

Huwendiek, S., Duncker, C., Reichert, F., De Leng, B. A., Dolmans, D., van der Vleuten, C. P., & Tönshoff, B. (2013). Learner preferences regarding integrating, sequencing and aligning virtual patients with

other activities in the undergraduate medical curriculum: A focus group study. *Medical Teacher, 35*(11), 920–929.

Huwendiek, S., Reichert, F., Bosse, H. M., De Leng, B. A., van der Vleuten, C. P., Haag, M., & Tönshoff, B. (2009). Design principles for virtual patients: a focus group study among students. *Medical Education, 43*(6), 580–588.

Issenberg, S. B., McGaghie, W. C., Petrusa, E., Gordon, D. L., & Scalese, R. J. (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Medical Teacher, 27*(1), 10–28.

Keller, J. M. (1987). Development and use of the ARCS model of instructional design. *Journal of Instructional Development, 10*(3), 2–10.

Keller, J. M. (2010). *Motivation design for learning and performance: The ARCS model approach.* New York: Springer.

Kirkpatrick, D. L. (1994). *Evaluating training programs: The four levels.* San Francisco: Berrett Koehler Publishers.

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development.* Englewood Cliffs: Prentice-Hall Inc.

Levett-Jones, T., & Lapkin, S. (2014). The effectiveness of debriefing in simulation-based learning for health professionals: A systematic review (Protocol). *JBI Database of Systematic Reviews and Implementation Reports, 9*, S81–S96.

Lindsey, L., & Berger, N. (2009). Experiential approach to instruction. In C. Reigeluth & A. Carr-Chellman (Eds.), *Instructional-design theories and models* (Vol. 3, pp. 117–142)., Building a common knowledge based New York: Routledge.

Maran, N. J., & Glavin, R. J. (2003). Low-to high-fidelity simulation – a continuum of medical education? *Medical Education, 37*(1), 22–28.

McGaghie, W., Issenberg, S., Cohen, E., Barsuk, J., & Wayne, D. (2011). Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence. *Academic Medicine, 86*(6), 706–711.

McGaghie, W. C., Issenberg, S. B., Petrusa, E. R., & Scalese, R. J. (2010). A critical review of simulation-based medical education research: 2003–2009. *Medical Educator, 44*, 50–63.

Paige, J. T., Arora, S., Fernandez, G., & Seymour, N. (2015). Debriefing 101: Training faculty to promote learning in simulation-based training. *The American Journal of Surgery, 209*(1), 126–131.

Pfeiffer, J. W., & Jones, J. E. (1975). Introduction to the structured experiences section. In J. E. Jones & J. W. Pfeiffer (Eds.), *The 1975 annual handbook for group facilitators.* La Jolla: University Associates.

Salem-Schatz, S., Ordin, D., & Mittman, B. (2010). Guide to the after action review v1.1. Accessed 27 September 2014 from http://as.vanderbilt.edu/overview/faculty/facultycouncil/archive/sitemason.vanderbilt.edu/files/cHpJCw/Guide%20to%20the%20After%20Action%20Review.pdf

Schank, R. C., Berman, T. R., & Macpherson, K. A. (1999). Learning by doing. In C. M. Reigeluth (Ed.), *Instructional design theories and models: A new paradigm of instructional theory* (pp. 161–179). Hillsdale: Lawrence Erlbaum Associates.

Stapleton, C., & Hirumi, A. (2011). Interplay instructional strategy: Learning by engaging interactive entertainment conventions. In M. Shaughnessy & S. Fulgham (Eds.), *Pedagogical models: The discipline of online teaching* (pp. 183–211). Hauppauge: Nova Science Publishers Inc.

Stapleton, C., & Hirumi, A. (2014). Designing InterPLAY learning landscapes to evoke emotions, spark the imagination, and promote creative problem solving. In A. Hirumi (Ed.), *Grounded designs for online and hybrid learning: Practical guidelines for educators and instructional designers* (pp. 159–190). Eugene: International Society for Technology in Education.

Atsusi Hirumi, PhD During Year 3, Dr. Hirumi was invited to join the team as the lead Instructional Designer. His research focuses on advancing experiential learning theory and the design of training stimulations. During the last year of the project, Dr. Hirumi's status was raised to Co-I as he provided leadership for design and testing of NERVE.

Teresa Johnson, PhD participated in study design, selection and consent of participants, and data management. She also completed statistical analyses. Her research interests include effective use of game-based learning, and utility of real-time, contextual tracking of learner performance for sampling of formative and summative assessment opportunities.

Ramsamooj Javier Reyes assisted with expert reviews, one-to-one evaluations, small-group evaluations, and field testing. Javier is an active duty Air Force officer pursuing a doctoral degree in Instructional Technology with interests in neuroscience research, virtual patient design and integration, and foreign language education.

Benjamin Lok, PhD Coordinated the acquisition of funding, personnel and other resources, and led the technical team in the development and direction of the project. Dr. Lok has developed and evaluated mixed reality and virtual reality interfaces for 13 years. He currently leads a virtual patient research group whose work has been integrated into eight universities across the country.

Kyle Johnsen, PhD Coordinated resources and personnel in the implementation of the virtual patient physical examination interface and Learning Resource center. Dr. Johnsen primary interest lies in promoting critical thinking and problem solving through learner-oriented virtual environments, and currently leads projects in science, engineering, and medical education.

Diego J. Rivera-Gutierrez Used CSS and HTML5 to develop and improve the Learning Resource Center, developed conversational models that control the virtual patients in the exam room, and created back-end to track student navigation. He has experience with virtual reality and modeling of physical systems and currently performs research in machine learning with multi-robotics.

Kenneth Bogert Designed and developed the cranial nerve model, the virtual patient control systems that interact with the nerve model, and the visualization for illustrating the motion of the extraocular muscles under various cranial nerve disorders. He has experience modeling physical systems and performs research in machine learning with multi-robotics.

Stacey Kubovec made major contributions to the development of content information included in the system as well as to testing and providing input to improve the design of key system features from a medical student's perspective. Stacey is pursuing a career in academic medicine, with interests in both clinical research and medical education.

Michael Eakins Assisted with design and development of 3D environment, characters, tools, and animations. Previous experience includes, game design, 2D/3D design, and 3D modeling and animation for games, simulation, and training. Professional areas of focus are around mobile games for training and immersive interactive storytelling.

Andrea Kleinsmith, PhD Coordinated the development process, worked with subject matter experts, and developed the virtual patients. Her research focus is in the area of affective computing and understanding how to create virtual humans in training systems that synthesize emotion and can recognize emotion from the user.

Michael Bellew, MD Facilitated integration of NERVE, including initial demo and subsequent AAR. Dr. Michael Bellew is a board-certified neurological surgeon with additional sub-specialization in management of complex cerebrovascular disease. At UCF, he serves as the Director for the Neuroscience Module as well as the Director for the Neurology Core Clerkship.

Juan Cendan, MD Coordinated the acquisition of funding and obtained access to the curriculum and medical students for VPs research. He currently Chairs the Medical Education Department at the UCF and has worked in clinical skills education with medical students utilizing the entire range of simulation platforms from standardized patients to manikins, task trainers and virtual patients.