An Empirical Evaluation of an Instrument to Determine the Relationship Between Second-Year Medical Students' Perceptions of NERVE VP Design Effectiveness and Students' Ability to Learn and Transfer Skills from NERVE

Ramsamooj Reyes
University of Central Florida

AN EMPIRICAL EVALUATION OF AN INSTRUMENT TO DETERMINE THE RELATIONSHIP BETWEEN SECOND-YEAR MEDICAL STUDENTS' PERCEPTIONS OF NERVE VP DESIGN EFFECTIVENESS AND STUDENTS' ABILITY TO LEARN AND TRANSFER SKILLS FROM NERVE

by

RAMSAMOOJ JAVIER REYES
B.S. University of Puerto Rico (UPR), 1995
Professional Certificate, UPR Medical School, 1997
M.S. University of Maryland University College, 2008

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the College of Education and Human Performance
at the University of Central Florida
Orlando, Florida

Spring Term
2016

Major Professor: Atsusi Hirumi

ABSTRACT

Meta-analyses and systematic reviews of literature comparing the use of virtual patients (VPs) to traditional educational methods support the efficacy of VPs (Cook, Erwin, & Triola, 2010; Cook & Triola, 2009; McGaghie, Issenberg, Cohen, Barsuk, & Wayne, 2011). However, VP design research has produced a variety of design features (Bateman, Allen, Samani, Kidd, & Davies, 2013; Botezatu, Hult, & Fors, 2010a; Huwendiek & De Leng, 2010), frameworks (Huwendiek et al., 2009b) and principles (Huwendiek et al., 2009a) that are similar in nature, but appear to lack consensus. Consequently, researchers are not sure which VP design principles to apply and few validated guidelines are available. To address this situation, Huwendiek et al. (2014) validated an instrument to evaluate the design of VP simulations that focuses on fostering clinical reasoning. This dissertation examines the predictive validity of one instrument proposed by Huwendiek et al. (2014) that examines VP design features. Empirical research provides evidence for the reliability and validity of the VP design effectiveness measure. However, the relationship between the design features evaluated by the instrument to criterion-referenced measures of student learning and performance remains to be examined.

This study examines the predictive validity of Huwendiek et al.'s (2014) VP design effectiveness measurement instrument by determining if the design factors evaluated by the instrument are correlated to medical students' performance in: (a) quizzes and VP cases embedded in Neurological Examination Rehearsal Virtual Environment (NERVE), and (b) NERVE-assisted virtual patient/standardized patient (VP/SP) differential diagnosis and SP checklists. It was hypothesized that students' perceptions of effectiveness of NERVE VP design

are significantly correlated to the achievement of higher student learning and transfer outcomes in NERVE.

The confirmatory factor analyses revealed the effectiveness of NERVE VP design was significantly correlated to student learning and transfer. Significant correlations were found between key design features evaluated by the instrument and students' performance on quizzes and VP cases embedded in NERVE. In addition, significant correlations were found between the NERVE VP design factors evaluated by Huwendiek et al.'s (2014) instrument and students' performance in SP checklists. Findings provided empirical evidence supporting the reliability and predictive validity of Huwendiek et al.'s (2014) instrument.

Future research should examine additional sources of validity for Huwendiek et al.'s (2014) VP design effectiveness instrument using larger samples and from other socio-cultural backgrounds and continue to examine the predictive validity of Huwendiek et al.'s (2014) instrument at Level 2 (Learning) and Level 3 (Application) of Kirkpatrick's (1975) four-level model of training evaluation.

This work is dedicated to my loving wife Brandy and my children Courtney, Ryan, Xander, Avery, Mallory, and Conley, whose unconditional love, support and understanding, gave me the strength to finish this study and inspires me everyday to do my best.

This work is also dedicated to my parents, Francisco and Judith, whose love and support enabled me to realize my childhood dreams.

ACKNOWLEDGMENTS

I would like to thank Dr. Atsusi Hirumi for his mentorship, guidance, and support during the last three years and in particular, in the completion of this study. Gratitude is also extended to the other members of my committee, Dr. Stephen Sivo, Dr. Juan Cendan, and Dr. Laurie Campbell for their support and substantial contributions that enabled me to complete my study. Finally, I would like to thank Dr. Teresa Johnson for her assistance in collecting and organizing the data for this study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS/ACRONYMS

AAR	After Action Report
BEME	Best Evidence Medical Education
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CI	Confidence Interval
CINAHL	Cumulative Index to Nursing and Allied Health Literature
CN	Cranial Nerve
COM	College of Medicine
CONSORT	Consolidated Standards of Reporting Trials
EFA	Exploratory Factor Analysis
EMBASE	Excerpta Medica Database
EMS	Electronic Management System
ERIC	Education Resources Information Center
ES	Effect Size
IQR	Interquartile Range
IRB	Institutional Review Board
IVIMEDS	International Virtual Medical School
MD	Medical Doctor
MEDLINE	U.S. National Library of Medicine bibliographic database
NERVE	Neurological Examination Rehearsal Virtual Environment
NID	Network Identification

NIH	National Institutes of Health
OR	Odd Ratio
PsychINFO	American Psychological Association Research Database
QUORUM	Quality of Reporting of Meta-Analyses
RMSEA	Root Mean Squared Error of Approximation
SBME	Simulation Based Medical Education
SD	Standard Deviation
SP	Standardized Patient
SRMR	Standardized Root Mean Square Residual
VIC	Virtual Interactive Case
VP	Virtual Patient
VPS	Virtual Patient Simulation

CHAPTER ONE
INTRODUCTION

Meta-analyses and systematic reviews of literature provide strong evidence for the use of virtual patient simulations (VPs) to facilitate medical education. Based on a meta-analysis of 48 studies, Cook, Erwin, and Triola (2010) concluded that the use of VPs is associated with higher learning outcomes than other educational methods used in medical schools. Similarly, in a meta-analytic review of 14 studies, McGaghie, Issenberg, Cohen, Barsuk, and Wayne (2011) found that simulation-based medical education (SBME) is more effective (ES=0.71) than traditional lecture-based methods, and Cook and Triola (2009) concluded that VPs are best suited to promote the development of clinical reasoning and decision making skills based on a systematic review of literature. A more recent meta-analysis of randomized controlled studies demonstrated a positive pooled overall effect for VPs compared to other educational methods (Consorti, Mancuso, Nocioni, & Piccolo, 2012), and another review of literature supported the integration of virtual patients into the medical curricula (Saleh, 2010). Taken together, syntheses of research comparing the use of VPs to the use of traditional educational methods support the efficacy of VPs. A growing body of research is now emphasizing the importance of systematically studying the design of VPs for facilitating medical education rather than comparing their effectiveness to alternative methods (Bateman et al., 2013; Cendan & Lok, 2012; Jäger, Riemer, Abendroth, Sehner, & Harendza, 2014).

Medical simulations research suggests that learning is less related to the technology, and more related to VP design and integration. To synthesize the results of VPs design studies, a Best Evidence Medical Education (BEME) study revealed that the characteristics of high-fidelity medical simulations (such as VPs) that lead to the most effective learning are (in order of importance): (a) providing feedback, (b) repetitive practice, (c) curriculum integration, (d) range of difficulty level, (e) multiple learning strategies, (f) capture clinical variation, (g) controlled environment, (h) individualized learning, (i) learning outcomes, and (j) simulation validity (e.g., realism, authenticity) (Issenberg, McGaghie, Petrusa, Lee Gordon, & Scalese, 2005). Further studies also supported these characteristics (Cendan & Lok, 2012; Edelbring et al., 2012; Huwendiek et al., 2009a). It appears that how VPs are designed (e.g., what design features are incorporated into the learning activities) and how VPs are integrated into the medical curricula (e.g., what learning activities are carried-out before, during, and after the VPs) may account for more variance in how students learn with VPs than the technology used (Hirumi, 2014; Salas & Gregory, 2011). Given the importance of identifying the design features that make VPs effective for teaching and learning clinical reasoning skills, researchers and practitioners alike may benefit from having a standardized, validated instrument for measuring the effectiveness of the design of VPs.

Subsequent VP design studies produced a variety of design features (Bateman et al., 2013; Botezatu et al., 2010a; Huwendiek & De Leng, 2010), frameworks (Huwendiek et al., 2009b) and principles (Huwendiek et al., 2009a) that are similar in nature, but appear to lack consensus and were derived under considerable research design limitations. In terms of lacking of consensus, Huwendiek et al.'s (2009a) study generated 10 VP design principles while the study by McGaghie, Issenberg, Petrusa, and Scalese (2010) yielded 12 best practices of SBME. Similarly, Botezatu et al. (2010a) produced five best practices for successful use of VPs in medical education while the study by Botezatu, Hult, Kassaye Tessma, and Fors (2010b) resulted in three recommendations for VP design. In addition, the study by Huwendiek and De Leng (2010) produced 10 recommendations for VP design and integration. In terms of limitations, studies by Huwendiek et al. (2009a), Botezatu et al. (2010b), and Huwendiek and De Leng (2010), were limited to measuring participants' reactions, which is the first level of Kirkpatrick's (1975) four-level model of training evaluation. Similarly, a study by Botezatu et al. (2010a) was limited by only measuring reactions from participants and by using a socio-culturally unique sample of Colombian medical students.

Even though VP adoption and use has been steadily increasing (Lang, Kogan, Berman, & Torre, 2013), researchers are not sure which VP design principles to apply to maximize development of clinical reasoning expertise and there are no validated guidelines available throughout the wider literature that medical educators can use to design effective VPs. The lack of uniform VP design principles makes it difficult to create a standardized tool to measure the effectiveness of VP design because without uniform principles, it would be almost impossible to provide a conceptual foundation for the development of an effective standardized tool to measure VP design. A standardized instrument to measure VP design effectiveness would allow researchers and practitioners to measure the effectiveness of their VP designs and make adjustments accordingly to maximize students' development of clinical reasoning and decision making skills. Moreover, "… a short and standardized evaluation tool for VP design is a prerequisite for achieving further expansion of VP use in medical education" (Huwendiek et al., 2014, p. 1).

So, what are the features of effective VP design, and how can the effectiveness of VP design be measured? To answer these questions, Huwendiek et al. (2014) validated an instrument to evaluate the design of virtual patient simulations that focuses on fostering clinical reasoning by examining three sources of evidence: (a) content validity using clinical reasoning theory and an international team of VP experts; (b) response process validity using think-aloud pilot studies and content analysis of answers to each free-text closed question associated with each item of the instrument; and (c) internal structure validity using exploratory factor analysis (EFA) and inter-rater reliability by generalizability analysis.

Using a large sample of medical students (N=2547) across three European countries (i.e., Germany, Poland, and the Netherlands) to evaluate 78 VPs, Huwendiek et al. (2014) found that the three sources of evidence reasonably supported the validity and reliability of the instrument, with at least 200 student responses per VP across all three identified factors (i.e., authenticity of patient encounter and the consultation, cognitive strategies in the consultation, and coaching during consultation). However, Huwendiek et al. (2014) chose not to consider the relationship of their instrument to other criterion-referenced measures.

Problem and Purpose Statement

The problem is that the predictive validity of Huwendiek et al.'s (2014) instrument for evaluating the effectiveness of virtual patient design remains to be examined. In other words, are the design factors evaluated by the instrument correlated to and/or predict students' learning from the VPs and ability to transfer skills learned from the VPs? It is reasoned that variables for informing the design of VPs are useful only to the extent to which they are related to students' subsequent learning and transfer outcomes. The current study aims to examine the predictive validity of Huwendiek et al.'s (2014) VP design effectiveness measurement instrument by determining if the design factors evaluated by the instrument are correlated to the students' ability to learn from the VPs and to transfer clinical reasoning skills learned from the VPs to other clinical situations. Failure to investigate the relationship between students' perceptions of VP design, and students' learning and transfer of clinical reasoning skills learned from VPs would leave important questions unanswered regarding the relationships between these variables and the predictive validity of Huwendiek et al.'s (2014) instrument. To examine the relationship between the students' perceptions of VP design effectiveness and the students' ability to learn and transfer clinical reasoning skills gained from the Neurological Examination Rehearsal Virtual Environment (NERVE) VPs, this dissertation seeks to answer two questions.

Research Questions

1. Does the effectiveness of NERVE VPs design correlate with higher student learning outcomes?

2. Does the effectiveness of NERVE VPs design correlate with higher student skills transfer outcomes?

Hypotheses

To answer the research questions, two hypotheses will be tested:

1. The second-year medical students' perceptions of the effectiveness of NERVE VPs design, as measured by Huwendiek et al.'s (2014) VP design effectiveness evaluation instrument, will correlate significantly with the skills and knowledge gained and sustained from interacting with the NERVE VPs, as measured by quizzes and required assessments of VP cases embedded in NERVE.

2. The second-year medical students' perceptions of the effectiveness of NERVE VPs design, as measured by Huwendiek et al.'s (2014) VP design effectiveness evaluation instrument, will correlate significantly with the ability to transfer skills gained and sustained from interacting with the NERVE VPs, as measured by performance checklists and patient assessments completed during a NERVE-assisted hybrid virtual patient/standardized patient (VP/SP) encounter.

Operational Definitions

For the purposes of this study, *virtual patients* are defined as, "interactive computer simulations of real-life clinical scenarios for the purpose of medical training, education, or assessment" (Ellaway, Poulton, Fors, McGee, & Albright, 2008, p. 170). *Clinical reasoning* is defined as, "the critical analysis of patients' symptoms, signs, laboratory results and imaging, to support the determination of a diagnosis, and the planning of appropriate treatment" (Posel, Mcgee, & Fleiszer, 2014, p. 1). To examine the design of VPs, students' perceptions of effectiveness of NERVE (a suite of VPs and tools designed to teach medical students how to interview, exam, and diagnose patients with cranial nerve disorders) will be measured by using the VP design evaluation instrument (Huwendiek et al., 2014). *Learning* will be measured by examining: (a) students' performance in quizzes embedded in NERVE, and (b) students' performance in required VP cases in NERVE, and *Transfer* will be measured by examining: (a) students' performance in conducting a differential diagnosis of a VP and a standardized patient (SP) presenting the same cranial nerve (CN) palsies, and (b) students' performance in standardized performance checklists completed by SPs.

Method

To help answer the research questions and test the hypotheses posited by this study, a correlational study with a purposive sample of 118 second-year medical students pursuing a Doctor of Medicine (MD) degree at a southeastern United States university's college of medicine (COM) will be examined using the NERVE educational tool and instruments internal and external to NERVE. A correlational research design was used to examine the relationship between the students' perceptions of VP design and two specific outcome measures: (a) clinical reasoning and decision making skills learning from the VPs, and (b) transfer of clinical reasoning and decision making skills acquired from the VP to other clinical situations. All 118 second-year medical students were scheduled to participate in the educational activity on February 2 and either February 3 or 4, 2015, according to their routine schedule. Data collection instruments used includes a VP design survey, quizzes and VP cases in NERVE, VP/SP performance assessments, and standardized performance checklists.

The validated VP design evaluation instrument (Huwendiek et al., 2014) was initially developed as part of the e-VIP project to develop effective VPs that could be shared among medical schools in Europe (Huwendiek & De Leng, 2010; Poulton et al., 2007). Huwendiek et al.'s (2014) instrument asks respondents to report the degree to which they agree or disagree with seven statements on 5-point Likert scales (ranging from strongly agree, agree, neutral, disagree and strongly disagree responses) across three factors (i.e., authenticity of patient encounter and the consultation, cognitive strategies in the consultation, coaching during consultation) and a global score area, to evaluate the design of the VP. There are 12 quizzes (one for each cranial nerve), each containing 10 true-or-false or multiple choice questions about the anatomy, physiology, symptoms, and pathology of cranial nerves to measure students' knowledge of cranial nerve palsies. There are six different VP cases in NERVE presented in two different formats (closed-menu and open-chat) to measure students' ability to perform a patient interview and differential diagnosis, and to formulate a treatment plan. The patient assessments ask students to provide information regarding: (a) localization of the problem, (b) differential diagnosis, (c) evaluation plan, and (d) management & counseling plan, after having experienced a hybrid VP/SP patient encounter. The standardized performance checklist is a list of 15 questions assessing specific behaviors students are expected to exhibit while performing the interview and VP/SP examination. Data will be analyzed using descriptive statistics, multiple regression, and multivariate analysis and results will be reported accordingly. Further details on the method used to complete the proposed correlational study, including additional information on the design of the intervention are provided under Chapter 3. This research has been approved by the IRB (Appendix G).

Conceptual Framework

There is a plethora of theories, models, and frameworks designed to promote the development of clinical reasoning skills. Huwendiek et al. (2014) used the work of Gruppen and Frohna (2002), Bowen (2006), Kim et al. (2006), and Huwendiek et al. (2009a) as the theoretical framework behind their study to validate the VP design evaluation instrument. Gruppen and Frohna (2002) proposed, “a model of clinical reasoning that integrates the key features of several theoretical frameworks” (p. 206). The model depicts an iterative six-step process for clinical reasoning. Bowen (2006) posited six educational strategies to promote development of clinical reasoning. In terms of outcomes, Gruppen and Frohna (2002)’s model appears to be similar to Bowen (2006)’s educational strategies. In other words, following the teaching strategies postulated by Bowen (2006) would make students follow Gruppen and Frohna’s (2002) clinical reasoning and decision making process. Similarly, Kim et al.’s (2006) recommendations for developing teaching cases appear to be very similar to Huwendiek et al.’s (2009a) VP design principles for the development of clinical reasoning skills.

Huwendiek et al.’s (2014) model for the design of the instrument serves as the conceptual framework for this dissertation study. Huwendiek et al.’s (2014) model consists of three variables (the respective elements of each variable in parenthesis): (a) authenticity of patient encounter and the consultation (patient, context characteristics (Gruppen & Frohna, 2002)), (b) cognitive strategies in the consultation (information gathering, problem representation, evaluation (Gruppen & Frohna, 2002), instructional methods (Huwendiek et al., 2009a; Kim et al., 2006)), and (c) coaching during consultation (prior knowledge (Gruppen & Frohna, 2002), instructional methods (Huwendiek et al., 2009a; Kim et al., 2006)). A detailed review of the theoretical as well as empirical foundations for this study is presented in Chapter Two.

Significance of Study

The current study has potential implications for researchers and practitioners. Potential implications for researchers include: (a) providing empirical evidence supporting the predictive validity of Huwendiek et al.'s (2014) VP design effectiveness evaluation instrument, and (b) encouraging additional research on examining the predictive validity of Huwendiek et al.'s (2014) instrument at Level 2 (Learning) and Level 3 (Application) of Kirkpatrick's (1975) four-level model of training evaluation. Potential implications for practitioners (i.e., medical educators, instructional designers) include: (a) providing empirical evidence supporting the use and integration of virtual patients in the medical curricula, (b) providing empirical evidence supporting the design of VPs based on the design features and principles evaluated by Huwendiek et al.'s (2014) instrument, (c) bridging the gap between the fields of instructional design and medical education, and (d) increasing collaboration between medical educators and instructional designers. Future research should examine additional sources of validity for Huwendiek et al.'s (2014) VP design effectiveness instrument using larger samples and from other socio-cultural backgrounds.

In summary, Chapter One established the context for this study by providing a brief review of the theoretical as well as empirical underpinnings of this study. The significance of examining the predictive validity of Huwendiek et al.'s (2014) VP design evaluation instrument as it relates to learning and transfer of clinical reasoning skills was discussed. The purpose for this dissertation was given followed by a rationale discussing the implications of this dissertation study for researchers and practitioners. A detailed review of the model for the design of Huwendiek et al.'s (2014) VP design evaluation instrument is presented in Chapter Two.

CHAPTER TWO
LITERATURE REVIEW

Chapter One established the context for this study by providing a brief review of the theoretical as well as empirical foundations for this study. Research regarding the benefits of VPs for medical education and availability of a standardized tool for measuring the effectiveness of VP design was briefly discussed, followed by explicit problem and purpose statements, and research questions, and a brief overview of the methods to be used to answer the research questions. Chapter One also discussed the significance of examining the predictive validity of Huwendiek et al.'s (2014) instrument. Chapter Two presents a synthesis of research on VP design followed by a detailed review of the model for the design of Huwendiek et al.'s (2014) VP design evaluation instrument.

Chapter Two begins with a discussion of the origin and evolution of medical simulations as a way to distinguish VPs from other types of simulations. Then, this chapter summarizes the research on VP design that provides empirical evidence supporting the effectiveness of VPs while highlighting the need to investigate the predictive validity of Huwendiek et al.'s (2014) instrument for VP design evaluation. Subsequently, this chapter describes the theoretical framework used to develop Huwendiek et al.'s (2014) VP design evaluation instrument. Finally, Chapter Two presents the model illustrating the relationship between the variables under examination in this dissertation and discusses additional empirical evidence supporting the relationship between the variables or factors in Huwendiek et al.'s (2014) instrument for VP design evaluation and the achievement of higher learning and transfer outcomes.

Evolution and Effectiveness of VP Simulations

Medical educators have long wrestled with the challenge of educating medical students while simultaneously providing high-quality healthcare for their patients. However, due to the lack of alternate methods of training (other than practicing with patients), education continues to advance at the expense of patient healthcare. It was not until the development of the *Resusci-Anne* manikin for Cardio-pulmonary resuscitation and later, the *Sim-One* manikin (1960s) and the explosion in high-fidelity mannequin simulators in the 1980's that lower-cost alternatives were developed (Bradley, 2006). High fidelity and lower cost options eventually led to widespread adoption. In turn, increased demand coupled with advancements in technology led to even more sophisticated models to include computer-based models such as the VP simulations available today. So, how are simulations defined and how do they compare to traditional methods of medical education?

Simulations are defined as "the technique of imitating the behaviour of some situation or process (whether economic, military, mechanical, etc.) by means of a suitably analogous situation or apparatus, especially for the purpose of study or personnel training" (Bradley, 2006, p. 254). Thus, simulations could be used for most instructional situations, including healthcare. For instance, in a meta-analytic review of 14 studies, McGaghie et al. (2011), who investigated the comparative effectiveness of simulation-based medical education (SBME) compared to traditional (e.g., lecture-oriented) medical education found that SBME is more effective (ES=0.71) than traditional, lecture-based methods. Healthcare simulations are not only more effective educational training tools but also safer and cheaper than other methods.

Advantages of Simulations-Based Training Systems for Medical Education

According to Chodos et al. (2010), simulation-based training systems (SBTS) in a virtual environment provide a safe, realistic, and more cost-effective alternative for medical students to practice when compared with the use of standardized patients (SPs) or professional actors. SPs, for example, must be continually trained to be able to present the clinical history, symptoms and pathology desired. Since SPs are paid for their time, to include training and performance, costs grow exponentially as SPs are increasingly used to present students with a variety of clinical cases. Similarly, some medical schools use professional actors (Shapiro & Hunt, 2003), which may be even costlier. Although SBTS may be expensive to develop, once operational, costs are abated as medical schools continue to reduce or even eliminate yearly recurring costs associated with training and performance of SPs and professional actors. It also helps improve patient safety as SBTS are being used today to present students with rare, difficult to imitate and diagnose conditions such as cranial nerve palsies, increasing student expertise in conditions students would have otherwise experienced for the first time in their clinical practice (Cendan & Lok, 2012). Given that VPs are examples of SBTS, one must wonder how VPs are defined and used in medical schools.

VPs, which are defined as "interactive computer simulations of real-life clinical scenarios for the purpose of medical training, education, or assessment" (Ellaway et al., 2008, p. 170), are effective for the development of clinical reasoning skills, and their pervasive use across borders is regulated by international technical standards (Bateman et al., 2013). However, VPs can be expensive, therefore, some medical schools may only have mannequins while others may have both or perhaps none. So, what characteristics set one apart from the other? What are the advantages and disadvantages of using one type of simulation over another?

Taxonomy of Simulation Based Training

To distinguish simulations and compare relative advantages and disadvantages, Van Hook (2004) as cited by Chodos et al. (2010), developed the *taxonomy of simulation-based training,* which classifies medical simulations under five categories based on fidelity: (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 and physiology, clinical reasoning and decision making). However, computer technology has changed significantly since 2004, making screen-based computer simulations highly interactive and realistic. Thus, while medical simulations using standardized patients would fall under the lower tier, realistic patient simulators like computer-driven mannequins and highly interactive VPs that can provide oral responses and follow learner commands (e.g., raise your hand, tilt your head, close one eye) would fall on the highest tier. Since more realistic simulations are associated with higher learning outcomes (Issenberg et al., 2005), high-fidelity simulations (e.g., computerized mannequins, highly interactive VPs) are expected to be more advantageous for teaching and learning.

Effectiveness of VP Simulations for Facilitating Clinical Reasoning

The purpose of this study is to examine the relationship between the students' perceptions of VP design effectiveness, as measured by Huwendiek et al.'s (2014) VP design evaluation instrument, and students' ability to learn and transfer clinical reasoning skills acquired from the VPs. Huwendiek et al.'s (2014) instrument was specifically devised to evaluate VPs designed to foster clinical reasoning, thus, clinical reasoning skills are first defined and strategies for developing clinical reasoning skills are discussed before examining the effectiveness of VP simulations.

Defining Clinical Reasoning and Decision Making Expertise

Posel et al. (2014) defined clinical reasoning as the "critical analysis of patients' symptoms, signs, laboratory results and imaging, to support the determination of a diagnosis, and the planning of appropriate treatment" (p. 1). Croskerry (2009) as cited by Posel et al. (2014) found that the fundamental components of clinical reasoning include: "(a) the availability of necessary domain knowledge, (b) association of this knowledge with evidenced-based research, and (c) its subsequent application through decision-making, high-quality clinical judgment and active problem-solving" (p. 1). Clinical reasoning skills expertise requires inductive, analogical or intuitive reasoning (Croskerry, 2009). For instance, experts, who possess a broad collection of mental representations (or knowledge base) of clinical cases based on experience rely more on analogical reasoning (e.g., heuristics, pattern recognition, unconscious thinking) to solve complex problems while novices mainly use analytical thinking (e.g., deductive, normative reasoning, deliberate thinking) (Croskerry, 2009). Norman (2005), who conducted a systematic review of the literature on clinical reasoning published during the previous 30 years found that clinical reasoning, "expertise is associated, not with a single basic representation but with multiple coordinated representations in memory, from causal mechanisms to prior examples" (p. 418) and that the development of clinical reasoning expertise requires "deliberate practice with multiple examples and feedback, both to facilitate effective transfer of basic concepts and to ensure an adequate experiential knowledge base" (p. 425). Norman (2005) also found that "expert clinical reasoning really amounts to expert diagnostic reasoning" (p. 425), underscoring the parallels between clinical reasoning and performing a differential diagnosis. Given the critical nature of clinical reasoning skills for effective diagnosis and treatment, researchers and practitioners began to explore educational tools for developing clinical reasoning expertise.

Educational Tools for Promoting Development of Clinical Reasoning Skills

Many educational tools and techniques have been posited for developing clinical reasoning skills. Typically, these techniques (e.g., instructional methods, strategies) can be implemented using a variety of tools. In an e-learning environment, for example, "learners interact with tools to complete tasks both within and outside the computer environment" (Hirumi, 2006, p. 51). The educational tools available to facilitate instructional or e-learning activities include: (a) telecommunication tools, (b) productivity tools, and (c) other tools used outside the computer environment (Hirumi, 2006). Similarly, tools used for the teaching and learning of clinical reasoning skills may include telecommunication and productivity tools (e.g., computers, phones, tablets, word processors, spreadsheets, databases, electronic mail, online discussion forums, learning management systems) as well as medical manikins, standardized patients, and virtual patient simulations, among others. However, inclusion of educational tools into the medical school curriculum, by itself, may not result in higher learning outcomes without properly applying instructional techniques or strategies that can leverage these tools to accomplish specified clinical reasoning learning objectives. In other words, even if medical educators understand the nature, scope, and particularities of what they want to teach (e.g., clinical reasoning and decision making skills) and have the proper educational tools to teach it (e.g., virtual patient simulations), their efforts may not be effective without using instructional methods that have been both empirically and practically supported for the development of clinical reasoning expertise.

VP Effectiveness in Fostering Development of Clinical Reasoning Expertise

Research indicates that VPs are more associated with higher learning outcomes compared to other medical scholastic methods (Consorti et al., 2012; Cook et al., 2010). For instance, Cook and Triola's (2009) systematic review of over 25 studies found significant positive differences in learning outcomes when using VPs compared to no intervention, but no differences were observed between VPs and other interventions. Similarly, a systematic review of literature and meta-analysis concluded that VPs are, "associated with large positive effects compared with no intervention" (Cook et al., 2010, p. 1589). In a meta-analytic review of 14 studies, McGaghie et al. (2011) found a large effect size 0.71 (95% confidence interval, 0.65–0.76; $P < .001$), and concluded that simulations, to include, VP simulations, are some of the most effective methods for the teaching and learning of clinical reasoning skills. Similarly, a recent meta-analysis of randomized controlled studies showed a positive pooled overall effect for VPs compared to other educational methods (Consorti et al., 2012), while another review of literature supported the integration of virtual patients into the medical curricula (Saleh, 2010).

Taken together, syntheses of research comparing VPs to the use of traditional educational methods indicate that VPs simulations are more effective than other educational methods, and best suited for developing clinical reasoning skills. According to Cook and Triola (2009), VPs "are likely to play an increasing role in medical education in coming years. However, their effective use requires evidence to guide design and integration. This evidence base is currently virtually non-existent, at least in published form" (p. 308). More research is needed examining the design of VP simulations to optimize learning.

Characterizing Research on VP Design for Medical Education

A growing body of research now emphasizes the importance of design for facilitating medical education. “Apparently, learning outcomes depend on the quality of instructional design methods rather than on the medium used” (Huwendiek et al., 2009a, p. 581). Research suggests that learning is less related to the technology, and more related to VP design (Cendan & Lok, 2012; Edelbring, Dastmalchi, Hult, Lundberg, & Dahlgren, 2011; Jäger et al., 2014).

The following section reviews VP design studies in chronological order to illustrate the evolution of design studies and how VP design variables have changed over time. At the end, a synthesis of VP design research is offered to examine the degree of consensus in published VP design research regarding the design features that should inform the design of VPs.

Issenberg et al.’s (2005) Best Evidence Medical Education (BEME) synthesis of research revealed that the characteristics of high-fidelity medical simulations (such as VPs) that lead to the most effective learning are: (a) providing feedback, (b) repetitive practice, (c) curriculum integration, (d) range of difficulty level, (e) multiple learning strategies, (f) capture clinical variation, (g) controlled environment, (h) individualized learning, (i) learning outcomes, and (j) simulation validity (e.g., realism, authenticity). Since Issenberg et al.’s seminal review of research, a number of investigations have been completed on VP design. To illustrate the evolution of VPs design studies, the following section reviews investigations published since 2005 in chronological order as illustrated in Table 1.

Cook and Triola (2009) synthesized research comparing VPs versus other methods, and distilled key design features based on their review. Researchers searched the MEDLINE, EMBASE, CINALH, and PsychINFO databases for articles using the terms *virtual patient*, *computer simulation*, *education professional*, and *clinical education.* Other studies were identified from other articles suggested by the search engine, reference lists inside the articles included in the search, and author's files. Researchers specifically selected quantitative comparative research studies that reported outcomes data, pre-assessment data, or rigorous qualitative studies. No information was provided regarding the exact number of articles selected for review but upon examination, the amount appears to be more than 25 articles.

Cook and Triola's (2009) systematic review of literature found significant positive differences in learning outcomes when using VPs compared to no intervention, but no differences were observed between VPs and other interventions. Moreover, VPs were found deficient for development and assessment of affective skills, which are important core skills in medical practice. In addition to comparing the relative effectiveness of VPs, Cook and Triola (2009) also indicated that "many principles identified in a recent review of high-fidelity simulation are probably applicable to VPs; these include the provision of: feedback; repetitive practice; progressive difficulty, and clinical variation (Table 1). These principles are well grounded in theory and other evidence" (p. 307). Cook and Triola (2009) concluded that VPs are best suited to promote the development of clinical reasoning and decision making skills (Table 1). Finally, Cook and Triola (2009) found that only a few studies examined presentation format, and concluded more research was needed to examine the design of VP simulations.

Huwendiek et al. (2009a) explored what students' perceive as ideal features of VPs that were developed using the CAMPUS-Classic VP authoring tool (http://www.virtualpatients.eu) and designed to foster clinical reasoning using a posttest-only design. The VPs were presented in two modes that differed in terms of the presence or absence of graphics support (authenticity of the user interface), long or short questions, and freedom of navigation as well as in the use of different media and questions with explanations. Participants belonged to a cohort of 104 fifth-year medical students completing their pediatric medicine training at the Heidelberg University in Germany. Students worked on eight VPs in four different designs followed by a 30-minute instructor-guided discussion of the cases. These students would later interact with real patients presenting the same conditions of the VPs before interviews were conducted. Twenty-seven students ($N = 27$) were randomly selected to participate in the five focus group interviews using a questioning route. The interviews were video recorded, transcribed, and analyzed using what appears to be *content analysis* resulting in 10 themes, and agreed upon by reaching consensus. Huwendiek et al. (2009a) found that the VP design principles that lead to higher learning outcomes are (Table 1): (a) a VP should be relevant, (b) of an appropriate level of difficulty, (c) highly interactive, (d) offer specific feedback, (e) make optimal use of media, (f) help students focus on relevant learning points, (g) offer recapitulation of key learning points, (h) provide an authentic web-based interface and student tasks, and (i) contain questions and explanations tailored to the clinical reasoning process.

Cook et al. (2010) conducted a systematic review and meta-analysis to examine the effects of VPs compared to no intervention or alternate methods of instruction to identify key VP design features. This review was conducted in accordance with the QUORUM, MOOSE, and PRISMA standards for reporting meta-analyses (Moher et al., 1999; Moher, Liberati, Tetzlaff, & Altman, 2009). Cook et al. (2010) searched the MEDLINE, EMBASE, ERIC, CINALH, Scopus and PsychINFO databases through February 2009 for studies using the terms *virtual patient*, *computer simulation*, *problem-based learning*, *case-based learning*, *medical education*, and *clinical simulation* using no beginning cut-off date. Other studies were identified from reference lists inside the articles included in the search. Researchers worked independently and in duplicate to screen and select all articles for inclusion into the study and to workout disagreements with a final inter-rater correlation coefficient of .69. Out of 151 potentially eligible articles, 49 (32%) met the selection criteria and were included. Studies were classified as *descriptive*, *no-intervention controlled*, *media-comparative*, or *computer-assisted comparative* as well as studies with rigorous qualitative analysis. In addition, an adaptation of the Newcastle-Ottawa scale for cohort studies was used to grade methodological quality. Researchers reported descriptive statistics and weighted and pooled effect size estimates. Cook et al.'s (2010) meta-analysis found large pooled effect sizes compared to no-intervention for: (a) knowledge outcomes (ES=.94, N=11); (b) clinical reasoning (ES=.80, N=5); and (c) other skills (ES=.90, N=9). Researchers found small average pooled effect sizes compared to non-computer instruction on: (a) satisfaction (ES=-.17, N=8); (b) knowledge (ES=.06, N=5); and (c) other skills (ES=.10, N=11). With regard to VP design features, Cook et al. (2010) found (see Table 1) "mastery learning, advance organizers, enhanced feedback, and explicitly contrasting cases improved learning outcomes in randomized trials, with ESs ranging 0.29 to 1.47" (p. 1599). Additionally, "studies further suggest that natural case evolution and working as groups are important" (Cook, et al., 2010, p. 1599).

Botezatu et al. (2010a) examined the students' perceptions of the educational use of VPs using focus group interviews and a posttest-only design. The linear, interactive VPs were designed to foster clinical reasoning and developed using a Spanish-only variant of the Web-SP virtual patient application from the Karolinska Institutet in Sweden. The Spanish-only variant had been in use in the curriculum since 2005 and the VPs contained patient photographs and diagnostic media. Upon VP case completion, students received expert feedback from a senior clinician in the form of a detailed case discussion and the actual patient follow up. Participants belonged to a cohort of 49 students completing their internal medicine rotation at the

Universidad el Bosque in Colombia. Sixteen students (N=16) were selected through simple randomization to participate in the interviews. The interviews were conducted in Spanish and audio and video recordings were made, transcribed using a non-verbatim approach, and translated back into English. The data was analyzed using *content analysis.* Eighteen categories were identified and clustered into five themes and agreed upon by reaching consensus. Botezatu et al. (2010a) found five factors associated with effective VP simulations use in medical education (Table 1): (a) learning, (b) teaching, (c) assessment, (d) authenticity, and (e) implementation, indicating that "medical students perceive VPs as important learning and assessment tools, fostering clinical reasoning, in preparation for the future clinical practice" (p. 1).

Botezatu et al. (2010a) argued that VPs are particularly suited for the teaching, learning, and assessment of clinical reasoning skills, and that skills acquired by learning with VPs are transferable and retained for longer periods. Findings suggest the design of VPs should include learning theories, and teaching and assessment strategies that promote the development of clinical reasoning skills. Including assessment into the design seem to be particularly important to enhance the effectiveness of the VP design as "students consider the VPS to be a more didactic form of evaluation and an intrinsically better evaluation tool than traditional exams" (Botezatu et al., 2010a, p. 5). Botezatu et al. (2010a) also found that authenticity has been associated with knowledge and skills transfer to real patients. Presumably, transferability is enhanced when using real patient cases, where students make associations between the VP cases and the real patients. However, authenticity should go beyond the clinical cases to include the provision of feedback regarding the evolution of patient condition and "to reflect the real clinical practice" (Botezatu et al., 2010a, p. 5). Finally, researchers recommended that VP simulations should be designed to provide clinical variation through implementation across "all major clinical specialties" and should include "frequent diseases" and "topics not included in the study plan" (Botezatu et al., 2010a, p. 4), which highlights *clinical variation* as an important design element of VP simulations (Table 1).

Botezatu et al. (2010b) set out to investigate the stakeholders' perceptions of the design of Web-based Simulation of Patients (Web-SP) at three universities in Colombia and Sweden using a cross-sectional survey with free-text questions and a posttest-only design. The linear, interactive Web-SP VPs were designed to foster clinical reasoning and developed using both a Spanish-only variant and the Swedish version of the Web-SP virtual patient application from the Karolinska Institutet in Sweden. The Spanish-only variant had been in use in the curriculum since 2005 and both VP suite versions contained patient photographs and diagnostic media. Upon VP case completion, students received expert feedback from a senior clinician in the form of a detailed case discussion and the actual patient follow up. Five faculty board members, seven college teachers, and 16 randomly chosen students from the Universidad el Bosque were invited to take the survey. In Sweden, three board members and eight professors were invited but no students were selected since Web-SP VPs were optional at the time a potential source of students for the sample had their rotation. The survey was written in both English and Spanish, quantitatively and qualitatively validated, and administered to participants (N=39). Scores were analyzed using descriptive and inferential statistics to determine the level of preference respondents had in rating the ranking of the variables and the concordance of their responses. No content analysis was needed since respondents provided very brief answers but the responses supported the statistical findings.

Botezatu et al. (2010b) found that VP design should (see Table 1): (a) allow extensive editing, (b) support case authenticity, and (c) enhance clinical reasoning abilities. According to Botezatu et al. (2010b), “the aspect of paying attention to VPS design, which should enhance clinical reasoning abilities and support case authenticity, cannot be overemphasized. Authenticity, however, extends well beyond the design of the interface. The users are more positive to the use of an application when the case content is robust, derived from everyday practice and supported by feedback providing an exposé of actual patient treatment and evolution” (p. 516).

Huwendiek and De Leng (2010) developed two standardized instruments to evaluate the design of VPs that are focused on the development of clinical reasoning skills. These instruments are: (a) a checklist that allows evaluators to characterize in detail the design of a VP, and (b) a questionnaire to assess the students' experience with the VP. The instruments were iteratively developed and optimized, and properly validated quantitatively and qualitatively. In addition, investigators used literature on teaching clinical reasoning and the design of teaching cases, and the results of a focus group study on key features of VP design to inform the development of a 14-item VP design questionnaire. Each VP design instrument contains open-ended questions and is available at (http://www.virtualpatients.eu/resources/evaluation-tool-kit/). Researchers reported that the first series of studies have confirmed the effectiveness of the instruments to evaluate VP design. The instruments are being tested at other European institutions. The results provide support for prior research concluding the VP design principles that lead to higher learning outcomes are (see Table 1): (a) a VP should be relevant, (b) of an appropriate level of difficulty, (c) highly interactive, (d) offer specific feedback, (e) make optimal use of media, (f) help students focus on relevant learning points, (g) offer recapitulation of key learning points, (h) provide an authentic web-based interface and student tasks, and (i) contain questions and explanations tailored to the clinical reasoning process (Huwendiek, et al., 2009a).

Adams, Rodgers, Harrington, Young, and Sieber (2011) investigated students' use and perceptions of video-enriched VP cases for primary care in medical education. Participants were undergraduate medical students at University of Oxford in the United Kingdom. The VPs were designed to guide students through a specific diagnostic strategy consisting of spot diagnosis, pattern recognition, patient self-report, classic presenting complain, restricted rule outs, red flags, probabilistic reasoning, test of treatment, test of time, and follow up and referral to secondary care, using patient video media files while prompting them, at times, with both true and false, and multiple choice questions. An independent reviewer was used for content accuracy. Researchers did not mention the VP authoring tool used but the VPs were created using decision-making software, linear, and video-based (e.g., videos clips) content only. Investigators collected data on three cohorts of students (N=82) during their primary care rotation, regarding their perceptions of accessibility, popularity, and disadvantages as well as number of attempts and VP case completion rates. Each VP also offered students an opportunity to provide additional comments via free-text boxes at the end of each VP case. Researchers analyzed the resulting data quantitatively for VP usage logs (i.e., frequency, duration) and qualitatively for answers to free-text questions. Adams et al. (2011) found that students liked the realistic and interactive nature of the VPs, which underscores the importance of *authenticity* and *interactivity* in VP design, supporting prior research by Botezatu et al. (2010a), Botezatu et al. (2010b), Cook and Triola (2009), and Huwendiek et al. (2009a). Moreover, Adams et al. (2011) found "VP cases can be tailored using decision-making software to give students experience of primary care specific clinical reasoning skills" (p. 273) and that students used VPs multiple times during their primary care rotation, primarily for revision purposes.

Gormley, Mcglade, Thomson, Mcgill, and Sun (2011) investigated the perceived *usability* of an e-learning package of 10 online VPs that was developed with the IVIMEDS (www.ivimeds.org) *Riverside* VP authoring tool and designed to foster clinical reasoning using a posttest-only design. The usability of a product primarily depends on content, interactivity, and user satisfaction (Gormley et al., 2011). Participants were 21-31 years-old undergraduate medical students at Queens's University in Belfast, Ireland. The VPs were designed to guide students through history taking, examination, investigation, and decision making stages while prompting them, at times, with both free-text and multiple choice questions. Investigators used an online questionnaire that was developed to capture the respondents': (a) opinions on the usefulness of the VP package, (b) opinions regarding how the VP package compares to other scholastic methods, (c) free-text comments regarding the usefulness of the VPs, and (d) perceptions of usability. The questionnaire was emailed to each of the six module groups upon completion of the general practice module and through an email reminder two weeks later. The data was analyzed using simple descriptive statistics and the free-text comments were analyzed using *content analysis* and themes were extracted and agreed upon by reaching consensus. Respondents (N=149) found the VP package had a high-level of usability (88.8%), as measured by the System Usability Scale (Brooke, 1996).

The 10-VP package used in Gormley et al.'s (2011) study provided a high degree of clinical variation and used *usability* as a measure of the quality of the content, user satisfaction, and interactivity. The findings suggest VP *case content* (e.g., what is presented, how it is presented, perceived authenticity), *clinical variation*, and *interactivity* are important VP design features (Table 1). In addition, students were asked to reflect about what they learned and how it will affect their clinical practice in the future, which also emphasizes *facilitating reflection* as an important VP design feature. Gormley et al. (2011) concluded that VPs are user-centered, user-friendly, and useful for development of clinical reasoning skills in primary care.

Cendan and Lok (2012) reviewed the literature on VP design and provided results from their research on VP design. Cendan and Lok (2012) developed the NERVE VP suite by applying Kolb (1984)'s experiential learning theory principles, which includes provisions for *repetitive practice*, *active participation*, *observation*, *individualized feedback* and *reflection* as the main characteristics of NERVE VP design. Researchers also presented a "visualization tool to allow the instructor to identify the manner in which a student using a VP has reached a final diagnosis" (Cendan & Lok, 2012, p. 50). According to Cendan and Lok (2012), VPs "facilitate the provision of feedback and represent a venue for safe and repetitive practice as well as a model where progressive clinical variation and difficulty can be presented" (p. 48). In addition "the main components of VPs include interactivity on the learner's part (as opposed to passively watching videos)" (Cendan & Lok, 2012, p. 48), leading to more efficient learning and higher retention rates compared to non-interactive education methods. Cendan and Lok (2012), also indicated "VPs can also provide students opportunities for self-directed learning, which leads to reflection, self-driven change, and more insight into performance (p. 49). Therefore, "educators should consider the learning objectives that they are trying to achieve as to motivate the design decisions in the creation of VP simulations" (Cendan & Lok, 2012, p. 51). Finally, Cendan and Lok (2012) found that students in small groups learned better from NERVE than individual users and posited that the "cognitive load of the activity may account for the difference in learning" (p. 51). Cendan and Lok (2012) concluded that higher quality research in VP design is needed to address the issues that hamper widespread use of VPs for the teaching and learning of clinical reasoning skills.

Wilson (2012) set out to explore students' perceptions of a VP for emergency medicine instruction. A convenient sample of seven participants worked on four low-fidelity branching narrative VPs created using the Open Labyrinth authoring tool (http://openlabyrinth.ca) as part of the requirements for the diploma in emergency medicine at the University of West Indies in Trinidad and Tobago. Participants (N=7) were interviewed about their experiences with the VP using questions that were borrowed from an instrument that was previously validated (see Huwendiek & De Leng, 2010) to evaluate the design of VPs. The interviews were recorded, transcribed, and analyzed using *thematic analysis* and agreed upon by reaching consensus. Wilson (2012) found that (see Table 1): (a) VPs are game-informed learning, (b) a VP tutorial is collaborative learning, (c) a VP is an authentic activity, (d) VPs encourage reflection, (e) VPs encourage clinical reasoning, and (e) VPs should be integrated into the curriculum.

According to Wilson (2012), VPs are game-informed learning because students thought VPs were "engaging, motivating, interesting, enjoyable, exciting, stimulating, interactive, and it made learning fun and easy" (p. 117). Moreover, "the characteristics of multiple paths, navigation, feedback, choices and consequences, which are also associated with game play, were mentioned with respect to virtual patients" (Wilson, 2012, p. 117). Wilson (2012) found that "learning was viewed as a social activity with the opportunity to talk and share diverse perspectives. There was general consensus that there was much to be learned from one's peers" (p. 118). Thus, findings suggest VPs may benefit from deliberate inclusion of constructivist theories of learning into the design. Wilson (2012) also found VPs were perceived as "realistic, relevant, and developed for the local environment. They felt that they were facing people and situations they would see in practice and reacting as they would in the emergency department" (p. 119).

The findings also underscore the importance of *authenticity* (e.g., content, language, clinical data, clinical context) as a key characteristic of VP design. Wilson (2012) found VPs encourage reflection through the provision of feedback. According to Wilson (2012), "providing feedback supports the reflective process. Reflective elements built into the VP that prompted reflection were: allowing the user to review choices and providing feedback. Feedback allowed the user to consider the consequences of their actions. Guided reflection by a mentor is particularly important in professional practice. The mentor facilitates reflection, while supporting and challenging the learner's beliefs and assumptions" (p. 121). Findings suggest *providing feedback* and *opportunities for reflection* are essential characteristics of VP design. Finally, Wilson (2012) found learners "agreed that their clinical reasoning and diagnostic skills were enhanced by working through VPs, and they felt better prepared to care for real patients with similar complaints" (p. 121) and that "VP provided structure and feedback, which helped them order their thoughts, proceed in a stepwise fashion, and refine their mental models. VP corrected them, prioritized information, and reinforced protocols and what they knew" (p. 121), which supports prior research regarding VPs best use to promote the development of clinical reasoning skills (e.g., Cook & Triola, 2009).

Bateman, Allen, Kidd, Parsons, and Davies (2012) investigated the effect of the presence or absence of two design variables, branching case pathways (Br) and structured clinical reasoning feedback (CRF) on clinical reasoning skills and student experience using a randomized 2x2 factorial pretest-posttest randomized controlled trial (RCT) design. The study was carried out using volunteer medical students (N=112) from the same year group at three different European medical schools. Students were randomly assigned to one of four groups using block randomization and each group was linearly exposed to one of four possible conditions (i.e., Br/CRF, Br/noCRF, noBr/CRF, noBr/noCRF) using 16 VPs created in accordance with the Medbiquitous international standard (http://www.medbiq.org/std_specs/standards/index.html). Each VP contained the same standardized 15-item clinical reasoning assessment and the same standardized 15-item self-reported evaluation. At one facility, researchers recorded patterns of use and pre- and post-intervention clinical reasoning scores using the Diagnostic Thinking Inventory (DTI) as well as formative and summative written and clinical assessments measured one week after intervention and several months later. The results of this study have not yet been published but researchers plan to analyze the data using descriptive and inferential statistics to identify main effects, effect sizes, and interactions among the design variables, namely feedback and branching. According to Bateman et al. (2012), "there are numerous VP design properties identified in the literature … of particular interest are, firstly, the use of branching case pathways and secondly, the role of structured feedback to promote clinical reasoning" (p. 2). Bateman et al. (2012) thus regarded *branching case pathways* and the *provision of structured feedback* as important VP design features that foster learning with VPs (Table 1).

Consorti et al. (2012) set out to explore the effects of using VPs as an alternative or additive to the curriculum versus conventional methods of medical education through a meta-analysis conducted in accordance with the QUORUM standard for reporting meta-analyses (Moher et al., 1999). Consorti et al. (2012) searched the MEDLINE, ERIC, Cochrane, and PsychINFO databases from January 2000 through July 2010 for studies using the terms *virtual patient*, *computer simulation*, *distance education*, *virtual system*, *distance training, and internet-based care.* Other studies were identified from reference lists inside the articles included in the search. Researchers worked independently and in duplicate to screen and select all articles for inclusion into the study and to workout disagreements. Only randomized studies with control groups were included. Out of 34 potentially eligible articles, 12 (35%) met the selection criteria and were included. Methodological quality was assessed by selecting only comparative studies that met at least 11 of 21 considered items in the CONSORT statement (Moher, Schulz, & Altman, 2001) and that reported the pre- and post-intervention figures required to calculate effect sizes. Researchers reported descriptive statistics and weighted and pooled effect size (i.e., odds ratios and confidence intervals) estimates. Odds ratios (ORs) are well-known measures of effect size used in meta-analyses and "represent the odds that an outcome will occur given a particular exposure, compared to the odds of the outcome occurring in the absence of that exposure … and an OR>1 means that the exposure is associated with higher odds of outcome" (Szumilas, 2010, p. 227). Consorti et al.'s (2012) meta-analysis found a clear positive pooled overall effect for VPs compared to other educational methods (Odds Ratio: 2.39; 95% C.I. 1.48 / 3.84), even when VPs were added to the curriculum (O.R.: 2.55; C.I. 1.36 / 4.79) or when compared to traditional methods of medical education (O.R.: 2.19; 1.06 / 4.52). Odds ratios are well-known measures of effect size used in meta-analyses and "represent the odds that an outcome will occur given a particular exposure, compared to the odds of the outcome occurring in the absence of that exposure … and an OR>1 means that the exposure is associated with higher odds of outcome" (Szumilas, 2010, p. 227).

Regarding VP design, Consorti et al. (2012) conceptualized VPs, as proposed by Ellaway and Davies (2011), "from a software artifact to an intrinsic part of an activity that mediates the ways that learners and their objectives interact" (p. 1002). This conceptualization, based on *activity theory*, stresses the value Consorti et al. (2012) placed in integrating *activity theory* and perhaps other types of *learning theories* into the design of VPs. Consorti et al. (2012), also regarded VPs as interactive, suggesting that "in being challenging for the learner, VPs are believed to foster … active learning" (p. 1005). According to Consorti et al. (2012), "it could be argued that at least part of the positive effect of VPs could rely on a longer period of work spent by the learners" (p. 1006), suggesting that VP learning activities designed to prolong *time-on-task* may result in higher learning outcomes. Finally, Consorti et al. (2012) indicated "clinical reasoning is … strongly dependent on the mental database of cases owned by healthcare professionals" (p. 1006) and that "if the assessment of clinical skills after the exposure to an even limited set of cases on a specific topic is highly consistent with the training cases themselves, we may expect a good performance" (P. 1006), suggesting that *clinical case content, variation,* and *specificity* are important VP design features to optimize learning (Table 1).

Bateman et al. (2013) explored students' perceptions of a model that explains the core phenomenon representing the students' interactions with VPs used in medical education to foster the development of clinical reasoning skills. Two VPs were created in accordance with the Medbiquitous international standard (http://www.medbiq.org/std_specs/standards/index.html) and integrated the following design variables: (a) branching and linear case narratives, (b) freedom of navigation through the case, (c) visible scoring systems, (d) different question types such as multiple-choice questions, (e) key feature problems, and (f) Bayesian reasoning (Bateman et al., 2013). Participants (N=46) selected through purposive iterative sampling worked on the VPs and then participated in an evaluation and six focus group interviews. The interviews were recorded, transcribed, and iteratively analyzed using *grounded theory analysis* resulting in a central core phenomenon named *learning from the VP* and four categories that were used to construct a three-layer model (Figure 1) representing how students learn from VPs. Specifically, Bateman et al. (2013) indicated "our work supports 10 general authoring recommendations produced from a thematic analysis of VPs in focus groups" (p. 602). Bateman et al. (2013) was referring to Huwendiek et al.'s (2009a) design principles: (a) a VP should be relevant, (b) of an appropriate level of difficulty, (c) highly interactive, (d) offer specific feedback, (e) make optimal use of media, (f) help students focus on relevant learning points, (g) offer recapitulation of key learning points, (h) provide an authentic web-based interface and student tasks, and (i) contain questions and explanations tailored to the clinical reasoning process.

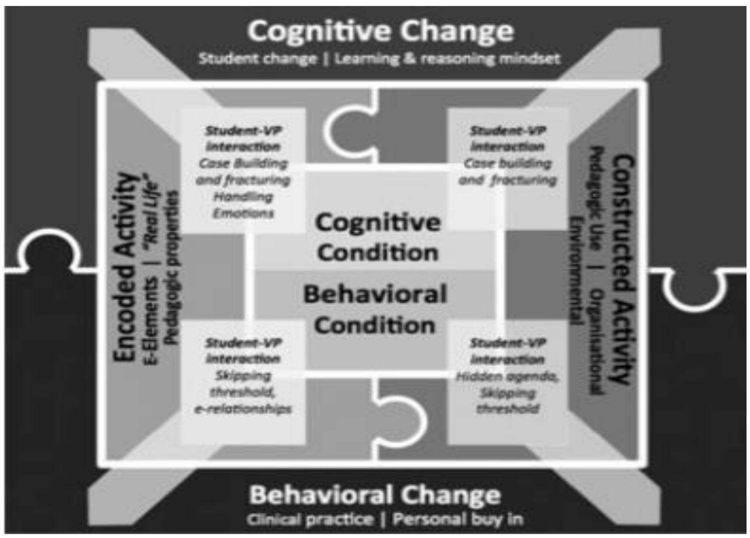


Figure 1: Virtual patient (VP) Implementation Model[a]
[a]Adapted from: Bateman, J., Allen, M., Samani, D., Kidd, J., & Davies, D. (2013). Virtual patient design: exploring what works and why. A grounded theory study. *Medical Education, 47*(6), 595-606.

Georg and Zary (2014) explored students' perceptions of a model for a VP to be used in nursing education that facilitates development of clinical reasoning skills using a posttest-only design. The VP model was developed using the Virtual Interactive Case (VIC) authoring tool (http://pie.med.utoronto.ca/VIC/) and the Outcome-Present State test theoretical framework that "emphasizes reflection, outcome specification, and tests for judgment within the context of the individual patient story" (Georg & Zary, 2014, p. 18). Two VP cases were created and iteratively developed and optimized. After completing the VPs, participants (N=102) volunteered to take a questionnaire that was validated as an instrument to evaluate the design of VPs (see Huwendiek and De Leng, 2010) to express their perceptions of the VP. Scores were analyzed using descriptive and inferential statistics. Georg and Zary (2014) found scores supported the validity of the model and that "students perceived the global linear VPs as a relevant learning activity for the integration of theory and practice" (p. 15). Georg and Zary (2014) argued "an essential characteristic of VPs is the interactive interface" (p. 17) and that VPs "support learning on clinical reasoning and decision making" (p. 17). According to Georg and Zary (2014), VPs are advantageous in that "students may be exposed to a large number of VP cases in a safe and controlled environment" (p. 17), that facilitates "repetitive practice and stand as a model where progressive clinical variation and difficulty can be presented" (p. 17). In short, this study highlighted the importance of the following design features (see Table 1) to enhance learning with VPs: (a) reflection, (b) outcome specification, (c) assessment (e.g., designing tests within the VPs to assess learner's clinical reasoning skills), (d) interactivity, (e) repetitive practice, (f) clinical variation, and (g) different levels of difficulty.

Salminen, Zary, Björklund, Toth-Pal, and Leanderson (2014) investigated students' perceptions of a model for a VP to be used in primary care that facilitates development of clinical reasoning skills using a posttest-only design. The VP model was developed using the OpenTUSK authoring tool (http://opentusk.org) and experiential learning theory (Kolb, 1984), and designed to foster clinical reasoning. The VPs were validated by a panel of 10 experts and pilot-tested at the Karolinska Institutet in Sweden. Three groups of participants (N=14) evaluated the VP followed by a semi-structured interview using a questioning route with each group focusing on the areas of technical design, authenticity, learning process, feedback, and feedback content. The interviews were video recorded, transcribed, and analyzed using *content analysis* and agreed upon by reaching consensus. According to Salminen et al. (2014), “for a student’s ability to apply concepts to solve new problems, active learning with multiple examples can have major effects” (P. 7). “The iterated self-regulated learning cycles, one of the essential parts of the VP model were inspired by Kolb’s learning cycle. For learning to occur, the student has to proceed through all stages of the cycle, which is based on a continuous flow of actions: Doing, Observing, Thinking, and Doing again are repeated. In our learning cycles, the student started out by planning their actions in the current VP section (doing), followed by a concrete experience via multimedia where they were prompted to reflective observation (observing), and after having read the preformulated feedback, wrote down their reflections (thinking)” (Salminen, et al., 2014, p. 7). Salminen et al. (2014) found the VP was well accepted by students and perceived as: (a) authentic, (b) good complement to their theoretical and clinical education, (c) interactive, (d) useful because it provided immediate feedback, (e) self-directed learning tool that promoted their reflective ability.

Posel et al.’s (2014) review of literature provided 12 recommendations for VP design to support the development of clinical reasoning skills. According to Posel et al. (2014), “VPs use authentic, relevant and comprehensive clinical scenarios to actively engage students in problem-solving exercises that emphasize critical analysis, pattern recognition through deliberate practice (Ericsson 2004; Bowen 2006; Bell 2009), require decision-making, hypothesis generation, treatment planning and care management skills (Kamin et al. 2003; Leong et al. 2003; Srinivasan et al. 2007), align decision-making with consequences, and allow for necessary repetition and deliberate practice in a safe and controlled environment (Cook & Triola 2009). They provide necessary immediate, continuous and iterative feedback (Ericsson 2004); as well as access to exemplars of expert practice and rationales (Cook & Triola 2009)” (p. 1). Moreover, “VP cases can integrate distracters such as false or misleading test results, delays and interruptions to further mimic authentic clinical environments. Case complexity can be adjusted to encourage learners to focus on tasks beyond their current levels of competence or comfort without impact on patient safety" (Posel et al., 2014, p. 1).

Based on Posel et al.'s (2014) review of literature, Posel et al. (2014) recommended: (a) creating VP cases that promote clinical variation to include uncommon, longitudinal, with differing levels of complexity, (b) promoting evidence-based decision making, (c) modeling data acquisition and organization, (d) highlighting the importance of summary statements, (e) providing opportunities to practice clinical semantics, (f) focusing on the differential diagnosis, (g) emphasizing the importance of communication in clinical reasoning (h) including continuous and immediate feedback, self- and formative assessment, (i) provide opportunities to learn from errors, (j) using VP cases in other settings, (k) make learners VP case authors, and (l) encouraging post-case reflection.

Finally, Huwendiek et al. (2014) validated an instrument to evaluate the design of virtual patient simulations that focuses on fostering clinical reasoning using three sources of validity evidence: (a) examining content validity using clinical reasoning theory and an international team of VP experts; (b) examining the validity of the response process using think-aloud pilot studies and content analysis of each free-text question associated with each item of the instrument; and (c) examining the internal structure validity using exploratory factor analysis (EFA) and inter-rater reliability by generalizability analysis. Researchers used a large sample of medical students (N=2547) across three European countries, namely Germany, Poland, and the Netherlands to evaluate 78 VPs. Investigators found that the three sources of validity evidence reasonably supported the validity and reliability of the instrument; provided that at least 200 student responses per VP are obtained across all three identified factors (i.e., authenticity of patient encounter and the consultation, cognitive strategies in the consultation, coaching during consultation) are obtained. Findings provided support for prior research concluding the VP design principles that lead to higher learning outcomes are: (a) a VP should be relevant, (b) of an appropriate level of difficulty, (c) highly interactive, (d) offer specific feedback, (e) make optimal use of media, (f) help students focus on relevant learning points, (g) offer recapitulation of key learning points, (h) provide an authentic web-based interface and student tasks, and (i) contain questions and explanations tailored to the clinical reasoning process (Huwendiek et al., 2009a).

Table 1

VP Design Recommendations

Issenberg, et al. (2005)	Cook and Triola (2009)	Huwendiek, et al. (2009)	Cook, et al. (2010)	Botezatu, Hult, Fors (2010)	Botezatu, et al. (2010)	Huwendiek and de Leng (2010)	Adams, et al. (2011)	Gormley, et al. (2011)	Cendan and Lok (2012)	Wilson 2012	Bateman, et al. (2012)	Consorti, et al. (2012)	Bateman, et al. (2013)	Georg and Zary (2014)	Salminem et al. (2014)	Posel et al. (2014)	Huwendiek, et al. (2014)	Total	Percent
Providing feedback	X	X	X	X	X	X			X	X	X		X		X	X	X	**13**	76
Repetitive practice	X		X						X					X		X		**5**	29
Curriculum integration										X								**1**	6
Range of difficulty	X	X				X			X				X	X		X	X	**8**	47

Issenberg, et al. (2005)	Cook and Triola (2009)	Huwendiek, et al. (2009)	Cook, et al. (2010)	Botezatu, Hult, Fors (2010)	Botezatu, et al. (2010)	Huwendiek and de Leng (2010)	Adams, et al. (2011)	Gormley, et al. (2011)	Cendan and Lok (2012)	Wilson 2012	Bateman, et al. (2012)	Consorti, et al. (2012)	Bateman, et al. (2013)	Georg and Zary (2014)	Salminem et al. (2014)	Posel et al. (2014)	Huwendiek, et al. (2014)	Total	Percent
Multiple learning strategies																		**0**	-
Clinical variation	X			X				X	X			X		X	X	X		**8**	47
Controlled environment																X		**1**	6
Individualized learning w/ interactivity		X				X	X	X	X	X		X	X	X	X		X	**11**	65

Issenberg, et al. (2005)	Cook and Triola (2009)	Huwendiek, et al. (2009)	Cook, et al. (2010)	Botezatu, Hult, Fors (2010)	Botezatu, et al. (2010)	Huwendiek and de Leng (2010)	Adams, et al. (2011)	Gormley, et al. (2011)	Cendan and Lok (2012)	Wilson 2012	Bateman, et al. (2012)	Consorti, et al. (2012)	Bateman, et al. (2013)	Georg and Zary (2014)	Salminem et al. (2014)	Posel et al. (2014)	Huwendiek, et al. (2014)	Total	Percent
Learning outcomes		X				X			X				X	X			X	**6**	35
Simulation validity (authenticity)		X		X	X	X	X	X	X	X		X	X		X	X	X	**13**	76
	Clinical Reasoning	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	**17**	100
		Relevant				X				X			X		X	X	X	**7**	41
		Questions & explanations				X							X				X	**4**	24
		Recapitulation				X							X				X	**4**	24

	Issenberg, et al. (2005)	Cook and Triola (2009)	Huwendiek, et al. (2009)	Cook, et al. (2010)	Botezatu, Hult, Fors (2010)	Botezatu, et al. (2010)	Huwendiek and de Leng (2010)	Adams, et al. (2011)	Gormley, et al. (2011)	Cendan and Lok (2012)	Wilson 2012	Bateman, et al. (2012)	Consorti, et al. (2012)	Bateman, et al. (2013)	Georg and Zary (2014)	Salminem et al. (2014)	Posel et al. (2014)	Huwendiek, et al. (2014)	Total	Percent
of key learning points																				
Optimal use of media							X							X				X	**4**	24
Authenticity of student tasks							X							X				X	**4**	24
Advanced organizers																			**1**	6
Contrasting																			**1**	6

	Issenberg, et al. (2005)	Cook and Triola (2009)	Huwendiek, et al. (2009)	Cook, et al. (2010)	Botezatu, Hult, Fors (2010)	Botezatu, et al. (2010)	Huwendiek and de Leng (2010)	Adams, et al. (2011)	Gormley, et al. (2011)	Cendan and Lok (2012)	Wilson 2012	Bateman, et al. (2012)	Consorti, et al. (2012)	Bateman, et al. (2013)	Georg and Zary (2014)	Salminem et al. (2014)	Posel et al. (2014)	Huwendiek, et al. (2014)	Total	Percent
Cases																				
Groups										X									**2**	12
Natural case evolution																			**1**	6
Assessment														X	X		X		**4**	24
Patient Videos																			**1**	6
Collaborative Lrng											X								**2**	12
Reflection											X				X	X	X		**5**	29
Iterative Pedagogical																X			**2**	12

Issenberg, et al. (2005)	Cook and Triola (2009)	Huwendiek, et al. (2009)	Cook, et al. (2010)	Botezatu, Hult, Fors (2010)	Botezatu, et al. (2010)	Huwendiek and de Leng (2010)	Adams, et al. (2011)	Gormley, et al. (2011)	Cendan and Lok (2012)	Wilson 2012	Bateman, et al. (2012)	Consorti, et al. (2012)	Bateman, et al. (2013)	Georg and Zary (2014)	Salminem et al. (2014)	Posel et al. (2014)	Huwendiek, et al. (2014)	Total	Percent
									Loops										
										Branched Casepathways	X		X					**3**	18
										Scoring Counters			X					**2**	12
										Game-informed learning								**1**	6
													Bayesian Reasoning					**1**	6
																Clinical Semantics		**1**	6

Issenberg, et al. (2005)	Cook and Triola (2009)	Huwendiek, et al. (2009)	Cook, et al. (2010)	Botezatu, Hult, Fors (2010)	Botezatu, et al. (2010)	Huwendiek and de Leng (2010)	Adams, et al. (2011)	Gormley, et al. (2011)	Cendan and Lok (2012)	Wilson 2012	Bateman, et al. (2012)	Consorti, et al. (2012)	Bateman, et al. (2013)	Georg and Zary (2014)	Salminem et al. (2014)	Posel et al. (2014)	Huwendiek, et al. (2014)	Total	Percent
																Distractors		**1**	6

VP Design Recommendations

Table one lists 31 VP design recommendations identified in VP design studies between 2005-2014. The following section discusses the most frequently cited design features but the intention is not to question the validity of these recommendations; rather to highlight the heterogeneity of findings regarding VPs design. Because some VP design features appeared to have been less popular than others, it does not mean that they are less important. For instance, specifying and/or measuring *learning outcomes* was explicitly recommended in six out of 17 articles (35%), yet specifying learning outcomes is of paramount importance to the effective design of instruction (Dick, Carey, & Carey, 2009), and *specifying and recapitulating key learning points* are two of the VP design principles that lead to higher learning outcomes (Huwendiek et al., 2009a).

Promoting Clinical Reasoning Expertise

All of the studies reviewed supported the idea that VPs are best used to develop *clinical reasoning skills*. Consorti et al. (2012), for example, concluded, “VPs are effective toward many different specific educational outcomes like clinical reasoning” (p.1007), which supported prior studies findings by Botezatu et al. (2010b), Cook et al. (2010), Cook and Triola (2009), and Huwendiek et al. (2009a), and was further supported by others (Georg & Zary, 2014; Huwendiek et al., 2014; Posel et al., 2014). For instance, Georg and Zary (2014) argued VPs “support learning on clinical reasoning and decision making” (p. 17). Therefore, there appears to be general consensus across the medical education community regarding the design of VPs to primarily develop clinical reasoning skills. Findings suggest that integrating learning theories (e.g., dual process theory) and instructional strategies for the development of clinical reasoning skills into the design of VPs may lead to higher learning outcomes.

Providing Feedback

Thirteen out of 17 studies (76%) underscored the significance of *feedback* as an important feature of VP design. Evidence suggests that “enhanced feedback” is associated with higher learning outcomes (Cook et al., 2010, p. 1599). In addition, feedback should be specific (Huwendiek et al., 2009a), immediate (Salminen et al., 2014) and continuous (Posel et al., 2014). Feedback or coaching during the consultation is also an important factor in instruments recently developed to measure VP design effectiveness (Huwendiek & De Leng, 2010; Huwendiek et al., 2014). Other studies are investigating the effect of providing (or not providing) structured feedback on student learning outcomes (Bateman et al., 2012).

Individualized Learning with Interactivity

Eleven out of 17 studies (65%) supported the notion of designing VPs that *individualize learning with interactivity*. Some argued that VPs should be highly interactive (Huwendiek et al., 2009a), while others found “students may be using the VPs in a formative manner” (Adams et al., 2011, p. 276), which denotes the students’ use of VPs as a tool to promote individualized learning. Consorti et al. (2012) also regarded VPs as interactive, suggesting, “in being challenging for the learner, VPs are believed to foster … active learning” (p. 1005). Additionally, Georg and Zary (2014) contended, “an essential characteristic of VPs is the interactive interface” (p. 17). Moreover, individualized learning also appears to be intricately related to other design features. For instance, Posel et al. (2014), reasoned in favor of providing specific feedback "aligned with individualized responses to key decision-making" (p. 4) and "allowing the students to learn from errors" (p. 4) as design features that would facilitate individualized learning with VPs. Findings suggest learning can be optimized by designing VPs to promote interactive learning.

Realism or Authenticity

Thirteen out of 17 studies (76%) agreed that *realism or authenticity* is a key VP design feature. For instance, Huwendiek et al. (2009a) indicated a VP should "provide an authentic web-based interface and students tasks" (p. 586), while Wilson (2012) concluded that a "VP is an authentic activity" (p. 121), and Salminen et al. (2014) revealed that "students found working with the primary care VP to be active and meaningful with a sense of authenticity" (p. 6). Wilson (2012) also found VPs were perceived as "realistic, relevant, and developed for the local environment. They felt that they were facing people and situations they would see in practice and reacting as they would in the emergency department" (p. 119). According to Botezatu et al. (2010b), "the aspect of paying attention to VPS design, which should enhance clinical reasoning abilities and support case authenticity, cannot be overemphasized" (p. 516). Moreover, authenticity has been associated with knowledge and skills transfer to real patients (Botezatu et al., 2010a). Findings underscore the importance of *authenticity* (e.g., content, language, clinical data, clinical context) as a key characteristic of VP design to enhance VP-mediated learning.

Other Design Features

Other VP design features (Table 1) include using, specifying or providing: (a) learning outcomes, relevant content, well-integrated media, and varying degrees of difficulty (Bateman et al., 2013; Huwendiek & De Leng, 2010; Huwendiek et al., 2009a); (b) clinical variation (Consorti et al., 2012; Cook & Triola, 2009; Gormley et al., 2011); (c) VPs as assessment tools (Botezatu et al., 2010a), particularly for non-analytical reasoning (Cook & Triola, 2009) (d) opportunities for reflection (Wilson, 2012) during case analysis (Georg & Zary, 2014; Salminen et al., 2014), and after case completion (Posel et al., 2014); (e) branched cases (Bateman et al., 2012); (f) repetitive practice (Cook et al., 2010); (g) game attributes like scoring counters (Bateman et al., 2012); (h) collaborative learning and VPs as group activities (Cendan & Lok, 2012); and (i) iterative pedagogical loops (Cendan & Lok, 2012), clinical semantics, and distractors (Posel et al., 2014).

Overall, research indicates that VP simulations are more effective for the teaching and learning of clinical reasoning skills than other scholastic methods but have resulted in a plethora of design principles and recommendations for VP design. For instance, the VP design studies reviewed in this chapter produced a variety of design features (Bateman et al., 2013; Botezatu et al., 2010a; Huwendiek & De Leng, 2010), frameworks (Huwendiek, et al., 2009b), and principles (Huwendiek et al., 2009a), that are similar in nature, but appear to lack consensus and in some instances were derived under considerable research design limitations. Croskerry (2009), for example, argued, “in more than four decades of research, a variety of approaches have been taken, but a consensus approach toward diagnostic decision making has not emerged” (p. 1022). In addition, studies by Huwendiek et al. (2009a), Botezatu et al. (2010b), Botezatu et al. (2010a), and Huwendiek and De Leng (2010), produced a myriad of design recommendations, and were limited to measuring participants’ reactions, which is the first level of Kirkpatrick’s (1975) four-level model of training evaluation. With so many studies producing different recommendations, researchers are not sure which VP design and integration principles to apply to maximize learners' development of clinical reasoning expertise. The lack of uniform VP design principles makes it difficult to create a standardized tool to measure the effectiveness of VP design because without uniform principles, it would be almost impossible to provide a conceptual foundation for the development of an effective standardized tool to measure VP design.

So, what are the features of effective VP design, and how can the effectiveness of VP design be measured? To answer these questions, Huwendiek et al. (2014) validated an instrument to evaluate the design of virtual patient simulations that focuses on fostering clinical reasoning. According to Huwendiek et al. (2014), "… a short and standardized evaluation tool for VP design is a prerequisite for achieving further expansion of VP use in medical education" (p.1). A standardized instrument to measure VP design effectiveness would allow researchers and practitioners to measure the effectiveness of their VP designs and make adjustments accordingly to maximize students' development of clinical reasoning and decision making skills. However, Huwendiek et al. (2014) chose not to consider the relationship of their instrument to other criterion-referenced measures. Since the purpose of this study is to examine the relationship between the students' perceptions of VP design effectiveness, as measured by Huwendiek et al.'s (2014) VP design evaluation instrument, and students' ability to learn and transfer clinical reasoning skills acquired from the VPs, Huwendiek et al.'s (2014) model for the design of the instrument serves as the conceptual framework for this dissertation.

Studies report low acceptance, usage and satisfaction among students when learning resources, such as VPs, are poorly designed or not well integrated with, or offered as an add-on to, other curricular components (Fischer et al., 2007; Haag et al., 2007). Therefore, how the VPs are designed (e.g., what design features are incorporated into the learning activities) and how VPs are integrated into the medical curricula (e.g., what learning activities are carried-out before, during, and after the VPs) may account for more variance in students' learning and satisfaction with VPs than the technology used (Hirumi, 2014; Salas & Gregory, 2011). This realization highlights the importance of Huwendiek et al.'s (2014) VP design evaluation instrument as a tool researchers and practitioners can use to guide the design of VPs that are intended for the development of clinical reasoning expertise. Huwendiek et al.'s (2014) model for the design of Huwendiek et al.'s (2014) VP design evaluation instrument serves as the conceptual framework for this dissertation. Consequently, understanding the model behind the design of Huwendiek et al.'s (2014) VP design evaluation instrument becomes necessary to describe the relationships between the design variables in Huwendiek et al.'s (2014) model and the criterion referenced variables investigated in this study. The following section describes Huwendiek et al.'s (2014) model and illustrates the relationships between the variables in Huwendiek et al.'s (2014) model and the criterion referenced measures examined in this dissertation.

Huwendiek et al. (2014) Model

Huwendiek et al.'s (2014) model for developing the Huwendiek et al.'s (2014) VP design evaluation instrument provides a strong theoretical and conceptual framework for studying the design of VP simulation and answering the research questions posed in this dissertation. Huwendiek et al.'s (2014) model (Figure 2) for developing the Huwendiek et al.'s (2014) VP design evaluation instrument is based on: (a) Gruppen and Frohna's (2002) model for clinical reasoning, (b) Bowen's (2006) teaching strategies, (c) Kim et al.'s (2006) recommendations for building VP cases, and (d) Huwendiek et al.'s (2009a) VP design principles (Table 2). The discussion of Huwendiek et al.'s (2014) theoretical model will be organized based on the three variables or factors of Huwendiek et al.'s (2014) VP design evaluation instrument: (a) authenticity of the patient encounter and the consultation, (b) cognitive strategies in the consultation, and (c) coaching during the consultation.

Table 2

Theoretical model used for Huwendiek et al.'s (2014) VP design evaluation instrument

Models	Strategies	Principles	Description
Gruppen and Frohna (2002)			Students develop clinical reasoning skills by: 1. Evaluating patient situation/characteristics. 2. Considering the context behind clinical symptoms. 3. Gathering information. 4. Developing problem representations. 5. Evaluating problem representations. 6. Comparing to prior knowledge to reach diagnosis or returning to prior steps. Several iterations may be needed before a successful diagnosis can be reached.
	Bowen (2006)		Teaching clinical reasoning by: 1. Asking open-ended questions. 2. Asking for single-sentence summaries of patient problems in abstract terms. 3. Asking for discriminating features of a set of diagnostic hypotheses. 4. Comparing and contrasting diagnostic hypotheses based on real clinical data. 5. Demonstrating typical presentations of different diagnostic hypotheses and the relative probability of different diagnose.
	Kim et al. (2006)		VP teaching cases should be: 1. Relevant. 2. Realistic. 3. Engaging. 4. Challenging. 5. Instructional.
		Huwendiek et al. (2009a)	The VP design principles that lead to higher learning outcomes are that VPs should: 1. Be relevant. 2. Be of an appropriate level of difficulty. 3. Be highly interactive. 4. Offer specific feedback. 5. Make optimal use of media. 6. Help students focus on relevant learning points. 7. Offer recapitulation of key learning points. 8. Provide an authentic web-based interface and student tasks. 9. Contain questions and explanations tailored to the clinical reasoning process.

Factor One - Authenticity of the Patient Encounter and the Consultation

The first variable in Huwendiek et al.'s (2014) theoretical model (Figure 2) underlying Huwendiek et al.'s (2014) VP design evaluation instrument is based on Gruppen and Frohna's (2002) model for clinical reasoning. Gruppen and Frohna (2002) proposed, "a model of clinical reasoning that integrates the key features of several theoretical frameworks" (p. 206). Gruppen and Frohna's (2002) clinical reasoning model has six elements: (a) patient situation and characteristics, (b) context behind clinical symptoms, (c) information gathering, (d) problem representation, (e) evaluation, and (f) prior knowledge (Table 2). According to Gruppen and Frohna (2002), the first element, *patient situation/characteristics*, refers to describing the patient in detail, to include symptoms and physical/mental conditions. The second element, *context*, refers to describing the conditions associated with the patient's ailments and concerns. The third element, *information gathering*, refers to the process of collecting relevant of information to make informed judgments of the patient's condition. The fourth element, *problem representation*, refers to the process of creating mental representations based on prior knowledge and information acquired regarding the patient's condition. These mental representations continue to change as more information about the patient's condition (e.g., blood tests, imaging) becomes available. The fifth element, *evaluation*, refers to the process of making continuous comparisons between existing mental representations of the patient's problems and the actual patient condition. Comparisons between students' mental model and the actual patients' conditions allow students to decide if more information is needed to effectively diagnose the patient's condition. Finally, the sixth element, *prior knowledge*, refers to the process of organizing, storing, and retrieving clinical knowledge, which appears to play a key role in the development of clinical reasoning expertise (Gruppen & Frohna, 2002). Gruppen and Frohna's

(2002) clinical reasoning model is cyclical so several iterations of the process are needed before a successful diagnosis can be reached. The elements of Gruppen and Frohna's (2002) model for clinical reasoning that appear to be most relevant to *establish the authenticity of the patient encounter and the consultation* are: (a) patient situation/characteristics, and (b) context behind clinical symptoms. Application of the evaluation and prior knowledge elements of Gruppen and Frohna's (2002) model is expected to result in students generating detailed descriptions of the patient's symptoms, physical/mental conditions, and the conditions associated with the patient's problems and concerns, which sets the foundation for the next phase in Huwendiek et al.'s (2014) theoretical model for designing Huwendiek et al.'s (2014) VP design evaluation instrument—cognitive strategies in the consultation.

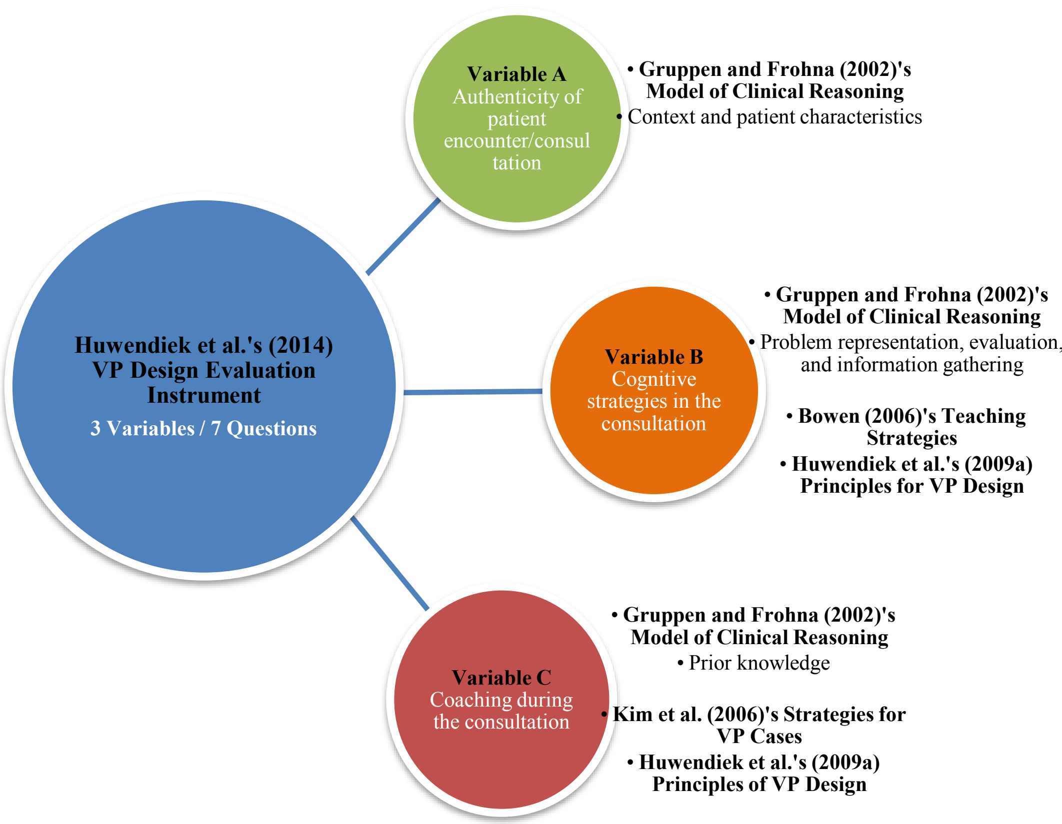


Figure 2: Huwendiek et al.'s (2014) theoretical model for the design of Huwendiek et al.'s (2014) VP design evaluation instrument

Factor Two – Cognitive Strategies in the Consultation

The second variable in Huwendiek et al.'s (2014) theoretical model is based on: (a) Gruppen and Frohna's (2002) model for clinical reasoning, (b) Bowen's (2006) teaching strategies for the development of clinical reasoning expertise, and (c) Huwendiek et al.'s (2009a) VP design principles (Table 2). The VP design principles proposed by Huwendiek et al. (2009a) that appear to be most relevant to facilitate the employment of *cognitive strategies in the consultation* are: (a) relevancy, (b) appropriate level of difficulty, (c) offering specific feedback, (d) helping students focus on relevant learning points, (e) recapitulating key learning points, and (f) providing questions and explanations tailored to the clinical reasoning process. Bowen (2006) posited six educational strategies to promote development of clinical reasoning expertise: (a) asking open-ended questions, (b) asking for single-sentence summaries of patient problems in abstract terms, (c) asking for discriminating features of a set of diagnostic hypotheses, (d) comparing and contrasting diagnostic hypotheses based on real clinical data, and (e) demonstrating typical presentations of different diagnostic hypotheses and the relative probability of different diagnoses. The elements of Bowen's (2006) instructional strategies that appear to be most relevant to enable effective employment of *cognitive strategies in the consultation* are that teachers asks students for: (a) single-sentence summaries of patient problems in abstract terms, (b) the discriminating features of a set of diagnostic hypotheses, (c) comparing and contrasting diagnostic hypotheses, and (d) demonstrating presentations of diagnostic hypotheses and the relative probability of different diagnoses. In terms of outcomes, Gruppen and Frohna's (2002) model appears to be very similar to Bowen's (2006) educational strategies. For instance, Bowen's (2006) recommendations require students to go through steps three to five of Gruppen and Frohna's (2002) model: (a) information gathering, (b) problem representation, and (c) evaluation, resulting in students making continuous comparisons between existing mental representations of the patient's problems and the actual patient condition, and making informed decisions regarding the type of information needed to effectively diagnose the patient's condition. The effective use of *cognitive strategies in the consultation* establishes the foundation for the next phase in Huwendiek et al.'s (2014) theoretical model for designing Huwendiek et al.'s (2014) VP design evaluation instrument—coaching during the consultation.

Factor Three – Coaching During the Consultation

The third variable in Huwendiek et al.'s (2014) theoretical model (Figure 2) underlying Huwendiek et al.'s (2014) VP design evaluation instrument is based on (a) Gruppen and Frohna's (2002) model for clinical reasoning, (b) Kim et al.'s (2006) recommendations for building VP cases, and (c) Huwendiek et al.'s (2009a) VP design principles (Table 2). According to Kim et al. (2006), teaching cases for the development of clinical reasoning skills should be: (a) relevant, (b) realistic, (c) engaging, (d) challenging, and (e) instructional (Table 2). Kim et al.'s (2006) recommendations for developing teaching cases appear to be very similar to Huwendiek et al.'s (2009a) VP design principles. For instance, Kim et al. (2006) recommends cases be *relevant* and *realistic*, which resembles Huwendiek et al.'s (2009a) design principles of *relevancy* and *providing authentic web-based interface and student tasks*. Kim et al. (2006) also recommended that cases be *challenging* and *engaging*, which parallels Huwendiek et al.'s (2009a) VP design principles of *high interactivity* and *appropriate level of difficulty*. Finally, Kim et al. (2006) recommends cases be *instructional*, which compels students to access *prior knowledge* (Gruppen & Frohna, 2002) and resembles Huwendiek et al.'s (2009a) design principles of *making optimal use of media*, *using questions and explanations*, *offering specific* *feedback,* and *focusing* on *and summarizing key learning points*. Altogether, using *authentic, interactive* VP cases that are *relevant*, *realistic*, *challenging*, and *engaging*, while highlighting and summarizing key learning points, offering specific feedback, asking questions, and providing rich explanations induces students to access *prior knowledge* and ascribe meaning to their experiences. Gruppen and Frohna's (2002) model for clinical reasoning is iterative, thus, requiring various iterations to reach successful diagnosis and treatment.

Huwendiek et al.'s (2014) theoretical model provides a theoretical foundation for the design of Huwendiek et al.'s (2014) VP design evaluation instrument and serves as the conceptual framework for this study. The following section illustrates the relationships between the variables in Huwendiek et al.'s (2014) model and the criterion referenced measures examined in this dissertation (Figure 3).

Relationship between Perceptions of VP design and Criterion Variables

The purpose of this dissertation is to examine the relationship between the students' perceptions of VP simulations design effectiveness and students' ability to learn and transfer clinical reasoning skills acquired from the VPs (Figure 3). It is reasoned that variables for informing the design of VP simulations are useful only to the extent to which that they are related to students' subsequent learning and transfer outcomes. This study focuses on studying the students' perceptions of VP design for two reasons: (a) "VP design is essential for the educational success of VPs" (Edelbring et al., 2012; Huwendiek et al., 2009a) as cited by Huwendiek et al. (2014, p. 1); and (b) evidence indicates that "to uncover the educational value of VPs, knowledge about how students make use of them and what qualities students ascribe to them is needed" (Boud & Prosser, 2002; Norman, 2006) as cited by Edelbring et al. (2011, p.

332). If VP design is paramount to the scholastic success of VP simulations and research indicates it is important to study the learners' perceptions of VPs, then it may be useful to examine what is known in the wider literature about the relationship between students' perceptions of VP design (i.e., Huwendiek et al.'s (2014) variables A, B, and C illustrated in Figure 3) and achievement of higher learning and transfer outcomes (Variables D, E, F and G illustrated in Figure 3).

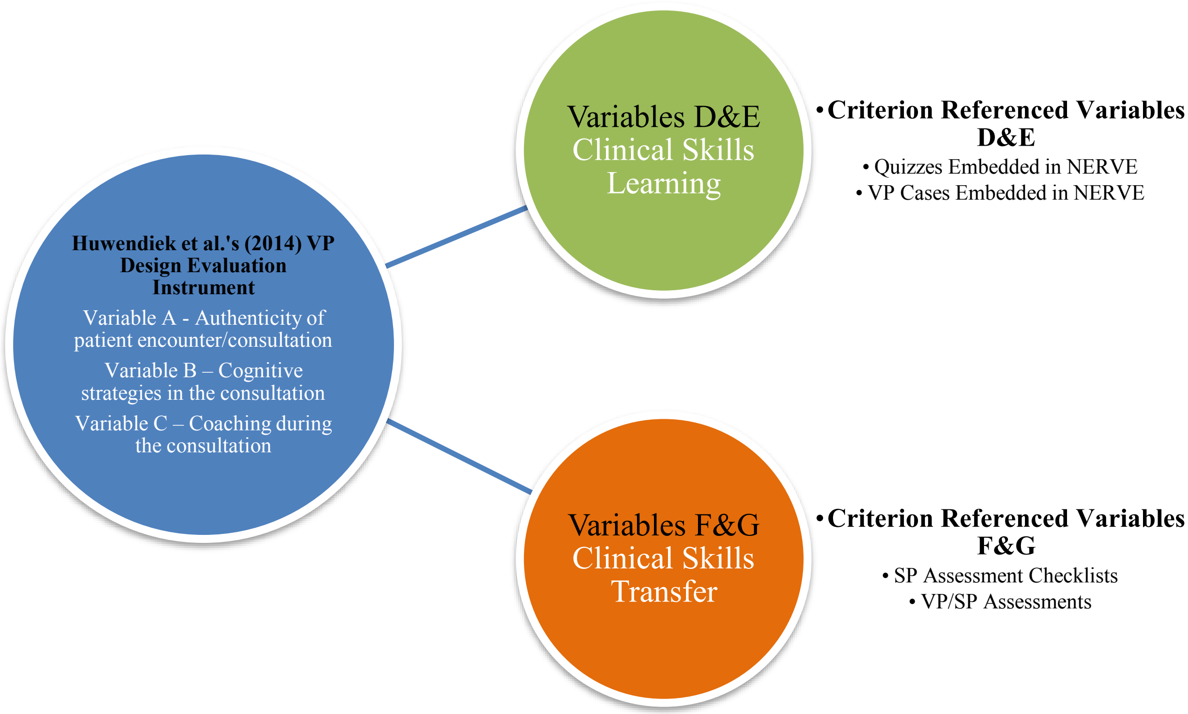


Figure 3: Model illustrating the relationship between students' perceptions of VP design and criterion reference variables.

Authenticity of the Patient Encounter and the Consultation

Research suggests that validity or authenticity (a perceived feature) is one of the key characteristics of high-fidelity medical simulations that lead to effective learning (Issenberg at al., 2005). The authenticity of VP simulations has been associated with knowledge and skills transfer to real patients (Botezatu et al., 2010a). According to Kim et al. (2006), the relevance and realism of VP teaching cases is vital for the development of clinical reasoning expertise. Moreover, according to Huwendiek et al.'s (2009a) VP design principles, VP simulations should "provide an authentic web-based interface and students tasks" (p. 586) to facilitate development of clinical reasoning skills. Finally, the degree of authenticity perceived in medical simulations appears to have a positive effect on patient outcomes, which reflects on the quality of educational outcomes. For instance, according to McGaghie et al. (2011), "a growing body of evidence shows that clinical skills acquired in medical simulation laboratory settings transfer directly to improved patient care practices and better patient outcomes. Examples of improved patient care practices linked directly to SBME include studies of better management of difficult obstetrical deliveries (e.g., shoulder dystocia), laparoscopic surgery, and bronchoscopy. Better patient outcomes linked directly to SBME have been reported in several studies using historical control groups that address reductions in catheter-related bloodstream infections and postpartum outcomes (e.g., brachial palsy injury, neonatal hypoxic–ischemic encephalopathy) among newborn infants" (p. 708).

Cognitive Strategies in the Consultation

Research supports the importance of implementing effective cognitive strategies during the consultation as a way to develop clinical reasoning expertise. For instance, research suggests case-based reasoning and problem solving promotes metacognition and conceptual knowledge (i.e., skills transfer) through deliberate practice with multiple examples (Adkins, 2005; Kolodner, Gray, & Fasse, 2003; Norman, 2005). This is important since working with VP cases is a form of case-based reasoning. Moreover, Norman (2005), who conducted a systematic review of the literature on clinical reasoning published during the previous 30 years found that the development of clinical reasoning expertise requires “deliberate practice with multiple examples and feedback, both to facilitate effective transfer of basic concepts and to ensure an adequate experiential knowledge base” (p. 425). Other approaches include learners using both analogical and analytical approaches for the development of clinical reasoning skills. Croskerry (2009), for example, proposed a universal model for clinical reasoning based pattern recognition and dual-process theory, which is centered on using both analogical and analytical reasoning while Gawronski and Creighton (2013) discussed several dual process theories, including the one “generically described as System 1 and System 2… that are assumed to underlie intuition versus reasoning” (p. 297). Similarly, Eva (2005) combined model for clinical reasoning (Figure 4), illustrates the interplay between the patient symptoms, evolving mental representations of the patient’s real condition, and the hypotheses being tested, requiring learners to manage the clinical reasoning process by employing diverse analogical and analytical cognitive strategies to reach a successful diagnosis. In terms of outcomes, Eva’s (2005) model (Figure 4) appears similar to Gruppen and Frohna’s (2002) model of clinical reasoning (Table 2). Eva’s (2005) model appears to require learners to go through all of the steps in Gruppen and Frohna’s (2002) model: (a) patient situation and characteristics, (b) context behind clinical symptoms, (c) information gathering, (d) problem representation, (e) evaluation, and (f) prior knowledge, resulting in students making continuous comparisons between existing mental representations of the patient's problems and the actual patient condition, and making informed decisions regarding the type of information needed to effectively diagnose the patient's condition.

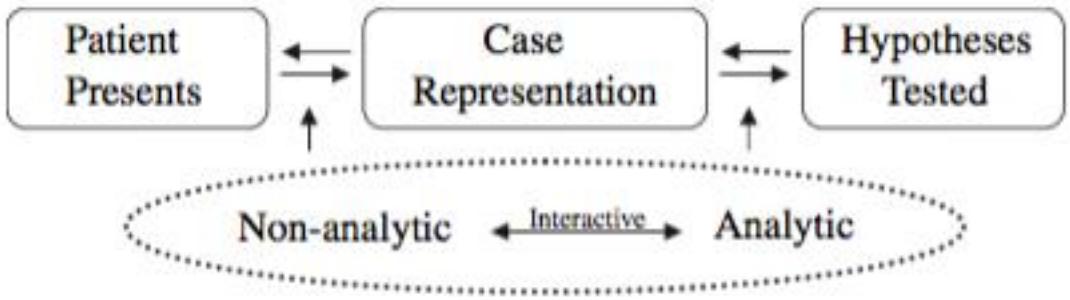


Figure 4: A combined model of clinical reasoning[1]
Adapted from: Eva, K. W. (2005). What every teacher needs to know about clinical reasoning. *Medical Education, 39*(1), 98-106.

Coaching During the Consultation

The importance of coaching during the consultation cannot be understated. Research suggests that the provision of *feedback* is one of the key characteristics of high-fidelity medical simulations that lead to effective learning (Issenberg et al., 2005). Norman (2005) found that *feedback* is required for the development of clinical reasoning expertise. Moreover, *feedback* should be specific, immediate, and continuous (Huwendiek et al., 2009a; Posel et al., 2014; Salminen et al., 2014). Eva (2005)'s teaching strategies for the development of clinical reasoning expertise include directing educators to provide *structure and guidance* as students consider clinical problems. Finally, Eva, Neville, and Norman (1998) recommended educators become active participants, providing *immediate feedback and guidance* through the range of meaningful comparisons that could potentially be drawn.

Altogether, research on VP design supports the relationship between the three variables or factors present in Huwendiek et al.'s (2014) VP design evaluation instrument and the criterion-referenced variables examined in this study (Figure 3). Still, the predictive validity of

Huwendiek et al.'s (2014) VP design evaluation instrument remains to be examined, which is the purpose of this dissertation. Therefore, it is hypothesized that a statistically significant positive correlation between the students' perceptions of NERVE VP design and NERVE-related learning and transfer outcomes will be found.

Summary

Chapter Two discussed the origin and evolution of medical simulations as a way to distinguish VPs from other types of simulations used for facilitating the development of clinical reasoning and decision making skills. Then, this chapter summarized the research on VP design that provides empirical evidence supporting the effectiveness of VPs while highlighting the need to investigate the predictive validity of Huwendiek et al.'s (2014) instrument for VP design evaluation. Subsequently, this chapter described the theoretical framework used to develop Huwendiek et al.'s (2014) VP design evaluation instrument. Finally, Chapter Two presented the model illustrating the relationship between the variables under examination in this dissertation and discussed additional empirical evidence supporting the relationship between the variables in Huwendiek et al.'s (2014) instrument for VP design evaluation and the achievement of higher learning and transfer outcomes. Specifically, the review of literature presented in Chapter 2 suggests that: (a) authenticity of the patient encounter and the consultation, (b) cognitive strategies in the consultation, and (c) coaching during the consultation, are key design features that affect student learning and transfer with VPs. Chapter Three will describe the research design and the intervention, including how the variables or factors in Huwendiek et al.'s (2014) instrument for VP design evaluation, and the variables representing achievement of learning and transfer outcomes, will be measured.

CHAPTER THREE
METHODOLOGY

Setting and Participants

This study was conducted at a medical education institution in Southeastern Florida. Some research activities were carried out inside the main lecture hall (e.g., orientation, consent forms, after action review) while other research activities (i.e., SP/VP interview and differential diagnosis) were conducted inside 12 physical examination rooms at the Clinical Skills and Simulation Center (CSSC).

Sampling

This research study was conducted using a *purposive sample* of second-year medical students (N=120). According to Fraenkel, Wallen, and Hyun (1993), purposive sampling differs from *convenience sampling* in that "researchers do not simply study whoever is available, but use their judgment to select a sample that they believe … will provide the data they need" (p. 114). In other words, purposive sampling uses specific criteria to select specially qualified individuals from a given population while convenient sampling selects individuals that are easily accessible. According to Fraenkel et al. (1993), the main disadvantage of purposive sampling is that "the researcher's judgment may be in error" (p. 114), alluding to the degree by which the sample selected represents the population to which researchers wish to make inferences to. Second-year medical students were chosen for this study because they are at the right educational level to benefit from NERVE compared to other students. First-year students, for example, may not have enough knowledge and clinical experience to successfully complete the VP clinical cases while third-year and beyond medical students may have superior clinical reasoning expertise to benefit from the instructional features of the VPs as much as second-year medical students would. In addition, at the COM, the cranial nerves are studied during the second year of the medical education curriculum so it made more sense to select second-year medical students for this study's sample. Given the specific criteria used to purposively select the sample, this study aimed to make inferences to medical students learning with VPs and in particular, to that specific population of second-year medical school students learning about cranial nerves palsies.

Subjects

The subjects were COM male and female medical students in their second year of the medical school curriculum. The majority of students attended school full time (e.g., credit-hours). It was assumed that students came from different educational and sociocultural backgrounds and were similarly motivated to complete the module as part of the requirements to complete their MD program. According to Gall, Borg, and Gall (1996), when selecting research participants for correlational design studies, "it is important to select a group of participants that is reasonably homogeneous" (p. 338). Despite the assumptions regarding minor individual differences aforementioned, the group of participants in this study was considered to be reasonably homogeneous with regard to the specific criteria (i.e., second-year medical student) used to select this purposive sample.

Instrumentation and Other Measures

Data collection instruments used included a VP design survey, quizzes and VP cases in NERVE, VP/SP performance assessments, and standardized performance checklists.

NERVE VP Suite Educational Tool

NERVE is a suite of VPs and tools designed to teach medical students how to interview, exam, and diagnose patients with cranial nerve disorders funded by the National Institutes of Health (NIH). NERVE contains a set of six virtual patients iteratively developed over the span of five years by a joint research team consisting of medical educators from two universities in Florida, developers from an institution in Georgia, and instructional designers from one university in southeastern Florida using NIH funding. NERVE is currently available online at http://nervesim.com. NERVE has two main areas, a learning center and an exam room (Figure 5).

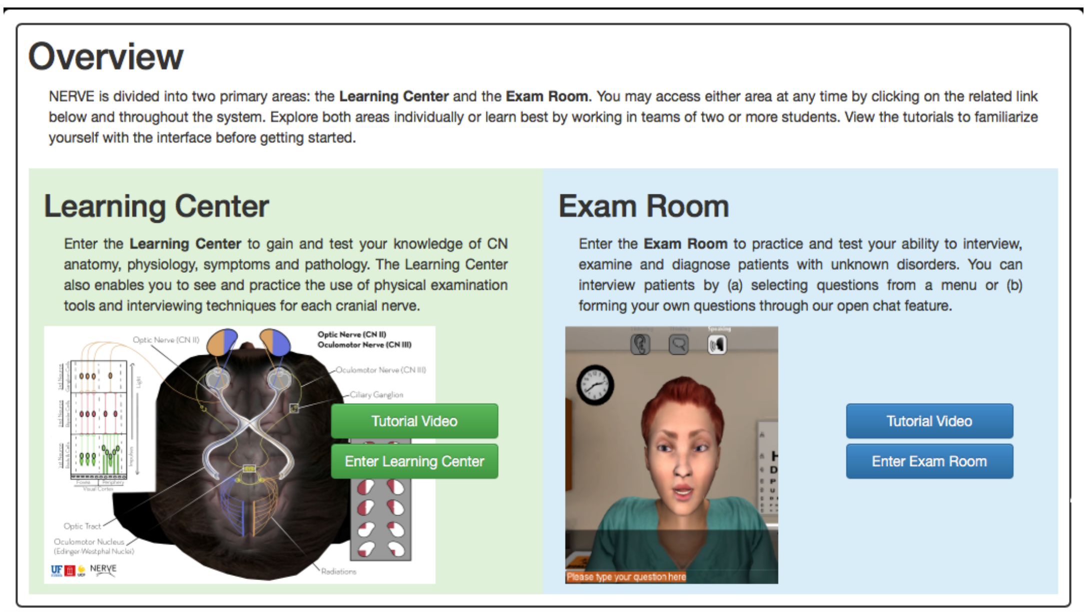


Figure 5: Screen shot illustrating NERVE's two primary areas, the learning center and the exam room

In the learning center students explored information about each of the 12 cranial nerves, to include anatomy, physiology, symptoms, pathology, and published clinical cases (Figure 6).

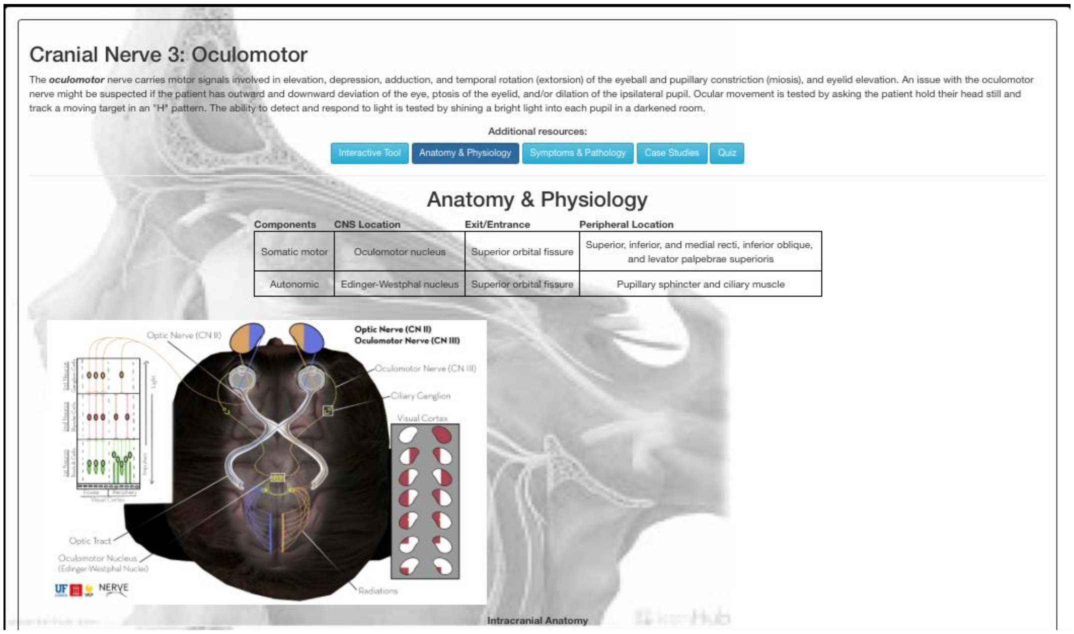


Figure 6: Screen shot illustrating presentation of information on CN 3 anatomy and physiology in NERVE Learning Center

Students also became familiar with the simulated tools (e.g., eye chart, ophthalmoscope, hand tool) used to perform routine patient examinations before proceeding to the exam room (Figure 7).

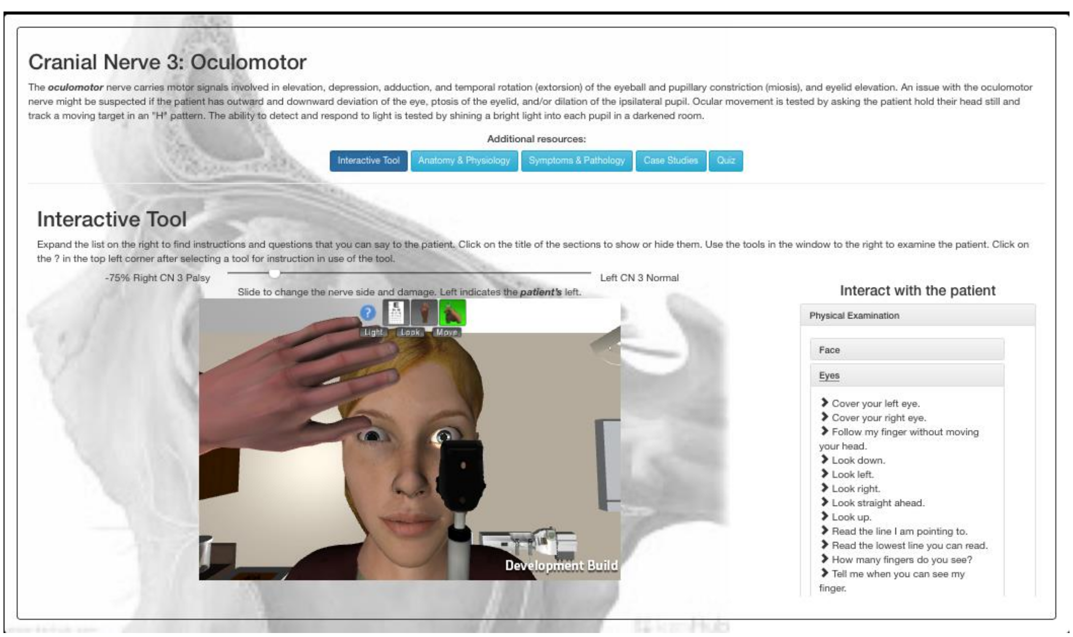


Figure 7: Screen shot illustrating how learners may practice use of examination tools in NERVE Learning Center

In addition, students tested their knowledge of cranial nerve palsies using quizzes embedded into the NERVE Learning Center (Figure 8).

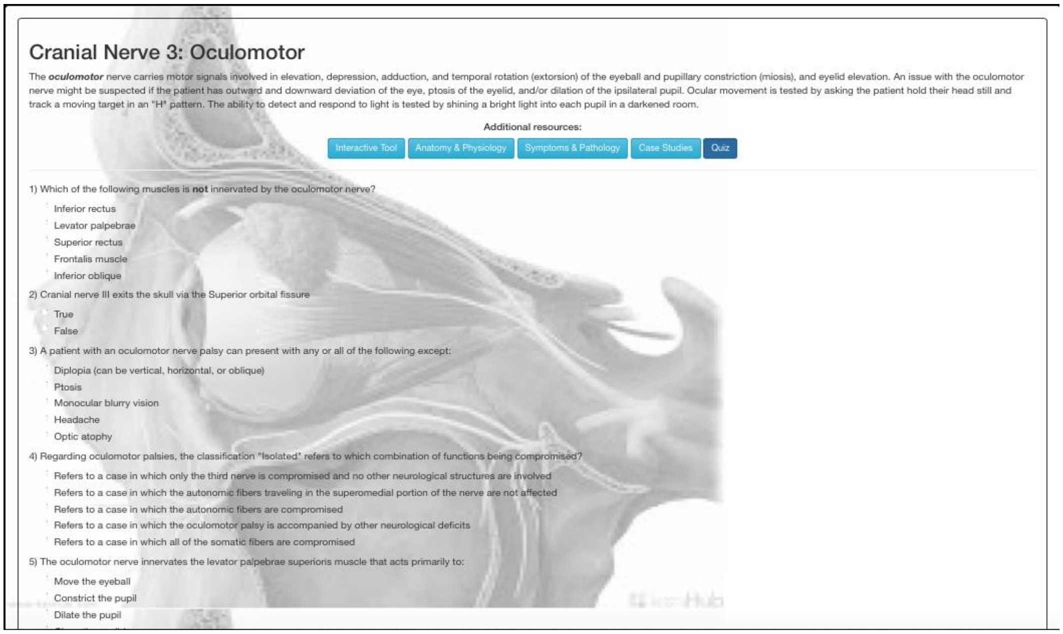


Figure 8: Screen shot illustrating how learners may assess their mastery of CN 3 palsies using embedded quizzes in NERVE Learning Center

In contrast, the exam room offered students the opportunity to practice performing a clinical interview and differential diagnosis of a specific set of VPs, each presenting a different pathology. The exam room had two modalities: (a) *selection NERVE*, which contained drop-down menus with pre-formulated questions to assist novice learners, and (b) *chat NERVE*, which allowed expert learners to type questions as they saw fit (Figure 9).

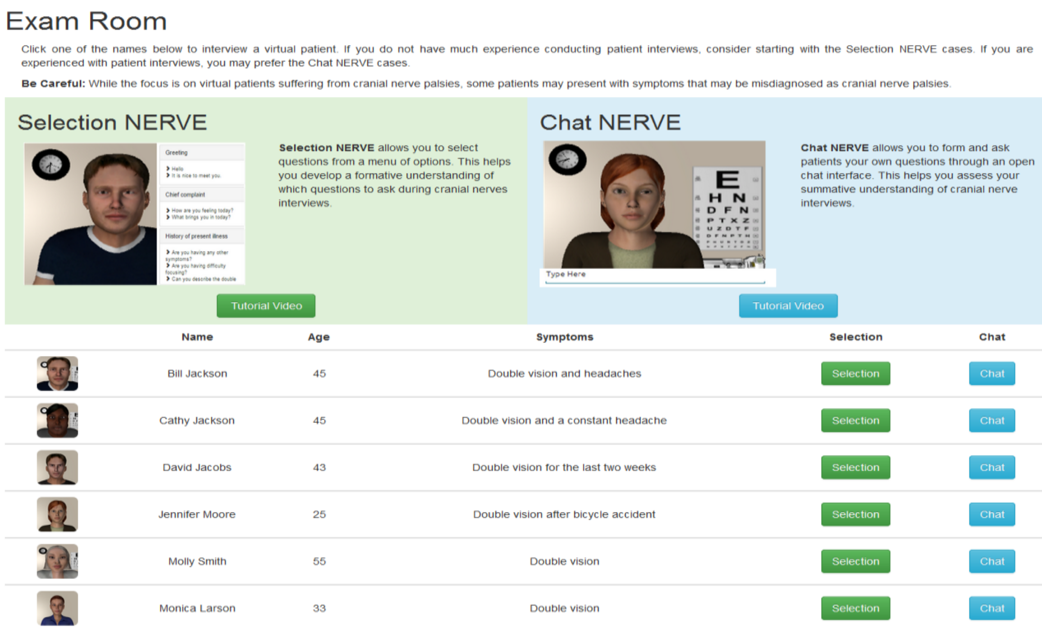


Figure 9: Screen shot illustrating VP cases available in NERVE Exam Room

During the interview, the VPs offered the same verbal and written responses back to learners to provide a heightened sense of realism. Finally, the exam room allowed students to monitor their progress towards a successful differential diagnosis (Figure 10) and provided individualized feedback after cases are completed. Clinical Neurology experts evaluated the NERVE learning center content, quizzes, and exam room VP cases.

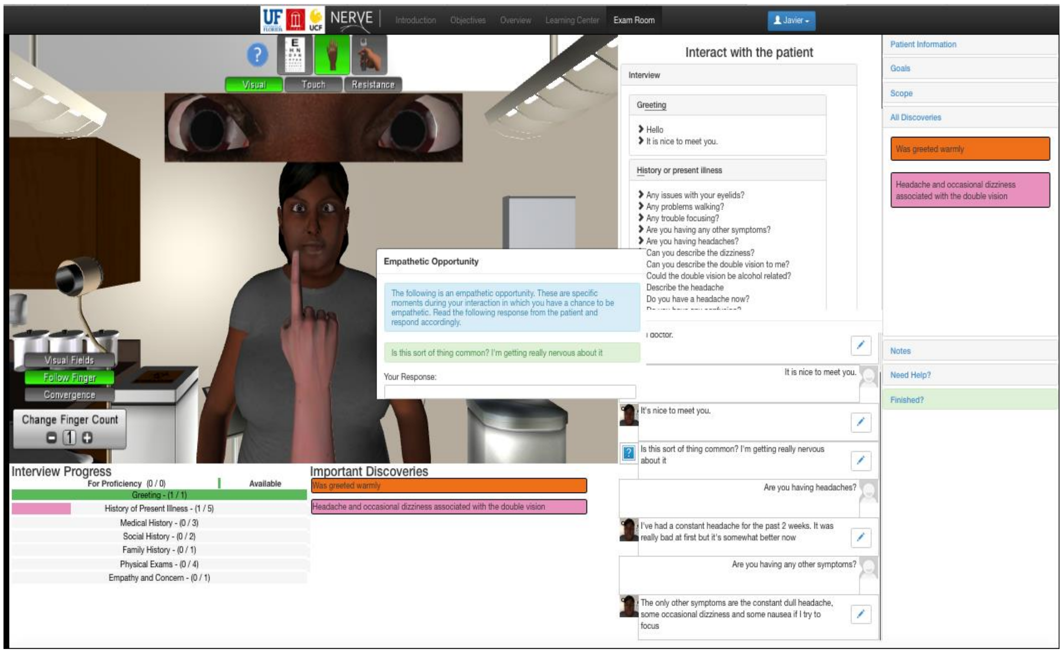


Figure 10: Screen shot illustrating key features of Selection NERVE interface in Exam Room

VP Design Evaluation Survey

Huwendiek et al. (2014) validated an instrument (see Appendix A) to evaluate the design of virtual patient simulations that are focused on fostering clinical reasoning by examining three sources of evidence: (a) content validity using clinical reasoning theory and an international team of VP experts; (b) response process validity using think-aloud pilot studies and content analysis of answers to each free-text closed question associated with each item of the instrument; and (c) internal structure validity using exploratory factor analysis (EFA) and inter-rater reliability by generalizability analysis. Using a large sample of medical students (N=2547) across three European countries (i.e., Germany, Poland, and the Netherlands) to evaluate 78 VPs. Huwendiek et al. (2014) found that the three sources of evidence reasonably supported the validity and reliability of the instrument, with at least 200 student responses per VP across all three identified factors (i.e., authenticity of patient encounter and the consultation, cognitive strategies in the consultation, and coaching during consultation). Huwendiek et al.'s (2014) VPs design evaluation instrument asked respondents to report the degree to which they agreed or disagreed with seven statements on 5-point Likert scales (ranging from strongly agree, agree, neutral, disagree and strongly disagree responses) across three factors (i.e., authenticity of patient encounter and the consultation, cognitive strategies in the consultation, coaching during consultation) and a global score area to evaluate the design of the VP.

Quizzes Embedded in NERVE

There were 12 quizzes (one for each cranial nerve), each contained 10 true-or-false or multiple choice questions about the anatomy, physiology, symptoms, and pathology of cranial nerves to measure students' knowledge of cranial nerve palsies (see Figure 8).

VP Cases in NERVE

There were six different VP cases in NERVE presented in two different formats (closed-menu and open-chat) to measure students' ability to perform a patient interview and differential diagnosis, and to formulate a treatment plan (see Figure 9).

VP/SP Assessments

The VP/SP assessments (see Appendix E) asked students to provide information regarding: (a) localization of the problem, (b) differential diagnosis, (c) evaluation plan, and (d) management & counseling plan, after having experienced a hybrid VP/SP patient encounter.

SP Standardized Performance Checklist

The standardized performance checklist (see Appendix B) was a list of 15 questions assessing specific behaviors students were expected to exhibit while performing the interview and VP/SP examination. Clinical medicine experts from two medical institutions validated this list and SPs were trained on how to use this tool to assess student performance during the examination.

Research Design

A correlational research design will be used to examine the relationship between the students perceptions of VP design and two specific outcome measures: (a) clinical reasoning and decision making skills learned from the VPs, and (b) transfer of clinical reasoning and decision making skills acquired from the VP to other clinical situations. According to Gall et al. (1996), the purpose of correlational research is to search for variables, measured at one point in time, that: (a) predict a criterion variable measured at another point in time, or (b) search for possible causal relationships among variables. Prediction studies, in particular, provide information regarding: (a) the extent to which a criterion behavior pattern can be predicted, (b) data for developing a theory that explains the relationship between the predictors and the criterion behavior pattern, and (c) evidence about the predictive validity of the measures that were correlated to the criterion behavior pattern (Gall et al., 1996, p. 342). Since the present study sought to examine the predictive validity of Huwendiek et al.'s (2014) VP design evaluation instrument, a correlational research design was considered as the best fit.

Procedure and Intervention

All 120 second-year medical students were scheduled to participate in the educational activity on February 2 and either February 3 or 4, 2015, according to their routine schedule for participating in *Practice of Medicine* class sessions. Participation in the educational activity was required of all 120 second-year medical students as part of the planned curriculum in the *Practice of Medicine* module of the COM MD program. As illustrated in Figure 11, a neurosurgery instructor at COM provided a 15-20 minute demonstration of NERVE to all 120 students one week prior to the class session on February 2, 2015, which was the regularly scheduled class on January 26, 2015.

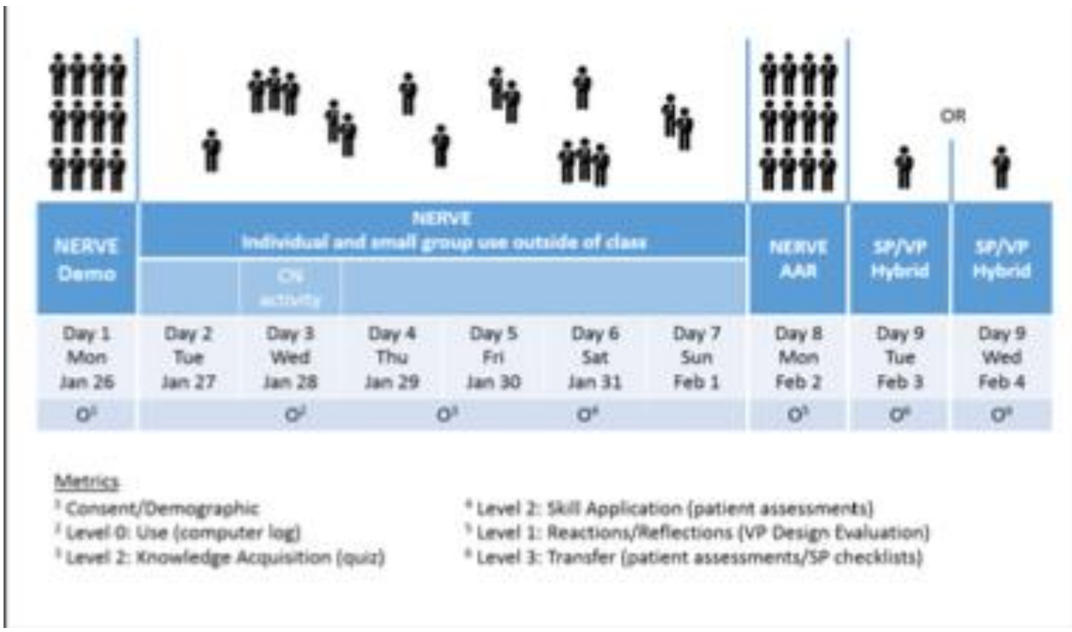


Figure 11: Diagram depicting research and learning activities during the intervention

Students were informed about expectations and what would happen during the class session on February 2, 2015. Another medical instructor informed students about the research component of the scheduled *Practice of Medicine* activities and asked for their voluntary participation on January 26, 2015, following the explanation of the upcoming educational activities (see Figure 11). The research component of the educational activity (i.e., completion of pre-training instruments; use of data for analysis and reporting) was estimated to take less than 5 minutes of the two-hour educational session. The instructor described the option to participate in the research component of the activity and ensured each student received a paper copy of the

Explanation of Research Form (Appendix F) reviewed during this session. This study did not include any vulnerable populations. Additionally, to minimize any possibility of coercion, it was made clear to students that participation in the study was optional, and that they were still required to participate in the course activities regardless of their decision to contribute data to the study. If a participant wanted to stop participating in the study, the student was allowed to terminate participation at any time without penalty. Participants were not compensated and participation in this study was not a component of the module grade. As part of the course activities (see Figure 11), all students were given access to the pre-training instruments (see Appendices C, D, and E) in Qualtrics (a web-based survey software tool), on their COM-issued laptops via an email link provided to them on January 26, 2015. The first page of the instrument asked students to indicate whether or not they consented to participate in the study (Appendix C). Students who provided their consent were directed to another webpage to complete the demographic portion of the pre-training instrument (Appendix D) and students who did not consent were directed to a thank-you page. However, documented consent is not required for exempt research. Though the consent of the participants was not documented in writing, the explanation of research form (Appendix F) was distributed informing students of the option to participate. Students were also given access to NERVE using their school network IDs (NIDs) to participate in the simulation component of the course activity from 26 Jan 2015 to 2 Feb 2015. Students were allowed to spend as much time as they wanted exploring NERVE whenever they wanted, individually or in teams, based on their preferences, prior to the class session on February 2, 2015 (see Figure 11). Data during the interaction on the student's performance was recorded within the NERVE system. This data included the topics covered by the students, the type and number of discoveries the students made, the patient history findings, the physical findings, diagnosis, and treatment.

To answer the first research question, students were asked to interact with NERVE individually and/or in small groups on their own time for the week leading up to and including February 2, 2015, to include completing the required quizzes and required VP cases (see Figure 11).

Then, during the class session on February 2, 2015, instructors facilitated an after-action review on: (a) what students learned about CNs and CN disorders, (b) what questions the students may have regarding CNs and the use of the NERVE system, and (c) what other CN information the students should know and learn. In particular, the After-Action Review (AAR) centered on answering four questions: (a) what actually happened? (eight items that students answered individually), and three items that students responded to in teams of six: (b) what did you learn? (c) what went well and why? (d) what can be improved and how? The instructor began the session by reviewing the goals, features and ground rules for the AAR and then asked students to complete Huwendiek et al.'s (2014) VP design evaluation questionnaire using a link that was emailed to the students during the session and to respond electronically to the four questions using a series of prompts. After each question, students were given several minutes to formulate their response and then submit their response online. Responses to the last three items were displayed anonymously on a giant screen for everyone to review, including the facilitator, whose purpose at this juncture was to react to the responses and ask follow-up questions as appropriate. All responses were compiled and recorded for further analysis after the AAR.

To answer the second research question, 60 students participated in the NERVE VP/SP interview and differential diagnosis educational activity of an SP with a CN6 injury for approximately one hour from 8am to 12pm on February 3, 2015, while 60 other students participated in the NERVE VP/SP interview and differential diagnosis educational activity of an SP with a CN6 injury for approximately one hour from 8am to 12pm on February 4, 2015, as pre-determined by the curriculum schedule (see Figure 11). A hybrid approach (e.g., real patient and virtual patient) was used in which the students performed patient interview on the SP and the physical examination on a VP through the physical examination version of NERVE. The physical exam portion of the standardized patient interview took place on a laptop using the physical exam version of NERVE. All students were given access to the post-encounter note or patient assessment (see Appendix E) in Qualtrics on a desktop computer in the clinical skills area at COM.

Inclusion and Exclusion Criteria

Only second-year medical students who agreed to participate in the study were included in the final study sample and data was recorded and analyzed for these students. Any student who did not consent to participate in the study was excluded (i.e., data was not analyzed) even though they were still required to participate in the educational activity as part of the planned curriculum in the *Practice of Medicine* module of the COM M.D. program. One student that was part of the NERVE research team was also excluded. Study participants were at least 18 years old.

Data Collection

According to Gall et al. (1996), data collection methods used in correlational studies include: (a) self-report measures, (b) questionnaires, (c) standardized tests, (d) interviews, and (e) observational techniques. This dissertation used the following instruments: (a) Huwendiek et al.'s (2014) VP design evaluation questionnaire (interval/ratio data), (b) quizzes in NERVE (interval/ratio data), (c) VP cases in NERVE (nominal data [correct/incorrect]), (d) VP/SP performance assessments (nominal data [correct/incorrect]), and (e) standardized performance checklists (nominal data [behavior observed/not observed]).

The students' perceptions of VPs design effectiveness as it relates to development of clinical reasoning and decision making skills were measured by using Huwendiek et al.'s (2014) VP design evaluation instrument. Student outcomes assessed include: (a) students' acquisition of skills and knowledge (i.e., learning) from the system as measured by embedded quizzes and required VP cases in NERVE, and (b) students' ability to transfer skills and knowledge acquired from NERVE as measured by VP/SP assessments (i.e., differential diagnosis and treatment), and a standardized patient checklist.

Data Management and Confidentiality

All data (i.e., consent responses, demographic responses, interaction data and survey responses) was accessed and managed by research team personnel only. Data relating to each virtual patient interaction (e.g., topics covered by the students, number and type of discoveries made, physical findings) was captured and recorded in the NERVE system by individual student login names (i.e., each student's NID). Research team personnel who had access to NERVE logs and internal data was able view all data by these NIDs but did not have the list that matches students' NIDs to student names. Only one research team member had access to the list that joins NIDs and student names so for other research team members, the NIDs simply served as unique identifiers and allowed controlled access to NERVE. Controlling access to the data that linked

NIDs to students' names was important to maintain the participants' privacy and confidentiality. Student responses on the pre- and post-training instruments were automatically collected and stored in a password-protected environment by Qualtrics (i.e., in a password-protected Qualtrics account), except for the team responses to the three team items during AAR, which were captured and stored in another research team member password-protected Qualtrics account to allow the facilitator and students to view that information in real-time during the 2 Feb 2015 session. Responses on the instruments that required individual students input were initially linked to individual student electronic mail account addresses. A research team member managed and accessed these Qualtrics data, exported all responses from Qualtrics and stored the exported file in a password-protected folder on a secure COM server. To protect participant confidentiality, a research team member assembled one complete data file in Excel/SPSS wherein each student's responses were matched across measures and randomly assigned a unique number from 1-120 to each student and removed all student names and electronic mail account addresses from the file. The de-identified data set file was stored in a password-protected folder on a secure COM server. From this point forward, research team personnel were only able to view and use this de-identified data set. Only individuals directly involved in study management had access to this de-identified data set. While the research team member that had access to the identifiable data set had no involvement in student assessment or grading for the *Practice of Medicine* module, these safeguards ensured that research team members who were involved in grading were not able to view student performance by name or electronic mail account address. All data will be maintained on a secure COM server for five years in compliance with the APA. There were no foreseeable risks to participants in this experiment, thus no medical care and personal injury compensation was needed. There were no costs and no direct benefits to students associated with participation in this trial. There were no drugs or devices used in this experiment. All research activities related to this dissertation were conducted at COM only and the results of the current study will be made available to participants upon request once data analysis has been completed.

Data Analysis

Data was analyzed using descriptive statistics and confirmatory factor analysis (CFA) and results were reported accordingly in Chapter Four. Likert-type rating scale data was treated as interval-level data. Since assumptions of statistical tests were met (e.g., normality, homogeneity of variance), parametric approaches were selected for data analysis. Descriptive data (e.g., means and frequencies) was reported for demographic items. Quantitative data was analyzed using SPSS 22.0 (IBM; Chicago, IL) and SAS 9.4 (SAS Institute; Cary, NC).

Several outcome variables were eliminated from the analysis due to: (a) very low or zero variance, and (b) poor fit with the specified model. The outcome variables eliminated from the analysis were: CN 3, CN 4, CN 5, CN Side, CN Left, CN Right, CN Bilateral, CN Right, SPQ3, SPQ5, SPQ6, and SPQ9. Finally, the predictor variable QS_31 (question number seven in Huwendiek et al.'s (2014) instrument) was also excluded because this question is not part of the three factors specified for the instrument: (a) authenticity of the patient encounter and the consultation, (b) cognitive strategies in the consultation, and (c) coaching during the consultation.

CHAPTER FOUR
RESULTS

Four different sets of analyses were conducted to evaluate the hypotheses and research questions posited in this study. The first set of analyses examined the characteristics of the sample used in this dissertation study to include participant demographics, student consent, exclusion criteria and measures of normality such as skewness and kurtosis. The second set of analyses investigated the validity and reliability of Huwendiek et al.'s (2014) VP design evaluation instrument as it relates to the scores obtained from the intervention. The third and fourth sets of analyses addressed the hypotheses and research questions postulated in this study. The third set of analyses assessed the correlation between NERVE VPs design effectiveness and higher student clinical reasoning learning outcomes, while the fourth set of analyses examined the correlation between the effectiveness of NERVE VPs design and higher student clinical reasoning skills transfer outcomes. Normal interval/ratio continuous variables are displayed as mean and standard deviation (SD) with 95% confidence interval (CI). Categorical variables are presented as frequency and percentage while ordinal and non-normal continuous variables are reported as median and interquartile (IQR) range and/or minimum-maximum range. Instrument reliability was evaluated as internal consistency using the Cronbach's Alpha coefficient. Statistical analyses were conducted using SPSS 23.0 (IBM; Armonk, NY) and SAS University Edition (SAS Institute; Cary, NC).

Sample Characteristics

This section examines the sample characteristics using measures of central tendency, percentages, and frequency counts regarding participants' self-reported age, gender, and race, as well as students' consent for research. According to Gall et al. (1996), when selecting research participants for correlational design studies, "it is important to select a group of participants that is reasonably homogeneous" (p. 338). Thus, investigating the homogeneity of the sample is necessary to confirm the appropriateness of the correlational design selected for this dissertation study.

Participant Demographics and Student Consent

One hundred eighteen of 120 second-year medical students (98.3%) consented to the use of their data for analysis and reporting purposes (Table 3).

Table 3

Consent of Participants

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	I consent to participate in the research component of the educational activity.	118	100.0	100.0	100.0

Participants (Table 4) included 58 males (49.2%) and 60 females (50.8%), median age (Table 6) of 24 years old (IQR = 23-25.25; range = 22-38), representing various ethnic groups

(Table 6)—80 White or Caucasian (67.8%), 34 Asian (28.8%), 7 Hispanic or Latino (5.9%), and 5 Black or African-American (4.2%).

Table 4

Gender of participants

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	58	49.2	49.2	49.2
	Female	60	50.8	50.8	100.0
	Total	118	100.0	100.0	

Table 5

Age of participants

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	22	3	2.5	2.5	2.5
	23	29	24.6	24.6	27.1
	24	36	30.5	30.5	57.6
	25	21	17.8	17.8	75.4
	26	13	11.0	11.0	86.4
	27	5	4.2	4.2	90.7
	28	3	2.5	2.5	93.2
	30	3	2.5	2.5	95.8
	31	1	.8	.8	96.6
	34	2	1.7	1.7	98.3
	36	1	.8	.8	99.2
	38	1	.8	.8	100.0
	Total	118	100.0	100.0	

Despite the 16-year range difference between participants (Table 6), 107 (91%) of participants in the sample are between the ages of 22 and 27 while over two-thirds of participants

(67.8%) reported their race as white (Table 6), which altogether supports the notion of *reasonable homogeneity* of participants in the sample as posited by Gall et al. (1996).

Table 6

Sample statistics

		Student Consent	Gender	Age	American Indian or Alaska Native	Asian	Black or African-American	Hispanic or Latino	Native Hawaiian or Other Pacific Islander	White
N	Valid	118	118	118	0	34	5	7	0	80
	Missing	0	0	0	118	84	113	111	118	38
Mean		1.00	.51	24.93		1.00	1.00	1.00		1.00
Std. Error of Mean		.000	.046	.241		.000	.000	.000		.000
Median		1.00	1.00	24.00		1.00	1.00	1.00		1.00
Mode		1	1	24		1	1	1		1
Std. Deviation		.000	.502	2.617		.000	.000	.000		.000
Variance		.000	.252	6.850		.000	.000	.000		.000
Std. Error of Skewness		.223	.223	.223		.403	.913	.794		.269
Std. Error of Kurtosis		.442	.442	.442		.788	2.000	1.587		.532
Range		0	1	16		0	0	0		0
Minimum		1	0	22		1	1	1		1
Maximum		1	1	38		1	1	1		1
Percentiles	25	1.00	.00	23.00		1.00	1.00	1.00		1.00
	50	1.00	1.00	24.00		1.00	1.00	1.00		1.00
	75	1.00	1.00	25.25		1.00	1.00	1.00		1.00
Skewness			-.034	2.679						
Kurtosis			-2.034	8.819						

Data Skewness and Kurtosis

Data regarding participants' self-reported age, gender, and race was found to have some levels of skewness, and kurtosis. According to Sematech (2006), "skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the center point" (p. 261). Moreover, the "skewness for a normal distribution is zero, and any symmetric data should have a skewness near zero. Negative values for the skewness indicate data that are skewed left and positive values for the skewness indicate data that are skewed right" (Sematech, 2006, p. 261). The skewness of the data related to the participants' gender in the sample was -0.34 (Table 6) suggesting a near-zero skewness while slightly skewed to the left to reflect the marginally higher amount of females 60 (50.8%) versus males 58 (49.2%) but still consistent with the characteristics of a normally distributed sample. However, gender variables are categorical in nature and are thus reported as frequencies and percentages. In contrast, the skewness of the data related to the participants' age in the sample was 2.679 (Table 6 & Figure 12) suggesting a moderately skewed to the right data set reflecting the preponderant amount of participants whose reported age (Table 5) was at the median value of 24 years old (IQR = 23-25.25; range = 22-38) or higher (Table 6). The degree of skewness of the data related to the participants' age in the sample is not consistent with the characteristics of a normally distributed sample and are thus considered non-normal continuous variables. The data related to the participants' age is thus reported as median and interquartile (IQR) range and/or minimum-maximum range.

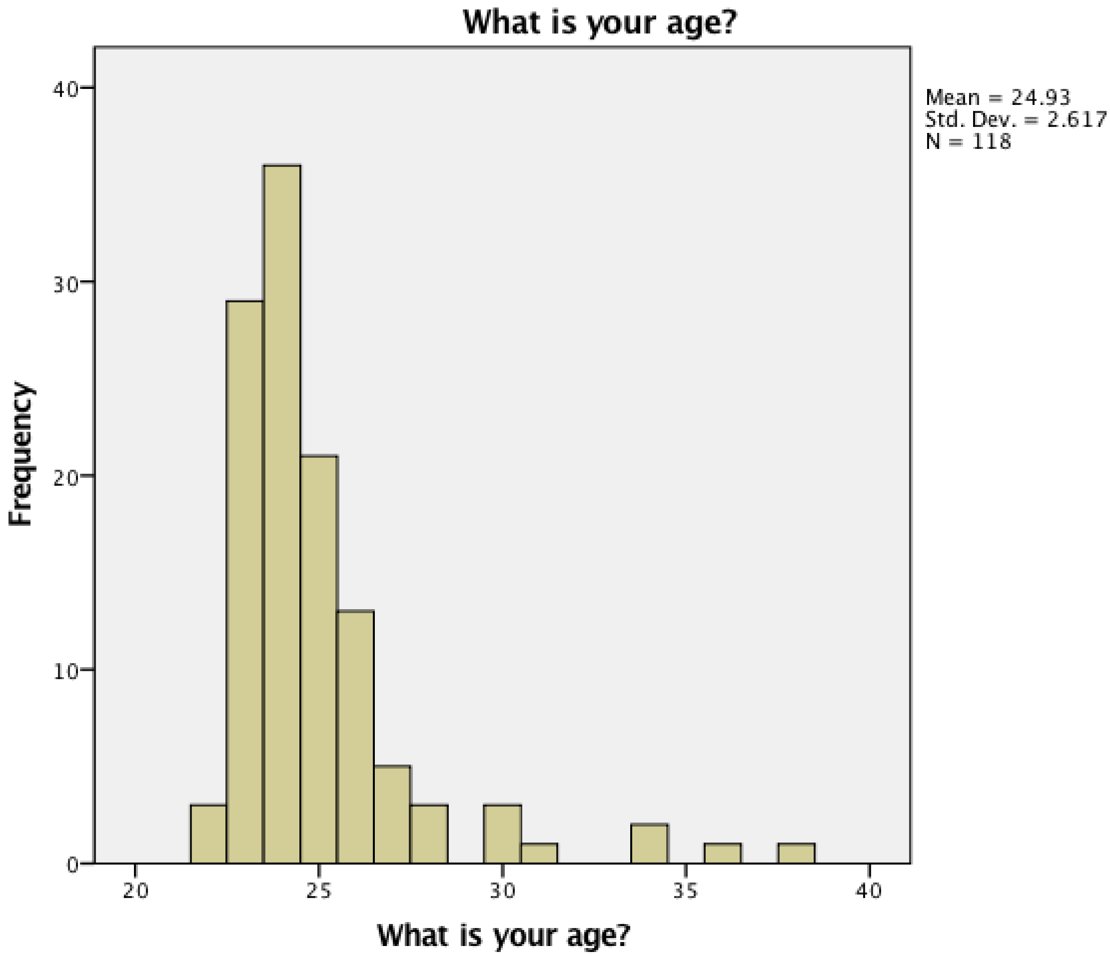


Figure 12: Histogram depicting the age of participants

According to Sematech (2006), "kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. That is, data sets with high kurtosis tend to have a distinct peak near the mean, decline rather rapidly, and have heavy tails. Data sets with low kurtosis tend to have a flat top near the mean rather than a sharp peak" (p. 261). Moreover, "the kurtosis for a standard normal distribution is three … positive kurtosis indicates a "peaked" distribution and negative kurtosis indicates a "flat" distribution" Sematech (2006, p. 262). The kurtosis of the data related to the participants' gender in the sample was -2.034 (Table 6) suggesting low negative kurtosis with a flat distribution as suggested by the 1.6% differential between the amount of females 60 (50.8%) and males 58 (49.2%) in the sample (Table 4) and illustrated in Figure 13.

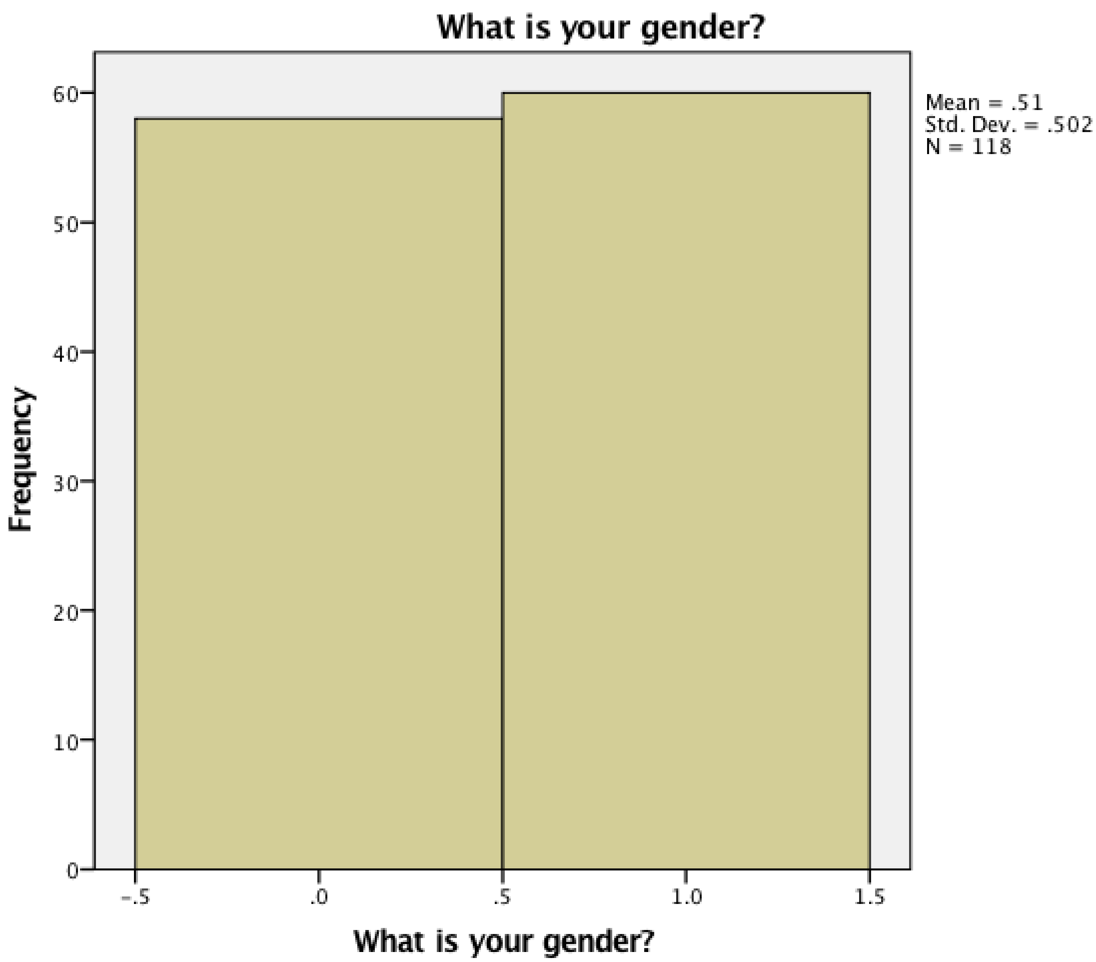


Figure 13: Histogram depicting the gender of participants

Even though the data related to the participants' gender in the sample appears to be consistent with the characteristics of a normally distributed sample, gender variables are categorical and associated values are thus reported as frequencies and percentages. In contrast, the kurtosis of the data related to the participants' age in the sample was 8.819 (Table 6) suggesting high positive kurtosis with a distinct peak distribution that declines rapidly and has a heavy tail as defined by Sematech (2006) and illustrated in Figure 12. The degree of kurtosis of the data related to the participants' age in the sample is not consistent with the characteristics of a normally distributed sample and are thus considered non-normal continuous variables. The data related to the participants' age is thus reported as median and interquartile (IQR) range and/or minimum-maximum range.

Altogether, participant self-reported demographic data was found to be reasonably homogeneous in terms of age as 91% of participants were between the ages of 22 and 27 (Figure 12) while participants' race and gender data was also found to be reasonably homogeneous as over two-thirds of participants (67.8%) reported their race as white (Table 6) and gender differences were less than two percent (Figure 13), which altogether supports the notion of *reasonable homogeneity* of participants in the sample for correlational design studies as posited by Gall et al. (1996).

Data Cleaning

Data cleaning is critical step in preserving accuracy throughout the data analysis and reporting process as it "deals with detecting and removing errors and inconsistencies from data in order to improve the quality of data. Data quality problems are present in single data collections, such as files and databases, e.g., due to misspellings during data entry, missing information or other invalid data" (Rahm & Do, 2000, p. 1). The establishment of exclusion criteria further facilitates the data cleaning process. The exclusion criteria for participant data in this dissertation study included: (a) students who did not provide consent for the use of their data for research purposes, (b) students who consented to the use of their data but did not provide, or researchers were not able to obtain from, data for all measures (i.e., surveys, quizzes, VP cases, VP/SP assessment, and SP checklists) used in this study, and (c) students who were members of the

NERVE research team. Therefore, from an original sample of 120 students, data from two participants was excluded because they did not provide consent for the use of their data for research purposes. From the remaining sample of students who provided consent (n=118), data from eight participants was excluded because they did not provide data for all measures or because the participant was a member of the NERVE research team, resulting in the final sample (n=110) used for analyses described in the rest of Chapter 4.

Validity and Reliability of Huwendiek et al.'s (2014) VP Design Evaluation Instrument

According to Messick (1995), "validity is an overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions on the basis of test scores or other modes of assessment. Validity is not a property of the test or assessment as such, but rather of the meaning of the test scores. These scores are a function not only of the items or stimulus conditions, but also of the persons responding as well as the context of the assessment. In particular, what needs to be valid is the meaning or interpretation of the score; as well as any implications for action that this meaning entails" (p. 741). This section examines the validity and reliability of the scores obtained from administering Huwendiek et al.'s (2014) VP design evaluation instrument to the sample of second-year medical students after interacting with NERVE. First, reliability is assessed as internal consistency based on the Cronbach's Alpha coefficient. Then, construct validity is examined using confirmatory factor analysis. Measures of validity and reliability will assist in ascribing meaning to the scores obtained from Huwendiek et al.'s (2014) VP design evaluation instrument as a reflection of the purpose for which it was designed, the constructs it is intended to measure, the participants, and the context in which Huwendiek et al.'s (2014) VP design evaluation instrument was administered.

Reliability of Huwendiek et al.'s (2014) VP Design Evaluation Instrument

According to Cronbach (1951), "any research based on measurement must be concerned with the accuracy or dependability or, as we usually call it, reliability of measurement. A reliability coefficient demonstrates whether the test designer was correct in expecting a certain collection of items to yield interpretable statements about individual differences" (p. 297). The reliability of Huwendiek et al.'s (2014) VP design evaluation instrument was assessed using the scores obtained from the sample of participants (n=110) on the first six questions of the instrument (Appendix A), since question seven is a global score question that is not part of the three design factors evaluated by Huwendiek et al.'s (2014) VP design evaluation instrument and the results are presented below.

Table 7

Valid cases used for estimating the Cronbach's Alpha

		N	%
Cases	Valid	110	93.2
	Excluded[a]	8	6.8
	Total	118	100.0

Note. [a] Listwise deletion based on all variables in the procedure.

A sample of participants (n=110) was used for estimating the Cronbach's Alpha reliability coefficient (Table 7). Respondent ratings of effectiveness of NERVE VP design obtained from Huwendiek et al.'s (2014) VP design evaluation instrument were considered to be highly reliable for the sample of second-year medical school students to whom it was given (Table 8), which yielded a Cronbach's Alpha reliability coefficient of .87 (N of items = 6).

According to Nunally (1978), “the minimum recommended instrument reliability is .70” (p. 245) so a reliability coefficient of .87 is considered to be synonymous to very good reliability. In addition, the internal consistency (.87) obtained with this dissertation’s sample (n=110) appears to be similar to the one obtained in the original study (n=2,547), which ranged from .74 to .82 for each factor in the instrument (Huwendiek at al., 2014, p. 5). Results suggest that similar results would be obtained if Huwendiek et al.’s (2014) instrument were administered to a different sample of second-year medical students after completing similar learning activities in NERVE. In addition, a .87 reliability coefficient also support the assertions made in the preceding section regarding sample homogeneity since scores obtained from heterogeneous groups within the same sample would be expected to yield a low (< .70) reliability coefficient. According to Cronbach (1951), “Alpha estimates, and is a lower bound to, the proportion of test variance attributable to common factors among the items” (p. 331). Therefore, results suggest that at least 87% of the variance in students’ perceptions of NERVE VP design can be attributed to the design factors or variables evaluated by Huwendiek et al.’s (2014) VP design evaluation instrument.

Table 8

Reliability statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.872	.871	6

Table 9

Item-total statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Q2_25	16.25	14.297	.742	.675	.837
Q2_26	16.32	14.035	.766	.694	.833
Q2_27	16.02	14.825	.705	.605	.844
Q2_28	15.89	15.786	.652	.538	.854
Q2_29	15.74	16.471	.595	.402	.863
Q2_30	16.15	15.691	.582	.405	.866

Further, inspection of the corrected item-total correlations (Table 9) revealed strong item-total correlations and that no items have a zero or negative corrected item-total correlations. Moreover, the Cronbach's Alpha decreases if any item is removed from the scale (Table 9) so the scale has maximized parsimony and no items will be removed.

Reliability estimates are essential to assess the validity of psychometric measures as "no validity coefficient and no factor analysis can be interpreted without some appropriate estimate of the magnitude of the error of measurement" (Cronbach, 1951, p. 297). Therefore, the item reliability estimates discussed in this section provide the foundation for the next section, evaluating the validity of Huwendiek et al.'s (2014) VP design evaluation instrument based on the sample and intervention using CFA.

Validity of Huwendiek et al.'s (2014) VP Design Evaluation Instrument

The purpose of CFA is to confirm if a set of latent (i.e., hidden, unobservable) factors underlies the data as hypothesized based on theoretically grounded descriptions of the relationship between the factors and observable data (i.e., manifest variables). CFA differs from exploratory factor analysis (EFA) in that EFA (used in Huwendiek et al.'s (2014) original study) tries to discover these relationships while CFA seeks to confirm the presence and strength of these known or assumed relationships. The assumptions for CFA are: (a) the existence of a theory-grounded model, (b) relatively large sample size, and (c) using interval continuous variables. CFA can be used to investigate the relationship between theory-specified factors and manifest variables to include the estimation of measurement and latent errors and the degree of fit of the proposed model using fit indices. For the purposes of this dissertation study, the following fit indices will be interpreted: (a) Standardized Root Mean Square Residual (SRMR), (b) Root Mean Squared Error of Approximation (RMSEA), and (c) Bentler Comparative Fit Index (Bentler CFI). The SRMR represents the standardized difference between the observed correlation and the predicted correlation and a value of zero represents a perfect fit. The RMSEA is a measure of the degree to which the proposed model does not fit the empirical data it is being tested against and a value of zero represents a perfect fit. The Bentler CFI examines inconsistencies between the proposed model and the data representing how much better the proposed model fits the data compared to the null model and its values ranges from zero (lack of fit) to one (perfect fit). According to Hu and Bentler (1999), "for the ML method, a cutoff value close to .95 for CFI, … a cutoff value close to .08 for SRMR; and a cutoff value close to .06 for RMSEA are needed before we can conclude that there is a relatively good fit between the hypothesized model and the observed data" (p.1). However, some of these indices and in particular, the RMSEA, are sensitive to sample size as "the cut-off value … actually *decreases* for incorrect models as sample size increases. This may suggest that power calculations are more likely to be optimal when based on those indices" (Sivo, Fan, Witta, & Willse, 2006, p. 267). Moreover, their findings suggest that researchers may be better able to distinguish between correct and incorrect models when making judgments based on the RMSEA.

The purpose of this analysis was to confirm that the latent factors described in the theoretical model for the design of Huwendiek et al.'s (2014) VP design evaluation instrument (Figure 2) indeed underlie the scores obtained from the instrument. Since Huwendiek et al.'s (2014) VP design evaluation instrument was validated using EFA and CFA assumptions were met, a CFA procedure was conducted using maximum likelihood (ML) rotation, which is a type of orthogonal rotational procedure that assumes factors are not correlated. The maximum likelihood procedure properly converged in four iterations (Figure 14) yielding the following fit indices: (a) SRMR=.00; (b) RMSEA=.00; and (c) Bentler CFI=1.00 (Figure 15).

Optimization Results			
Iterations	4	Function Calls	13
Jacobian Calls	6	Active Constraints	0
Objective Function	0.0115058966	Max Abs Gradient Element	9.5553103E-6
Lambda	0	Actual Over Pred Change	1.1095811199
Radius	0.0001896609		

Convergence criterion (ABSGCONV=0.00001) satisfied.

Figure 14. Optimization Results

Results indicate a perfect fit between Huwendiek et al.'s (2014) VP design evaluation instrument and the data as it was originally conceptualized (Figure 15). Moreover, inspection of the Lambda standardized coefficients (Figure 16), revealed Lambdas t values were all > 1.96 (critical value), therefore, the Lambdas, which represent the relationship between the three factors in Huwendiek et al.'s (2014) instrument and the manifest variables (i.e., scores obtained from participants for questions 26-30) were all significant ($p<.0001$). Results confirm the presence of three VP design factors (i.e., authenticity of patient encounter and consultation, cognitive strategies in the consultation, and coaching during consultation) underlying the participant data obtained from the intervention. Since these three VP design factors were derived from Huwendiek et al.’s (2014) theoretical framework for the instrument based on Gruppen and Frohna’s (2002) model of clinical reasoning, Bowen’s (2006) teaching strategies, Kim et al.’s (2006) strategies for building VP cases, and Huwendiek et al.’s (2009a) VP design principles, the findings also provide support for the relationships between the variables espoused by Huwendiek et al.’s (2014) theoretical framework as well as for the construct validity of Huwendiek et al.’s (2014) instrument. However, validity goes beyond the measure to include the participants and the context in which it was administered. Therefore, assertions made in this dissertation about the validity of Huwendiek et al.’s (2014) instrument will be confined to these parameters. Consequently, Huwendiek et al.’s (2014) VP design effectiveness evaluation instrument was found to be a valid and reliable tool to measure the effectiveness of NERVE VP design using second-year medical students and the learning objectives and educational activities described in the intervention (see Chapter 3). Since Huwendiek et al.’s (2014) instrument was found to be a valid measure; the students’ perceptions of the effectiveness of NERVE VP design (i.e., scores obtained from the instrument) are expected to reflect the degree of effectiveness of the NERVE VP suite in promoting the development of clinical reasoning and decision making skills.

Students' Perceptions of NERVE VP Design as measured by Huwendiek et al.'s (2014) Three-Factor VP Design Evaluation Instrument

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

Fit Summary		
Modeling Info	Number of Observations	110
	Number of Variables	6
	Number of Moments	21
	Number of Parameters	15
	Number of Active Constraints	0
	Baseline Model Function Value	3.2107
	Baseline Model Chi-Square	349.9634
	Baseline Model Chi-Square DF	15
	Pr > Baseline Model Chi-Square	<.0001
Absolute Index	Fit Function	0.0115
	Chi-Square	1.2541
	Chi-Square DF	6
	Pr > Chi-Square	0.9741
	Z-Test of Wilson & Hilferty	-1.9199
	Hoelter Critical N	1095
	Root Mean Square Residual (RMR)	0.0100
	Standardized RMR (SRMR)	0.0097
	Goodness of Fit Index (GFI)	0.9962
Parsimony Index	Adjusted GFI (AGFI)	0.9868
	Parsimonious GFI	0.3985
	RMSEA Estimate	0.0000
	RMSEA Lower 90% Confidence Limit	0.0000
	RMSEA Upper 90% Confidence Limit	0.0000
	Probability of Close Fit	0.9871
	ECVI Estimate	0.3056
	ECVI Lower 90% Confidence Limit	0.3529
	ECVI Upper 90% Confidence Limit	0.3529
	Akaike Information Criterion	31.2541
	Bozdogan CAIC	86.7613
	Schwarz Bayesian Criterion	71.7613
	McDonald Centrality	1.0218
Incremental Index	Bentler Comparative Fit Index	1.0000

Figure 15. Fit indices for Huwendiek et al.'s (2014) VP design evaluation instrument

Standardized Effects in Linear Equations						
Variable	Predictor	Parameter	Estimate	Standard Error	t Value	Pr > |t|
Q2_25	F1	L1	0.88383	0.03476	25.4281	<.0001
Q2_26	F1	L2	0.89270	0.03410	26.1813	<.0001
Q2_27	F2	L3	0.90245	0.03950	22.8489	<.0001
Q2_28	F2	L4	0.78896	0.04752	16.6015	<.0001
Q2_29	F3	L5	0.77578	0.06507	11.9220	<.0001
Q2_30	F3	L6	0.72572	0.06699	10.8328	<.0001

Figure 16. Standardized effects in linear equations

The current study aims to examine the predictive validity of Huwendiek et al.'s (2014) VP design effectiveness measurement instrument by determining if the design factors evaluated by Huwendiek et al.'s (2014) instrument are correlated to criterion referenced measures such as the students' ability to learn from the NERVE VPs and to transfer clinical reasoning skills learned from the NERVE VPs to other clinical situations. The next section presents results regarding the analysis of the relationship between the participants' perceptions of NERVE VP design and the students' ability to learn from the NERVE VPs.

Correlation between NERVE VPs Design Effectiveness and Higher Student Clinical Reasoning Learning Outcomes

The correlation between the students' perceptions of effectiveness of NERVE VPs design and the achievement of higher clinical reasoning learning outcomes was measured using CFA to explore the relationship between the design factors evaluated by Huwendiek et al.'s (2014) VP design effectiveness measurement instrument and student performance in both quizzes and required VP cases embedded in NERVE. The following section presents results regarding the analysis of the relationship between the participants' perceptions of NERVE VP design and student performance in quizzes embedded NERVE. It was hypothesized (the specified model) that students' perceptions of effectiveness of NERVE VPs design would be correlated to the achievement of higher student clinical reasoning learning outcomes in NERVE.

Correlation between NERVE VPs Design Effectiveness and Student Performance in Quizzes Embedded in NERVE

The maximum likelihood procedure properly converged in 10 iterations (Figure 17) yielding the following fit indices: (a) SRMR=.02; (b) RMSEA=.06; and (c) Bentler CFI=.99

(Figure 18). Results indicate a very good fit for the specified model suggesting the existence of a relationship between the students' perceptions of NERVE VP design effectiveness and student performance in quizzes embedded in NERVE, as originally conceptualized (Figure 18). However, further inspection of the standardized coefficients (Lambdas or Beta Weights) revealed Lambdas t values were not all > 1.96 (critical value), therefore, the Lambdas (Figure 19), which represent the relationship between the three factors in Huwendiek et al.'s (2014) instrument and the manifest variables (i.e., scores obtained from participants in quizzes embedded in NERVE) were not all significant at the $p \leq .05$ level.

Optimization Results			
Iterations	10	Function Calls	25
Jacobian Calls	12	Active Constraints	0
Objective Function	0.0275451134	Max Abs Gradient Element	0.0000221585
Lambda	1.110223E-14	Actual Over Pred Change	1.146872716
Radius	1482075.9469		

Convergence criterion (GCONV=1E-8) satisfied.

Figure 17. Optimization Results

Relationship between Students' Perceptions of NERVE VP Design and Students' Clinical Reasoning Skills Learning as measured by Quizzes Embedded in NERVE

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

Fit Summary		
Modeling Info	Number of Observations	110
	Number of Variables	5
	Number of Moments	15
	Number of Parameters	13
	Number of Active Constraints	0
	Baseline Model Function Value	1.9581
	Baseline Model Chi-Square	213.4301
	Baseline Model Chi-Square DF	10
	Pr > Baseline Model Chi-Square	<.0001
Absolute Index	Fit Function	0.0275
	Chi-Square	3.0024
	Chi-Square DF	2
	Pr > Chi-Square	0.2229
	Z-Test of Wilson & Hilferty	0.7684
	Hoelter Critical N	218
	Root Mean Square Residual (RMR)	0.0040
	Standardized RMR (SRMR)	0.0237
	Goodness of Fit Index (GFI)	0.9897
Parsimony Index	Adjusted GFI (AGFI)	0.9225
	Parsimonious GFI	0.1979
	RMSEA Estimate	0.0678
	RMSEA Lower 90% Confidence Limit	0.0000
	RMSEA Upper 90% Confidence Limit	0.2140
	Probability of Close Fit	0.3108
	ECVI Estimate	0.2800
	ECVI Lower 90% Confidence Limit	0.2718
	ECVI Upper 90% Confidence Limit	0.3655
	Akaike Information Criterion	29.0024
	Bozdogan CAIC	77.1087
	Schwarz Bayesian Criterion	64.1087
	McDonald Centrality	0.9955
Incremental Index	Bentler Comparative Fit Index	0.9951

Figure 18. Fit indices for the correlation between NERVE VP design effectiveness and students' performance in quizzes embedded in NERVE

Standardized Effects in Linear Equations						
Variable	Predictor	Parameter	Estimate	Standard Error	t Value	Pr > |t|
Quizzes_3	F2	L1	0.82128	0.37011	2.2190	0.0265
Quizzes_4	F1	L2	0.26640	0.16619	1.6030	0.1089
Quizzes_5	F1	L3	0.29279	0.17739	1.6506	0.0988
Quizzes_7	F3	L4	0.84098	0.03824	21.9897	<.0001
Quizzes_10	F3	L5	0.83950	0.03839	21.8692	<.0001

Figure 19. Standardized effects in linear equations

The next section presents results regarding the analysis of the relationship between the participants' perceptions of NERVE VP design and student performance in VP cases embedded NERVE. It was hypothesized (the specified model) that students' perceptions of effectiveness of NERVE VPs design would be correlated to the achievement of higher student clinical reasoning learning outcomes in NERVE.

Correlation between NERVE VPs Design Effectiveness and Student Performance in VP Cases Embedded in NERVE

The maximum likelihood procedure properly converged in six iterations (Figure 20) yielding the following fit indices: (a) SRMR=.01; (b) RMSEA=.00; and (c) Bentler CFI=1.00 (Figure 21). Results indicate a very good fit for the specified model suggesting the existence of a relationship between the students' perceptions of NERVE VP design effectiveness and student performance in VP cases embedded in NERVE, as originally conceptualized (Figure 21). Moreover, further inspection of the standardized coefficients (Lambdas or Beta Weights) revealed Lambdas t values were all > 1.96 (critical value), therefore, the Lambdas (Figure 22), which represent the relationship between the three factors in Huwendiek et al.'s (2014) instrument and the manifest variables (i.e., scores obtained from participants in VP cases embedded in NERVE) were all significant (p<.0001).

Optimization Results			
Iterations	6	Function Calls	24
Jacobian Calls	8	Active Constraints	0
Objective Function	0.0065754885	Max Abs Gradient Element	8.7087551E-6
Lambda	0	Actual Over Pred Change	0.961583256
Radius	0.0511208842		

Convergence criterion (ABSGCONV=0.00001) satisfied.

Figure 20. Optimization Results

Relationship between Students' Perceptions of NERVE VP Design and Student Clinical Reasoning Skills Learning as measured by VP Cases Embedded in NERVE

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

Fit Summary		
Modeling Info	Number of Observations	110
	Number of Variables	5
	Number of Moments	15
	Number of Parameters	13
	Number of Active Constraints	0
	Baseline Model Function Value	1.3145
	Baseline Model Chi-Square	143.2803
	Baseline Model Chi-Square DF	10
	Pr > Baseline Model Chi-Square	<.0001
Absolute Index	Fit Function	0.0066
	Chi-Square	0.7167
	Chi-Square DF	2
	Pr > Chi-Square	0.6988
	Z-Test of Wilson & Hilferty	-0.5358
	Hoelter Critical N	912
	Root Mean Square Residual (RMR)	3.3374
	Standardized RMR (SRMR)	0.0149
	Goodness of Fit Index (GFI)	0.9974
Parsimony Index	Adjusted GFI (AGFI)	0.9802
	Parsimonious GFI	0.1995
	RMSEA Estimate	0.0000
	RMSEA Lower 90% Confidence Limit	0.0000
	RMSEA Upper 90% Confidence Limit	0.1400
	Probability of Close Fit	0.7600
	ECVI Estimate	0.2590
	ECVI Lower 90% Confidence Limit	0.2718
	ECVI Upper 90% Confidence Limit	0.3119
	Akaike Information Criterion	26.7167
	Bozdogan CAIC	74.8230
	Schwarz Bayesian Criterion	61.8230
	McDonald Centrality	1.0059
Incremental Index	Bentler Comparative Fit Index	1.0000

Figure 21. Fit indices for the correlation between NERVE VP design effectiveness and students' performance in VP cases embedded in NERVE

Standardized Effects in Linear Equations						
Variable	Predictor	Parameter	Estimate	Standard Error	t Value	Pr > |t|
cn_3_take_1	F1	L1	0.37403	0.09403	3.9777	<.0001
cn_4_take_1	F1	L2	0.79138	0.10238	7.7299	<.0001
cn_5_take_1	F2	L3	0.51477	0.08972	5.7372	<.0001
cn_7_take_1	F2	L4	0.61776	0.08924	6.9223	<.0001
cn_10_take_1	F3	L5	1.01894	0.00257	396.5	<.0001

Figure 22. Standardized effects in linear equations

Students were also required to apply the skills learned while interacting with NERVE to other clinical situations. The following section presents results regarding the analysis of the relationship between the participants' perceptions of NERVE VP design and the achievement of higher students' clinical reasoning skills transfer outcomes.

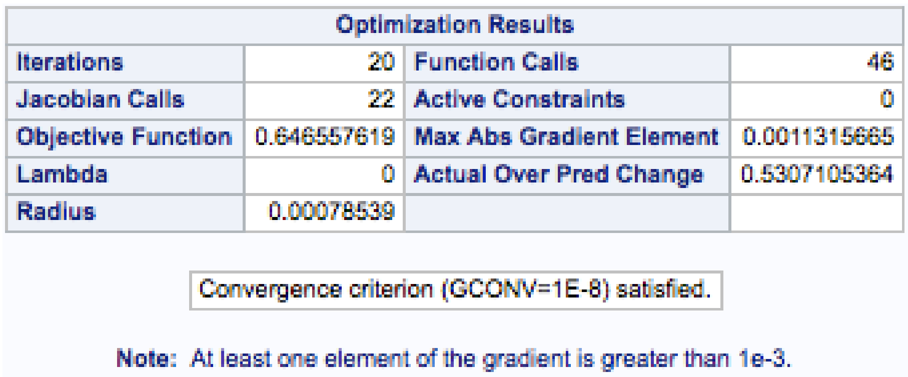
Correlation between NERVE VPs Design Effectiveness and Higher Student Clinical Reasoning Skills Transfer Outcomes

The correlation between the students' perceptions of effectiveness of NERVE VPs design and the achievement of higher clinical reasoning skills transfer outcomes was measured using CFA to explore the relationship between the design factors evaluated by Huwendiek et al.'s (2014) VP design effectiveness measurement instrument and student performance in both NERVE-assisted VP/SP assessment and SP checklists. The following section presents results regarding the analysis of the relationship between the participants' perceptions of NERVE VP design and student performance in the NERVE-assisted VP/SP assessment and SP checklists. It was hypothesized (the specified model) that students' perceptions of effectiveness of NERVE VPs design would be correlated to the achievement of higher student clinical reasoning skills transfer outcomes in NERVE.

Correlation between NERVE VPs Design Effectiveness and Student Performance in VP/SP Differential Diagnosis and SP Checklists

The maximum likelihood procedure properly converged in 31 iterations (Figure 23) yielding the following fit indices: (a) SRMR=.07; (b) RMSEA=.03; and (c) Bentler CFI=.96 (Figure 24). Results indicate a good fit for the specified model suggesting the existence of a relationship between the students' perceptions of NERVE VP design effectiveness and student performance in both NERVE-assisted VP/SP assessment and SP checklists, as originally conceptualized (Figure 24). However, further inspection of the standardized coefficients

(Lambdas or Beta Weights) revealed Lambdas t values were not all > 1.96 (critical value), therefore, the Lambdas (Figure 25), which represent the relationship between the three factors in Huwendiek et al.'s (2014) instrument and the manifest variables (i.e., scores obtained from participants in NERVE-assisted VP/SP assessments and SP checklists) were not all significant at the $p \leq .05$ level. All other relationships were not found to be significant.



Optimization Results			
Iterations	20	Function Calls	46
Jacobian Calls	22	Active Constraints	0
Objective Function	0.646557619	Max Abs Gradient Element	0.0011315665
Lambda	0	Actual Over Pred Change	0.5307105364
Radius	0.00078539		

Convergence criterion (GCONV=1E-8) satisfied.

Note: At least one element of the gradient is greater than 1e-3.

Figure 23. Optimization Results

Relationship between Students' Perceptions of NERVE VP Design and Students' Clinical Reasoning Skills Transfer as measured by NERVE-assisted VP/SP Differential Diagnosis and SP Checklists

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

Fit Summary		
Modeling Info	Number of Observations	110
	Number of Variables	13
	Number of Moments	91
	Number of Parameters	29
	Number of Active Constraints	0
	Baseline Model Function Value	2.8556
	Baseline Model Chi-Square	311.2613
	Baseline Model Chi-Square DF	78
	Pr > Baseline Model Chi-Square	<.0001
Absolute Index	Fit Function	0.6466
	Chi-Square	70.4748
	Chi-Square DF	62
	Pr > Chi-Square	0.2153
	Z-Test of Wilson & Hilferty	0.7887
	Hoelter Critical N	126
	Root Mean Square Residual (RMR)	0.0066
	Standardized RMR (SRMR)	0.0759
	Goodness of Fit Index (GFI)	0.9144
Parsimony Index	Adjusted GFI (AGFI)	0.8744
	Parsimonious GFI	0.7269
	RMSEA Estimate	0.0354
	RMSEA Lower 90% Confidence Limit	0.0000
	RMSEA Upper 90% Confidence Limit	0.0703
	Probability of Close Fit	0.7167
	ECVI Estimate	1.2571
	ECVI Lower 90% Confidence Limit	1.2632
	ECVI Upper 90% Confidence Limit	1.4937
	Akaike Information Criterion	128.4748
	Bozdogan CAIC	235.7887
	Schwarz Bayesian Criterion	206.7887
	McDonald Centrality	0.9622
Incremental Index	Bentler Comparative Fit Index	0.9637

Figure 24. Fit indices for the correlation between NERVE VP design effectiveness and students' performance in NERVE-assisted VP/SP assessments and SP checklists.

Standardized Effects in Linear Equations						
Variable	Predictor	Parameter	Estimate	Standard Error	t Value	Pr > |t|
CNType_Recoded	F2	L1	-0.09941	0.10543	-0.9429	0.3457
CN_Diagnosis_Recoded	F3	L2	0.14349	0.08758	1.6384	0.1013
SPQ1	F1	L3	-0.03126	0.07384	-0.4234	0.6720
SPQ2	F1	L4	-0.05735	0.07478	-0.7669	0.4431
SPQ4	F2	L5	0.14018	0.10453	1.3410	0.1799
SPQ7	F1	L6	0.70159	0.11128	6.3050	<.0001
SPQ8	F3	L7	0.55327	0.07269	7.6119	<.0001
SPQ10	F3	L8	0.20953	0.08631	2.4276	0.0152
SPQ11	F1	L9	0.26580	0.09580	2.7746	0.0055
SPQ12	F3	L10	0.71812	0.06137	11.7009	<.0001
SPQ13	F2	L11	0.80195	0.06707	11.9561	<.0001
SPQ14	F2	L12	0.36498	0.09421	3.8740	0.0001
SPQ15	F2	L13	0.81369	0.06684	12.1734	<.0001

Figure 25. Standardized effects in linear equations

The results presented in the previous sections of this chapter support the validity and reliability of Huwendiek et al.'s (2014) VP design evaluation instrument to measure the design effectiveness of NERVE. Since the design effectiveness of NERVE is related to the students' scores on Huwendiek et al.'s (2014) instrument, the following section presents the results of the analysis of students' scores on the instrument.

Students' Perceptions of NERVE VP Design

One hundred ten of 118 students (93.0%) completed Huwendiek et al.'s (2014) VP design effectiveness evaluation instrument (Table 10). Participant scores ranged from 1 to 5. Mean score and SD for participants' scores on Huwendiek et al.'s (2014) total (aggregated) scale was 3.21 ± .98 (95% CI 3.02-3.40). In comparison, the aggregated mean score and SD obtained in the original study (range=1-5) was 3.71 ± .81. Student responses and summary data by individual item are provided in Table 11. Results indicate that in average, only 24% of participants had an unfavorable view (i.e., disagree or strongly disagree) of NERVE VP design.

Table 10

Descriptive Statistics for Students' Perceptions of NERVE VP Design

	N	Range	Min.	Max.	Mean	Std. Deviation	Skewness		Kurtosis	
							Statistic	Std. Error	Statistic	Std. Error
Q2_25	110	4	1	5	3.03	1.053	-.343	.230	-.775	.457
Q2_26	110	4	1	5	2.95	1.070	-.229	.230	-.880	.457
Q2_27	110	4	1	5	3.25	1.009	-.697	.230	-.196	.457
Q2_28	110	4	1	5	3.38	.908	-.688	.230	.178	.457
Q2_29	110	4	1	5	3.54	.853	-.476	.230	-.069	.457
Q2_30	110	4	1	5	3.12	1.002	-.576	.230	-.209	.457
Valid N	110									

Table 11

Students' Perceptions of NERVE VP Design[a]

	Item	Strongly Disagree n (%)	Disagree n (%)	Neither Agree Nor Disagree n (%)	Agree n (%)	Strongly Agree n (%)	Median (IQR)
Q2_25	While working on the cases, I felt I had to make the same decisions a doctor would make in real life.	10 (8.5)	25 (21.2)	31 (26.3)	40 (33.9)	4 (3.4)	3.0 (2-4)
Q2_26	While working on the cases, I felt as if I were the doctor caring for the patients.	11 (9.3)	28 (23.7)	30 (25.4)	37 (31.4)	4 (3.4)	3.0 (2-4)
Q2_27	While working through the cases, I was actively engaged in revising my initial image of the patients' problems as new information became available.	8 (6.8)	16 (13.6)	31 (26.3)	50 (42.4)	5 (4.2)	3.5 (3-4)
Q2_28	While working through the cases, I was actively engaged in thinking about which findings supported or refuted each diagnosis in my differential diagnosis.	4 (3.4)	14 (11.9)	34 (28.8)	52 (44.1)	6 (5.1)	4.0 (3-4)
Q2_29	The questions that were given to me while working through the cases were helpful in enhancing my diagnostic reasoning in these cases.	1 (0.8)	13 (11.0)	32 (27.1)	54 (45.8)	10 (8.5)	4.0 (3-4)
Q2_30	The feedback I received was helpful in enhancing my diagnostic reasoning in the cases.	10 (8.5)	15 (12.7)	41 (34.7)	40 (33.9)	4 (3.4)	3.0 (3-4)

Note. [a]Items were measured on a 5-point Likert-type scale of agreement, where 1 = Strongly Disagree and 5 = Strongly Agree.

Students' Performance in Criterion Reference Measures

Since the purpose of this study was to examine the predictive validity of the Huwendiek et al.'s (2014) instrument, the following section presents the results of the analysis of students' performance on the following criterion reference measures: (a) CN quizzes, (b) VP cases, (c) SP/VP differential diagnosis, and (d) SP checklists.

Student Performance in CN Quizzes

Students were required to complete a minimum of 5 quizzes for CNs 3, 4, 5, 7, & 10. Each quiz had 10 questions on CN topics learned while exploring the NERVE Learning Center. Students were allowed to take quizzes again an unlimited amount of times. Eighty-seven of 118 students (73.7%) completed the CN 3 quiz at least once, while nine students (7.6%) took it for a second time and three students (2.5%) took it three times (Table 12). Ninety-two of 118 students (78%) completed the CN 4 quiz at least once while five students (4.2%) took it for a second time, one student (.8%) completed it three times, and two students (1.7%) took it four times (Table 13). Ninety of 118 students (76.3%) completed the CN 5 quiz at least once, while five students (4.2%) took it for a second time, and one student (.8%) completed it three times (Table 14). Eighty-five of 118 students (72%) completed the CN 7 quiz at least once, while 10 students (8.5%) took it for a second time and one student (.8%) took it three times (Table 15). Ninety-one of 118 students (77.1%) completed the CN 10 quiz at least once while six students (5.1%) took it for a second time, and one student (.8%) completed it three times (Table 16).

Table 12

Student frequency counts for CN 3 quiz

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	18	15.3	15.4	15.4
	1	87	73.7	74.4	89.7
	2	9	7.6	7.7	97.4
	3	3	2.5	2.6	100.0
	Total	117	99.2	100.0	
Missing	System	1	.8		
Total		118	100.0		

Table 13

Student frequency counts for CN 4 quiz

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	17	14.4	14.5	14.5
	1	92	78.0	78.6	93.2
	2	5	4.2	4.3	97.4
	3	1	.8	.9	98.3
	4	2	1.7	1.7	100.0
	Total	117	99.2	100.0	
Missing	System	1	.8		
Total		118	100.0		

Table 14

Student frequency counts for CN 5 quiz

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	21	17.8	17.9	17.9
	1	90	76.3	76.9	94.9
	2	5	4.2	4.3	99.1
	3	1	.8	.9	100.0
	Total	117	99.2	100.0	
Missing	System	1	.8		
Total		118	100.0		

Table 15

Student frequency counts for CN 7 quiz

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	21	17.8	17.9	17.9
	1	85	72.0	72.6	90.6
	2	10	8.5	8.5	99.1
	3	1	.8	.9	100.0
	Total	117	99.2	100.0	
Missing	System	1	.8		
Total		118	100.0		

Table 16

Student frequency counts for CN 10 quiz

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	19	16.1	16.2	16.2
	1	91	77.1	77.8	94.0
	2	6	5.1	5.1	99.1
	3	1	.8	.9	100.0
	Total	117	99.2	100.0	
Missing	System	1	.8		
Total		118	100.0		

Students' quiz mean scores ranged from 74.9% to 85.8% (Table 17), which are considered low by COM standards. However, student quiz scores had no impact on their grades, which is unusual for graded activities in the COM curriculum.

Table 17

Students' scores on 5 required quizzes in NERVE Learning Center[a]

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

Note. [a]Adapted from: Hirumi, A., Johnson, T., Reyes, R., Lok, B., Johnson, K., Bogert, K., Kubovec, S., Eakins, M., R., Rivera-Gutierrez, D., Kleinsmith, A., Bellew, M., & Cendan, J. (in press). Advancing virtual patient simulations through design research and Inter*PLAY*: Part II – Integration and field test. *Educational Technology, Research & Development*.

The next section presents the results of the analysis of students' performance in the VP cases embedded in NERVE.

Student Performance on VP Cases Embedded in NERVE

Students were required to complete a minimum of 3-6 VP cases to prepare for the SP/VP differential diagnosis educational activity. Students were allowed to work on the VP cases again to increase their scores an unlimited amount of times. Ninety-nine of 118 students (83.9%) completed the CN 3 VP case at least once. The mean score was 78.79, $SD_{CN3\ VP\ Case} = 14.16$ while scores ranged from 40% to 100% (Table 18) and 65.6% of students scored 80% or more (Table 19). One hundred of 118 students (84.7%) completed the CN 4 VP case at least once. The mean score was 75.10, $SD_{CN4\ VP\ Case} = 17.14$ while scores ranged from 30% to 100% (Table 18) and 51% of students scored 80% or more (Table 20). Ninety-six of 118 students (81.4%) completed the CN 5 VP case at least once. The mean score was 84.06, $SD_{CN5\ VP\ Case} = 13.10$ while scores ranged from 40% to 100% (Table 18) and 83.4% of students scored 80% or more (Table 21). Ninety-six of 118 students (81.4%) completed the CN 7 VP case at least once. The mean score was 72.92, $SD_{CN7\ VP\ Case} = 20.46$ while scores ranged from 10% to 100% (Table 18) and 54.2% of students scored 80% or more (Table 22). Ninety-eight of 118 students (83.1%) completed the CN 10 VP case at least once. The mean score was 70.92, $SD_{CN10\ VP\ Case} = 23.64$ while scores ranged from 10% to 100% (Table 18) and 50% of students scored 80% or more (Table 23).

Table 18

Student performance on NERVE VP cases, take 1

		cn_3_ take_1	cn_4_ take_1	cn_5_ take_1	cn_7_ take_1	cn_10_ take_1
N	Valid	99	100	96	96	98
	Missing	19	18	22	22	20
Mean		78.79	75.10	84.06	72.92	70.92
Median		80.00	80.00	90.00	80.00	75.00
Mode		80	90	90	80	100
Std. Deviation		14.162	17.144	13.105	20.464	23.642
Skewness		-.595	-.404	-1.109	-.773	-.523
Std. Error of Skewness		.243	.241	.246	.246	.244
Kurtosis		.256	-.550	1.633	.231	-.681
Std. Error of Kurtosis		.481	.478	.488	.488	.483
Range		60	70	60	90	90
Minimum		40	30	40	10	10
Maximum		100	100	100	100	100
Percentiles	25	70.00	60.00	80.00	60.00	50.00
	50	80.00	80.00	90.00	80.00	75.00
	75	90.00	90.00	90.00	90.00	90.00

Table 19

Student performance on CN 3 VP case, take 1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	40	3	2.5	3.0	3.0
	50	2	1.7	2.0	5.1
	60	11	9.3	11.1	16.2
	70	18	15.3	18.2	34.3
	80	31	26.3	31.3	65.7
	90	22	18.6	22.2	87.9
	100	12	10.2	12.1	100.0
	Total	99	83.9	100.0	
Missing	System	19	16.1		
Total		118	100.0		

Table 20

Student performance on CN 4 VP case, take 1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	30	2	1.7	2.0	2.0
	40	1	.8	1.0	3.0
	50	11	9.3	11.0	14.0
	60	15	12.7	15.0	29.0
	70	20	16.9	20.0	49.0
	80	14	11.9	14.0	63.0
	90	26	22.0	26.0	89.0
	100	11	9.3	11.0	100.0
	Total	100	84.7	100.0	
Missing	System	18	15.3		
Total		118	100.0		

Table 21

Student performance on CN 5 VP case, take 1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	40	2	1.7	2.1	2.1
	50	1	.8	1.0	3.1
	60	6	5.1	6.3	9.4
	70	7	5.9	7.3	16.7
	80	30	25.4	31.3	47.9
	90	31	26.3	32.3	80.2
	100	19	16.1	19.8	100.0
	Total	96	81.4	100.0	
Missing	System	22	18.6		
Total		118	100.0		

Table 22

Student performance on CN 7 VP case, take 1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	10	1	.8	1.0	1.0
	20	1	.8	1.0	2.1
	30	4	3.4	4.2	6.3
	40	3	2.5	3.1	9.4
	50	10	8.5	10.4	19.8
	60	10	8.5	10.4	30.2
	70	15	12.7	15.6	45.8
	80	23	19.5	24.0	69.8
	90	16	13.6	16.7	86.5
	100	13	11.0	13.5	100.0
	Total	96	81.4	100.0	
Missing	System	22	18.6		
Total		118	100.0		

Table 23

Student performance on CN 10 VP case, take 1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	10	1	.8	1.0	1.0
	20	2	1.7	2.0	3.1
	30	6	5.1	6.1	9.2
	40	8	6.8	8.2	17.3
	50	9	7.6	9.2	26.5
	60	10	8.5	10.2	36.7
	70	13	11.0	13.3	50.0
	80	16	13.6	16.3	66.3
	90	14	11.9	14.3	80.6
	100	19	16.1	19.4	100.0
	Total	98	83.1	100.0	
Missing	System	20	16.9		
Total		118	100.0		

Since some students completed the NERVE CN VP cases more than once, Repeated Measures ANOVA or Paired Samples t-tests were conducted to examine student differences in performance over time as they returned to the NERVE Learning Center to learn more about CNs before completing the VP cases again.

Table 24

Descriptive statistics for CN 3 take 1-3

	Mean	Std. Deviation	N
cn_3_take_1	60.00	17.321	3
cn_3_take_2	80.00	10.000	3
cn_3_take_3	96.67	5.774	3

Three out of 99 students (3.1%) completed CN 3 VP case at least 3 times (Table 24). The mean score for CN 3 take 1 was 60.00, $SD_{CN3\ take\ 1} = 17.32$, while the mean score for CN 3 take 2 was 80.00, $SD_{CN3\ take\ 2} = 10.00$, and the mean score for CN 3 take 3 was 96.67, $SD_{CN3\ take\ 3} = 5.77$ (Table 24). Mauchly's Test of Sphericity was non-significant ($p > 0.05$), thus sphericity is assumed and the assumption of homogeneity of variance was met (Table 25).

Table 25

Mauchly's Test of Sphericity for CN 3, take 1-3

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
NERVELearningCenter_Effect	.879	.129	2	.938	.892	1.000	.500

Note. Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix. [a]Design: Intercept Within Subjects Design: NERVELearningCenter_Effect. [b]May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

However, there was no statistically significant difference in CN 3 take 1-3 test scores (sphericity assumed $F_{2, 4} = 5.687$; $p > 0.05$) as depicted in Table 26. Results indicate that even though students' performance improved between the first and the last time they completed the CN 3 VP case, those differences in performance were not significant. One plausible explanation for this observation may be the small sample size (n=3).

Table 26

Test of within-subjects effects for CN 3 take 1-3

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
NERVE Learning Center_ Effect	Sphericity Assumed	2022.222	2	1011.111	5.687	.068	.740	11.375	.519
	Greenhouse-Geisser	2022.222	1.78	1133.550	5.687	.079	.740	10.146	.471
	Huynh-Feldt	2022.222	2.00	1011.111	5.687	.068	.740	11.375	.519
	Lower-bound	2022.222	1.00	2022.222	5.687	.140	.740	5.687	.280
Error(NERVE Learning Center_ Effect)	Sphericity Assumed	711.111	4	177.778					
	Greenhouse-Geisser	711.111	3.56	199.306					
	Huynh-Feldt	711.111	4.00	177.778					
	Lower-bound	711.111	2.00	355.556					

Note. [a]Computed using alpha = .05

Table 27

Descriptive statistics for CN 4 take 1-3

	Mean	Std. Deviation	N
cn_4_take_1	63.33	23.094	3
cn_4_take_2	86.67	5.774	3
cn_4_take_3	93.33	5.774	3

Three out of 100 students (3%) completed CN 4 VP case at least 3 times (Table 27). The mean score for CN 4 take 1 was 63.33, $SD_{CN4\ take\ 1}$ = 23.09, while the mean score for CN 4 take 2 was 86.67, $SD_{CN4\ take\ 2}$ = 5.774, and the mean score for CN 4 take 3 was 93.33, $SD_{CN4\ take\ 3}$ = 5.774 (Table 27). Mauchly's Test of Sphericity was non-significant ($p > 0.05$), thus sphericity is assumed and the assumption of homogeneity of variance was met (Table 28).

Table 28

Mauchly's test of spericity for CN 4 take 1-3

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
NERVE Learning Center_ Effect	.102	2.284	2	.319	.527	.613	.500

Note. Tests the null hypothesis that the error covariance matrix of the orthonormalized. [a]Design: Intercept Within Subjects Design: NERVELearningCenter_Effect transformed dependent variables is proportional to an identity matrix. [b]May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

However, there was no statistically significant difference in CN 4 take 1-3 test scores (sphericity assumed $F_{2,4} = 2.851$; $p > 0.05$) as depicted in Table 29. Results indicate that even though students' performance improved between the first and the last time they completed the

CN 4 VP case, those differences in performance were not significant. One plausible explanation for this observation may be the small sample size (n=3).

Table 29

Test of within-subjects effects for CN 4 take 1-3

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
NERVE Learning Center_ Effect	Sphericity Assumed	1488.889	2	744.444	2.851	.170	.588	5.702	.295
	Greenhouse-Geisser	1488.889	1.05	1413.063	2.851	.229	.588	3.004	.180
	Huynh-Feldt	1488.889	1.22	1213.622	2.851	.216	.588	3.498	.201
	Lower-bound	1488.889	1.00	1488.889	2.851	.233	.588	2.851	.173
Error(NERVELearning Center_ Effect)	Sphericity Assumed	1044.444	4	261.111					
	Greenhouse-Geisser	1044.444	2.10	495.626					
	Huynh-Feldt	1044.444	2.45	425.673					
	Lower-bound	1044.444	2.00	522.222					

Note. [a]Computed using alpha = .05

Table 30

Descriptive statistics for CN 5 take 1-2

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	cn_5_take_1	81.67	6	7.528	3.073
	cn_5_take_2	90.00	6	10.954	4.472

Six out of 96 students (6.3%) completed CN 5 VP case at least 2 times (Table 30) so the sample was analyzed using a paired samples t-test. The mean score for CN 5 take 1 was 81.67,

$SD_{CN5\ take\ 1}$ = 7.528, while the mean score for CN 5 take 2 was 90.00, $SD_{CN5\ take\ 2}$ = 10.954 (Table 30). However, there was no statistically significant difference between CN 5 take 1 and CN 5 take 2 mean scores ($t = -1.74$, $df = 5$, $p > .05$) as depicted in Table 31. Moreover, the 95% confidence interval indicates the true mean difference between paired samples may range from $-20.60<\mu<3.93$, crossing through zero, which raises the possibility of zero true differences between the means (Table 31). Results indicate that even though students' performance improved between the first and the second time they completed the CN 5 VP case, those differences in performance were not significant. One plausible explanation for this observation may be the small sample size (n=6).

Table 31

Paired samples test for CN 5 take 1-2

		Paired Differences							
				Std. Error	95% Confidence Interval of the Difference				Sig. (2-
		Mean	Std. Deviation	Mean	Lower	Upper	t	df	tailed)
Pair 1	cn_5_take_1 - cn_5_take_2	-8.33	11.69	4.77	-20.60	3.93	-1.74	5	.141

Table 32

Descriptive statistics for CN 7 take 1-2

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	cn_7_take_1	67.27	11	17.373	5.238
	cn_7_take_2	85.45	11	16.348	4.929

Eleven out of 96 students (11.5%) completed the CN 7 VP case at least 2 times (Table 32) so the sample was analyzed using a paired samples t-test. The mean score for CN 7 take 1 was 67.27, $SD_{CN7\ take\ 1}$ = 17.373, while the mean score for CN 7 take 2 was 85.45, $SD_{CN7\ take\ 2}$ =

16.348 (Table 32). There was a statistically significant difference between CN 7 take 1 and CN 7 take 2 mean scores ($t = -3.50$, $df = 10$, $p < .05$) as depicted in Table 33. Moreover, the 95% confidence interval indicates the true mean difference between paired samples may range from $-29.74<\mu<-6.61$, never crossing through zero, which indicates the true mean difference between the means will never be zero (Table 33). Results suggest students benefited from using the NERVE Learning Center as a learning resource for completing the CN 7 VP case in NERVE.

Table 33

Paired samples test for CN 7 take 1-2

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	cn_7_take_1 - cn_7_take_2	-18.18	17.21	5.19	-29.74	-6.61	-3.503	10	.006

Table 34

Descriptive statistics for CN 10 take 1-2

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	cn_10_take_1	57.14	7	21.381	8.081
	cn_10_take_2	95.71	7	7.868	2.974

Seven out of 98 students (7.2%) completed the CN 10 VP case at least 2 times (Table 34) so the sample was analyzed using a paired samples t-test. The mean score for CN 10 take 1 was 57.14, $SD_{CN10\ take\ 1} = 21.381$, while the mean score for CN 10 take 2 was 95.71, $SD_{CN10\ take\ 2} =$ 7.974 (Table 34). There was a statistically significant difference between CN 10 take 1 and CN 10 take 2 mean scores ($t = -5.47$, $df = 6$, $p < .05$) as depicted in Table 35. Moreover, the 95%

confidence interval indicates the true mean difference between paired samples may range from -55.81<μ<-21.32, never crossing through zero, which indicates the true mean difference between the means will never be zero (Table 35). Results suggest students benefited from using the NERVE Learning Center as a learning resource for completing the CN 10 VP case in NERVE.

Table 35

Paired samples test for CN 10 take 1-2

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	cn_10_take_1 - cn_10_take_2	-38.57	18.64	7.04	-55.81	-21.32	-5.473	6	.002

Once students had interacted with NERVE and completed required quizzes and VP cases embedded in NERVE, researchers conducted and AAR followed by two NERVE-assisted VP/SP sessions to measure the students' ability to transfer skills learned from the NERVE VPs to other clinical situations. During these VP/SP sessions, students were required to interview the SP while conducting the physical exam using NERVE to provide a differential diagnosis of the patient's condition. SPs also evaluated the students using a standardized SP checklist. The next section presents the results of the analysis of students' performance in the NERVE-assisted VP/SP sessions.

Student Performance on VP/SP Differential Diagnosis

One hundred-seventeen students (100%) completed the NERVE-assisted VP/SP assessments and provided a differential diagnosis of the patient (Table 36). Results show that 108 out of 117 students (92.3%) correctly identified the damaged nerve (CN 6) as illustrated in Table 37, while 115 out of 117 students (98.3%) correctly identified the damaged CN side (left) as depicted in Table 38. Finally, results show that 102 out of 117 students (87%) provided a differential diagnosis congruent with the clinical SP case and VP examination hybrid encounter (Table 39). Results suggest students successfully transferred most clinical reasoning skills learned using NERVE to the NERVE-assisted SP/VP clinical assessment.

Table 36

Descriptive statistics for SP/VP Assessment

		Describe the localization of the problem. Cranial nerve (select one option from the drop-...	Describe the localization of the problem. Side (select one option from the drop-down list):	Primary diagnosis (select one option from the drop-down list):
N	Valid	117	117	117
	Missing	0	0	0
Mean		5.82	1.03	11.11
Median		6.00	1.00	11.00
Mode		6	1	8
Std. Deviation		.665	.206	3.952
Skewness		-3.907	8.646	.850
Std. Error of Skewness		.224	.224	.224
Kurtosis		15.311	78.051	-.538
Std. Error of Kurtosis		.444	.444	.444
Range		4	2	13
Minimum		2	1	7
Maximum		6	3	20
Percentiles	25	6.00	1.00	8.00
	50	6.00	1.00	11.00
	75	6.00	1.00	16.00

Table 37

Student performance in identifying damaged CN during SP/VP Assessment

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	CN2	1	.9	.9	.9
	CN3	2	1.7	1.7	2.6
	CN4	5	4.3	4.3	6.8
	CN5	1	.9	.9	7.7
	CN6	108	92.3	92.3	100.0
	Total	117	100.0	100.0	

Table 38

Student performance in identifying damaged CN side during SP/VP Assessment

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Left	115	98.3	98.3	98.3
	Right	1	.9	.9	99.1
	Bilateral	1	.9	.9	100.0
	Total	117	100.0	100.0	

Table 39

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	.9	.9
	Transient Ischemic Attack	1	.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	

Note. [a]Adapted from: Hirumi, A., Johnson, T., Reyes, R., Lok, B., Johnson, K., Bogert, K., Kubovec, S., Eakins, M., R., Rivera-Gutierrez, D., Kleinsmith, A., Bellew, M., & Cendan, J. (in press). Advancing virtual patient simulations through design research and Inter*PLAY*: Part II – Integration and field test. *Educational Technology, Research & Development*.

Student Performance on VP/SP Assessments Using SP Checklists

One hundred-seventeen students (100%) were evaluated by SPs using SP checklists and the median percent correct score was 93.3% or 14/15 items (IQR = 93.3-100; range = 66.7-100) as depicted in Table 40. One hundred-eleven out of 117 students (94.9%) scored at least 13 items correct (Table 41) or at least 80% (Table 42), which is the lower bound expected standard of performance at COM. Results suggest students successfully transferred most clinical reasoning and decision making skills learned using NERVE to the SP/VP clinical assessment.

Table 40

Student performance on SP checklists (*N*=117)[a,b]

Item	Completed n (%)	Not Completed n (%)
Greeted patient warmly and verified patient's identity	116 (99.1)	1 (0.9)
Washed hands before patient contact and maintained clean technique	112 (95.7)	5 (4.3)
Introduced him/herself to the patient (first and last name, full title)	116 (99.1)	1 (0.9)
Explained purpose of encounter within the first 1-2 minutes.	110 (94.0)	7 (6.0)
Treated the patient with respect.	117 (100.0)	0 (0.0)
Listened attentively.	116 (99.1)	1 (0.9)
Demonstrated genuineness, care, concern, empathy.	113 (96.6)	4 (3.4)
Expressed interest in the patient as a person.	109 (93.2)	8 (6.8)
Used open-ended techniques that encouraged the patient to tell his/her story	117 (100.0)	0 (0.0)
Explored the patient's worries/fears about cause(s)/implications.	101 (86.3)	16 (13.7)
Explored how health issues have affected the patient.	101 (86.3)	16 (13.7)
Communicated clearly, avoided medical jargon or explained terms when used.	116 (99.1)	1 (0.9)
Provided information related to the working diagnosis and/or next steps.	110 (94.0)	7 (6.0)
Encouraged patient to develop full and accurate understanding of key messages.	87 (74.4)	30 (25.6)
Asked if the patient has any other questions or concerns prior to leaving room.	109 (93.2)	8 (6.8)

Note. [a]Median percent correct score = 93.3% or 14/15 items; interquartile range = 93.3-100.0; minimum-maximum = 66.7-100.0. [b]Adapted from: Hirumi, A., Johnson, T., Reyes, R., Lok, B., Johnson, K., Bogert, K., Kubovec, S., Eakins, M., R., Rivera-Gutierrez, D., Kleinsmith, A., Bellew, M., & Cendan, J. (in press). Advancing virtual patient simulations through design research and Inter*PLAY*: Part II – Integration and field test. *Educational Technology, Research & Development*.

Table 41

Student performance on SP checklists (raw points)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	10	3	2.6	2.6	2.6
	11	3	2.6	2.6	5.1
	12	5	4.3	4.3	9.4
	13	15	12.8	12.8	22.2
	14	33	28.2	28.2	50.4
	15	58	49.6	49.6	100.0
	Total	117	100.0	100.0	

Table 42

Student performance on SP checklists (percentages)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	66.7	3	2.6	2.6	2.6
	73.3	3	2.6	2.6	5.1
	80.0	5	4.3	4.3	9.4
	86.7	15	12.8	12.8	22.2
	93.3	33	28.2	28.2	50.4
	100.0	58	49.6	49.6	100.0
	Total	117	100.0	100.0	

Chapter 4 presented the results of four different sets of analyses conducted to evaluate the hypotheses and research questions posited in this study. The first set of analyses examined the characteristics of the sample. The second set of analyses investigated the validity and reliability of Huwendiek et al.’s (2014) VP design evaluation instrument as it relates to the scores obtained from the intervention. The third and fourth sets of analyses addressed the hypotheses and research questions postulated in this study. Chapter 5 will discuss the results as it relates to the research questions and hypotheses posited in this dissertation, implications, and will provide recommendations for future research.

CHAPTER FIVE
DISCUSSION

The purpose of this study was to examine the predictive validity of Huwendiek et al.'s (2014) instrument for evaluating the design effectiveness of VPs. The study focused on VPs designed specifically to promote development of clinical reasoning. I examined its predictive validity by investigating if students' perceptions of NERVE VP design, based on Huwendiek et al.'s (2014) instrument, are correlated to students' performance on criterion reference measures.

Chapter 4 reported the results of the data analyses performed to test the research hypotheses and answer the research questions. Chapter 5 discusses the findings as it relates to the hypotheses and research questions. Chapter 5 is organized based on the two research questions and hypotheses postulated in this dissertation. Each research question is addressed by discussing the findings of the CFA, including fit indices related to each specified model and individual correlations (Lambdas) found among the independent and outcome variables. Then, this chapter discusses the findings in relation to prior research on the development of clinical reasoning expertise and the design VP simulations. Finally, limitations and implications are discussed and recommendations are made for future research.

Research Question One: Does the effectiveness of NERVE VPs design correlate with student learning?

Significant correlations were found between measures of NERVE VP design and student learning. Specifically, significant correlations were found between the NERVE VP design factors evaluated by Huwendiek et al.'s (2014) instrument and students' performance in quizzes and VP cases embedded NERVE. The following section discusses the first three significant correlations in detail to include the model fit indices and Lambdas representing the correlation between the effectiveness of NERVE VP design and student performance in quizzes. It was hypothesized that the specified model for students' perceptions of effectiveness of NERVE VPs design would be correlated to the achievement of higher student clinical reasoning learning outcomes in NERVE.

NERVE VP Design and Quizzes

Significant correlations were found between two of the three primary NERVE VP design factors specified by Huwendiek et al.'s (2014) instrument and students' performance in quizzes embedded NERVE. The CFA using Maximum Likelihood (ML) estimation yielded the following fit indices: (a) SRMR=.02; (b) RMSEA=.06; and (c) Bentler CFI=.99 (Figure 18). According to Hu and Bentler (1999), "for the ML method, a cutoff value close to .95 for CFI, … a cutoff value close to .08 for SRMR; and a cutoff value close to .06 for RMSEA are needed before we can conclude that there is a relatively good fit between the hypothesized model and the observed data" (p.1). Results indicate a very good fit for the specified model suggesting the existence of a correlation between the effectiveness of NERVE VP design and student performance in quizzes (Figure 18), as hypothesized. The SRMR of .02 for quizzes was well below the recommended cutoff value (.08), suggesting the presence of a significant correlation between the NERVE VP design variables and students' performance in quizzes. However, inspection of the Lambdas (or Beta weights) standardized coefficients revealed Lambdas (Figure 19), were not all significant at the $p \leq .05$ level.

Authenticity and Quizzes

No significant correlations were found between the first factor in Huwendiek et al.'s (2014) instrument (i.e., authenticity of the patient encounter and the consultation) and students' performance in quizzes embedded in NERVE. Moreover, inspection of the Lambdas (Figure 19) revealed no significant correlations (Lambda 2 = .26, $p = .10$ and Lambda 3 = .29, $p = .09$), suggesting the *authenticity of the patient encounter and the consultation* was not significantly correlated to students' performance in quizzes embedded in NERVE. This finding stands in stark contrast to significant correlations found between *cognitive strategies* or *coaching during the consultation* as discussed in the next two sections of this chapter. Answering CN quiz questions may be more dependent on retrieval of facts from memory. Thus, one alternative explanation is that the *authenticity of the patient encounter and the consultation* may not be as relevant to figuring out the right answer to a quiz question. Another explanation may be that answering CN quiz questions may be more associated with *cognitive strategies* or *coaching during the consultation*. For instance, the use of effective cognitive strategies in the consultation requires students to engage in *gathering information* as a pre-requisite to the *development of mental representations of the patient's problem and situation* (Gruppen & Frohna, 2002). Mental representations are then *evaluated against the actual patient condition* and *prior knowledge* to reach a diagnosis (Gruppen & Frohna, 2002). Thus, information derived from quiz questions helps the learner engage in the *cognitive strategies* needed to diagnose the patient's condition.

<u>Cognitive Strategies and Quizzes</u>

A significant correlation was found between the second factor in Huwendiek et al.'s (2014) instrument (i.e., cognitive strategies in the consultation) and students' performance in quizzes embedded in NERVE. Specifically, *cognitive strategies in the consultation* were significantly correlated to students' performance in quizzes at the $p \leq .05$ level (Figure 19). Inspection of the Lambda revealed a significant correlation (Lambda 1 = .82, $p < .05$), indicating that *cognitive strategies in the consultation* were significantly correlated to students' performance in quizzes embedded in NERVE.

Coaching and Quizzes

Significant correlations were found between the third factor in Huwendiek et al.'s (2014) instrument (i.e., coaching during the consultation) and students' performance in quizzes embedded in NERVE. Inspection of the Lambdas (Figure 19) revealed two significant correlations (Lambda 4 = .84, $p < .0001$ and Lambda 5 = .83, $p < .0001$), indicating that *coaching during the consultation* was significantly correlated to students' performance in quizzes embedded in NERVE. No other significant correlations were found in this group of analyses.

NERVE VP Design and VP Cases

Significant correlations were found between all three primary VP design factors evaluated by Huwendiek et al.'s (2014) instrument and students' performance in VP cases embedded NERVE. The CFA yielded the following fit indices: (a) SRMR=.01; (b) RMSEA=.00; and (c) Bentler CFI=1.00 (Figure 21). Results indicate a very good fit for the specified model suggesting the existence of a correlation between NERVE VP design and students' performance in VP cases embedded in NERVE (Figure 21). The SRMR of .01 for VP cases was well below the recommended cutoff value (.08), suggesting the presence of a significant correlation between the NERVE VP design variables and students' performance in VP cases. Moreover, inspection of the Lambdas (or Beta weights) standardized coefficients revealed Lambdas (Figure 22) were all significant at the $p < .0001$ level. The next section discusses the significant correlations between the three factors in Huwendiek et al.'s (2014) instrument and students' performance in VP cases embedded in NERVE in more detail.

Authenticity and VP Cases

Significant correlations were found between the first factor in Huwendiek et al.'s (2014) instrument (i.e., authenticity of the patient encounter and the consultation) and students' performance in VP cases embedded in NERVE. Inspection of the Lambdas (Figure 22) revealed two significant correlations (Lambda 1 = .37, p = < .0001 and Lambda 2 = .79, p = < .0001), indicating that *authenticity of the patient encounter and the consultation* was significantly correlated to students' performance in VP cases embedded in NERVE.

Cognitive Strategies and VP Cases

Significant correlations were found between the second factor in Huwendiek et al.'s (2014) instrument (i.e., cognitive strategies in the consultation) and students' performance in VP cases embedded in NERVE. Inspection of the Lambdas (Figure 22) revealed two significant correlations (Lambda 3 = .51, p < .0001 and Lambda 4 = .61, p < .0001), indicating that *cognitive strategies in the consultation* were significantly correlated to students' performance in VP cases embedded in NERVE.

Coaching and VP Cases

A significant correlation was found between the third factor in Huwendiek et al.'s (2014) instrument (i.e., coaching during the consultation) and students' performance in NERVE VP cases. Inspection of the Lambda (Figure 22) revealed a significant correlation (Lambda 5 = 1.01, p < .0001), indicating that *coaching during the consultation* was significantly correlated to students' performance in VP cases embedded in NERVE. No other significant correlations were found in this group of analyses.

Research Question Two: Does the effectiveness of NERVE VPs design correlate with students' skills transfer?

Significant correlations were found between measures of NERVE VP design and student transfer. Specifically, significant correlations were found between the NERVE VP design factors evaluated by Huwendiek et al.'s (2014) instrument and students' performance in the NERVE-assisted VP/SP diagnosis and SP checklists. The following section discusses the significant correlations in detail to include the model fit indices and Lambdas representing the correlation between the effectiveness of NERVE VP design and student performance in VP/SP diagnosis. It was hypothesized that the specified model for students' perceptions of effectiveness of NERVE VPs design would be correlated to the achievement of higher student clinical reasoning transfer outcomes in NERVE.

NERVE VP Design and VP Diagnosis

Significant correlations were found between the NERVE VP design factors evaluated by Huwendiek et al.'s (2014) instrument and students' performance in VP/SP diagnosis and SP checklists. The CFA procedure yielded the following fit indices: (a) SRMR=.07; (b) RMSEA=.03; and (c) Bentler CFI=.96 (Figure 24). Results indicate a good fit for the specified model suggesting the existence of a correlation between NERVE VP design and students' performance in VP/SP diagnosis and SP checklists (Figure 24). The SRMR of .07 was below the recommended cutoff value (.08), suggesting the presence of a significant correlation between the NERVE VP design variables and students' performance in VP/SP diagnosis and SP checklists. However, inspection of the Lambdas (or Beta weights) standardized coefficients revealed Lambdas (Figure 25) were not all significant at the $p \leq .05$ level. Results suggest that although the overall model explained the relationships between the variables, there is room for improvement. In particular, the SP questions that were not significant (i.e., SPQ1 [greeted patient warmly], SPQ2 [washed hands], and SPQ4 [explained purpose of encounter]) appear to be *procedural* in nature and thus do not appear to be as associated to simulation *authenticity, employing cognitive skills,* and *coaching*, as do the others variables in the model. The next section discusses these significant correlations in more detail.

<u>NERVE VP Design and CN Type</u>

No significant correlations were found between any of the three NERVE VP design variables specified in Huwendiek et al.'s (2014) instrument and the CN Type outcome variable. Even though the specified model achieved its best level of fit to the data when the *CN Type* outcome variable was included (Figure 24), no significant relationships between NERVE VP design and *CN Type* (Figure 25) were observed (Lambda 1 = .09, p = .34). The *CN Type* outcome variable is closely related to the *CN Side* outcome variable*,* which was eliminated from the specified model during the analysis phase due to a poor fit. Moreover, although one hundred-eight out of 117 students (92.3%) correctly identified the *CN Type* (Table 40) and *CN Side* (Table 41), the correlation between *cognitive strategies (F2)* and *CN Type* specified in the model (Figure 25) was not significant. Finally, the model did not correlate *authenticity* and *coaching* to the *CN Type* outcome variable. A potential explanation may be the ceiling effect (low variance) observed from scores on the *CN Type* outcome variable. Results warrant further testing.

<u>NERVE VP Design and Differential Diagnosis</u>

No significant correlations were found between any of the three NERVE VP design variables specified in Huwendiek et al.'s (2014) instrument and the *Differential Diagnosis* outcome variable. Inspection of the Lambda (Figure 25) revealed no significant correlations

(Lambda 2 = .14, p = .10), indicating that none of the *three NERVE VP design* variables specified in Huwendiek et al.'s (2014) instrument was significantly correlated to students' performance in the NERVE-assisted *Differential Diagnosis*. A potential explanation may be the ceiling effect (low variance) observed from scores on the *Differential Diagnosis* outcome variable (Table 39). Results warrant further testing.

NERVE VP Design and SP Checklist

Significant correlations were found between the NERVE VP design factors evaluated by Huwendiek et al.'s (2014) instrument and students' performance in SP checklists. However, inspection of the Lambdas (or Beta weights) standardized coefficients revealed Lambdas (Figure 25) were not all significant at the $p \leq .05$ level. The following section discusses the significant correlations in detail.

Authenticity and SP Checklist Questions

Significant correlations were found between the first factor in Huwendiek et al.'s (2014) instrument (i.e., authenticity of the patient encounter and the consultation) and student performance in SP checklists. Specifically, *authenticity of the patient encounter and consultation* was significantly correlated to students' performance in SPQ7 and SPQ11 (Figure 25). Inspection of the Lambdas revealed a significant correlation between *authenticity of the patient encounter and consultation* (Lambda 6 = .70, $p < .0001$) and SPQ7 (i.e., demonstrated genuineness, care, concern, empathy). Inspection of the Lambdas also revealed a significant correlation between *authenticity of the patient encounter and consultation* (Lambda 9 = .26, $p \leq .005$) and SPQ11 (i.e., explored how health issues have affected the patient). Findings suggest that *authenticity of the patient encounter and consultation* was significantly correlated to students' performance in SP questions. Therefore, *authenticity of the patient encounter and the consultation* was significantly correlated to students' skills transfer using NERVE.

SPQ7 and SPQ11 appear to be related to each other and to the caregiver showing a genuine interest in understanding and empathizing with the patient's condition. These behaviors are consistent with projecting an image of caring, which relates to being authentic and to treating the situation, simulated or not, as authentic. Thus, the behaviors represented by SPQ7 and SPQ11 appear to be related to *authenticity*, as demonstrated by the findings. No other significant correlations were found in this group of analyses.

<u>Cognitive Strategies and SP Checklist Questions</u>

Significant correlations were found between the second factor in Huwendiek et al.'s (2014) instrument (i.e., cognitive strategies in the consultation) and student performance in SP checklists. Specifically, *cognitive strategies in the consultation* were significantly correlated to students' performance in SPQ13, SPQ14 and SPQ15 (Figure 25). Inspection of the Lambdas revealed a significant correlation between *cognitive strategies in the consultation* (Lambda 11 = .80, $p < .0001$) and SPQ13 (i.e., provided information related to the working diagnosis and/or next steps). Inspection of the Lambdas also revealed a significant correlation between *cognitive strategies in the consultation* (Lambda 12 = .36, $p \leq .0001$) and SPQ14 (i.e., encouraged patient to develop and demonstrate a full and accurate understanding of key messages). Finally, inspection of the Lambdas revealed a significant correlation between *cognitive strategies in the consultation* (Lambda 13 = .81, $p < .0001$) and SPQ15 (i.e., asked if the patient has any other questions or concerns prior to leaving the room). Findings suggest that *cognitive strategies in the consultation* were significantly correlated to students' performance in SP questions. Therefore, *cognitive strategies in the consultation* were significantly correlated to students' skills transfer using NERVE.

Providing information related to the diagnosis (SPQ13) could be considered an outcome of Bowen's (2006) instructional strategies that ask students for: (a) single-sentence summaries of patient problems in abstract terms, (b) the discriminating features of a set of diagnostic hypotheses, (c) comparing and contrasting diagnostic hypotheses, and (d) demonstrating presentations of diagnostic hypotheses and the relative probability of different diagnoses. Bowen's (2006) recommendations require students to go through steps three to five of Gruppen and Frohna's (2002) model of clinical reasoning: (a) information gathering, (b) problem representation, and (c) evaluation.

Encouraging the patient to develop and demonstrate a full and accurate understanding of key messages (SPQ14) is essentially the logical follow-up action to SPQ13. Once the student has *provided the patient with information related to the diagnosis (SPQ13)*, the student would ask for feedback from the patient to ensure the explanation was not too technical or complicated to be understood. *Offering specific feedback* is one of the VP design principles proposed by Huwendiek et al. (2009a). In addition, the exchange of feedback is a useful *cognitive strategy in the consultation* since: (a) students have to understand first what the patient's problem is before they can explain it clearly and correctly, (b) the patient not understanding well may be a sign that the student has not developed a full understanding the diagnosis, and (c) having to explain again the patient's problem may force the student to revisit their mental representations of the patient's condition. Students' asking for feedback is also related to the third step in Gruppen and Frohna's (2002) model of clinical reasoning (i.e., information gathering), which acts as a precursor to the next two steps in Gruppen and Frohna's (2002) model (i.e., problem representation and evaluation).

Finally, *asking if the patient has any other questions or concerns prior to leaving the room (SPQ15),* is closely related to and in a way a logical follow-up action to SPQ14. Once the student has ensured the patient understands the preliminary diagnosis (SPQ14), the student has to communicate what the next steps are in terms of diagnostic tests (e.g., blood work, imaging) and treatment (e.g., hospitalization, referral to a specialist). Consequently, the student wants to ensure the patient has understood these next steps (SPQ15). This is also another opportunity for *information gathering*, which could lead to the discovery of new information. Discovering new information may lead to another iteration of steps 3-5 of Gruppen and Frohna's (2002) model of clinical reasoning (i.e., information gathering, problem representation and evaluation). New iterations may result in the students' employment of *cognitive strategies in the consultation* by making continuous comparisons between existing mental representations of the patient's problems and the actual patient condition. Finally, students use new mental representations of patient's problems for making informed decisions regarding the type of information needed to effectively diagnose the patient's condition. No other significant correlations were found in this group of analyses.

<u>Coaching and SP Checklist Questions</u>

Significant correlations were found between the third factor in Huwendiek et al.'s (2014) instrument (i.e., coaching during the consultation) and student performance in SP checklists. Specifically, *coaching during the consultation* was significantly correlated to students' performance in SPQ8, SPQ10 and SPQ12 (Figure 25). Inspection of the Lambdas revealed a significant correlation between *coaching during the consultation* (Lambda 7 = .55, $p < .0001$) and SPQ8 (i.e., expressed interest in the patient as a person). Inspection of the Lambdas also revealed a significant correlation between *coaching during the consultation* (Lambda 8 = .20, $p \leq$

.01) and SPQ10 (i.e., explored the patient's perspective of illness or worries/fears about cause(s)/implications). Finally, inspection of the Lambdas revealed a significant correlation between *coaching during the consultation* (Lambda 10 = .71, $p < .0001$) and SPQ12 (i.e., communicated clearly, avoided medical jargon or explained terms when used). Therefore, *coaching during the consultation* was significantly correlated to students' clinical reasoning skills transfer using NERVE.

Expressing interest in the patient as a person (SPQ8) could be related to *coaching during the consultation*. During the NERVE-assisted SP/VP assessment, coaching took place from: (a) NERVE to the student, and (b) from the student to the SP. How NERVE provided coaching during the consultation was discussed previously in this chapter. However, how the student coached the SP during the consultation appears to be more relevant here since the SP questions were answered from the SPs' perspective. In a review of the doctor-patient communication literature, Ong, De Haes, Hoos, and Lammes (1995) indicated that "three different purposes of communication are identified, namely: (a) creating a good inter-personal relationship; (b) exchanging information; and (c) making treatment-related decisions" (p. 903). One could argue that *expressing interest in the patient as a person (SPQ8)* is related to creating a good inter-personal relationship but there may be other reasons. According to Ong et al. (1995), "another purpose of medical communication is to enable doctors and patients to make decisions about treatment. Traditionally the ideal doctor-patient relationship was paternalistic: the doctor directs care and makes decisions about treatment. During the past two decades, this approach has been replaced by the ideal of 'shared decision-making'. It appears logical that in order to make such decisions, patients need information" (p. 905). Thus, doctors also have a responsibility in *coaching the patients during the consultation* to assist them in making care, treatment, and at times, life-changing decisions. In addition, the feedback doctors receive from patients during the coaching process facilitates the doctors' clinical reasoning process. Feedback from patients allows doctors to continually refine their mental representations of the patients' problem to reach a successful diagnosis, which according to Gruppen and Frohna (2002) is an iterative process.

Exploring the patient's perspective of illness or worries/fears about cause(s)/implications (SPQ10) is also related to *coaching during the consultation*. As doctors inquire about the patients' worries or fears and receive feedback, doctors assume a coaching role (Ong et al., 1995) and try to put the patients at ease by offering positive information regarding treatment and expected outcomes.

Finally, *communicating clearly by avoiding medical jargon and explaining terms when used (SPQ12)* is also related to *coaching during the consultation*. For instance, the coaching role doctors assume during the consultation (Ong et al., 1995), would be severely hampered by unclear communication. Thus, avoiding medical jargon facilitates *coaching during the consultation.* No other significant correlations were found in this group of analyses.

Relationship with VP Design Effectiveness, Learning, and Transfer

The relationships between the VP design effectiveness are discussed in relation to student learning and transfer. The specific features of NERVE that were designed to promote authenticity, cognitive strategies, and coaching are identified, followed by a discussion of how the three VP design factors were related to student learning and transfer.

Relationship between Authenticity, Learning, and Transfer

This section discusses *authenticity of the patient encounter and the consultation* in relation to student learning and transfer to include: (a) NERVE features that promoted the students' perception of *authenticity* and how NERVE could be improved, and (b) relationship between *authenticity* and the learning and transfer of clinical reasoning skills.

NERVE Features that Promote Authenticity

According to a recent systematic review of 17 studies on VP design (Reyes, R.J., Hirumi, A., 2016), most VPs used in the reviewed studies offered wide-ranging levels of fidelity but did not include the ability to have a two-way dialogue with the simulated patient. NERVE is a highly interactive VP suite where students can ask written questions to the simulated patients and the VPs will offer verbal, non-verbal, and written responses back to learners. NERVE VP responses include information about family history, symptoms, and conditions simulating the real-life dialogue doctors would normally have with patients and providing a heightened sense of realism to the consultation. NERVE VPs can perform a series of movements to include eyes, head, mouth, tongue, lips, raising their arms, covering their eyes, and many others. NERVE VPs also provide real-time information regarding their ability to see, hear, smell, and feel. Students also become familiar with the simulated tools (e.g., eye chart, ophthalmoscope, hand tool) used to perform routine patient examinations before proceeding to the exam room. During the eye exam, for example, students can even examine the VP's cornea and the optic nerve using the simulated ophthalmoscope. Therefore, NERVE provides a series of unique interactive features that elevate the level of realism or authenticity. Another factor that contributed to the perception of authenticity of the patient encounter and consultation in NERVE was clinical variation. NERVE offered students the opportunity to examine up to six patients presenting six different CN palsies, emulating the variety of clinical conditions doctors experience in their clinical practice. In addition, the NERVE Learning Center offers a variety of published clinical cases for all CNs, greatly expanding the realm of CN-related clinical cases available for students to learn from and develop their clinical reasoning expertise. Finally, NERVE design was informed by Inter*PLAY* instructional theory (Hirumi et al., 2016), which at is core, is based on experiential learning precepts (Kolb, 1984), thus entailing authenticity. Altogether, the high fidelity, highly interactive and realistic features of NERVE facilitated the students' perceptions of *authenticity* to enhance clinical reasoning expertise but there is room for improvement.

NERVE could be enhanced by integrating real-life clinical cases where copies of real imaging, laboratory, and other diagnostic reports (without any personally-identifiable information) are presented to students, on demand, as they conduct the examination.

Authenticity, Learning, and Transfer

Research suggests there is a correlation between authenticity and achievement of targeted learning outcomes. According to Huwendiek et al.'s (2009a) VP design principles, VP simulations should "provide an authentic web-based interface and students tasks" (p. 586) to facilitate development of clinical reasoning skills. Moreover, a systematic review of 109 studies published between 1969 and 2003 concluded that *simulation validity or authenticity* is one of the key characteristics of high-fidelity medical simulations that lead to effective learning (Issenberg et al., 2005). Many other VP design studies since 2005 have also agreed that *realism or authenticity* is a key VP design feature. For instance, Kim et al. (2006) found that the relevance and realism of VP teaching cases is vital for the development of clinical reasoning expertise. In addition, Wilson (2012) found VPs were perceived as "realistic, relevant, and developed for the local environment. They felt that they were facing people and situations they would see in practice and reacting as they would in the emergency department" (p. 119) and concluded that a "VP is an authentic activity" (p. 121). Similarly, Salminen et al. (2014) revealed, "students found working with the primary care VP to be active and meaningful with a sense of authenticity" (p. 6). Moreover, Botezatu et al. (2010a) identified *authenticity* as one of the five factors associated with effective VP simulations use in medical education. Botezatu et al. (2010a) indicated, "medical students perceive VPs as important learning and assessment tools, fostering clinical reasoning, in preparation for the future clinical practice" (p. 1). Finally, Botezatu et al. (2010b) indicated, "the aspect of paying attention to VPS design, which should enhance clinical reasoning abilities and support case authenticity, cannot be overemphasized. Authenticity, however, extends well beyond the design of the interface. The users are more positive to the use of an application when the case content is robust, derived from everyday practice and supported by feedback providing an exposé of actual patient treatment and evolution" (p. 516). Altogether, research suggests there is a relationship between the *authenticity of the patient encounter and the consultation* and the development of clinical reasoning and decision making skills.

Research also suggests there is a correlation between *authenticity of the patient encounter and the consultation* and the transfer of skills learned from VPs to other clinical situations. For instance, according to Botezatu et al. (2010a), the "transfer of knowledge to the real patient is the ultimate goal of simulation technology" (P. 6). Moreover, the *authenticity* of VP simulations has been associated with knowledge and skills transfer to real patients (Botezatu et al., 2010a). Presumably, transferability is enhanced when using real patient cases, where students make associations between the VP cases and the real patients. However, authenticity should go beyond the clinical cases to include the provision of feedback regarding the natural evolution of the patient condition and "to reflect the real clinical practice" (Botezatu et al., 2010a, p. 5). Similarly, Issenberg et al.'s (2005) systematic review of 109 high-fidelity simulation studies concluded *authenticity* serves as a source of concurrent validity, where the "ability on simulator transfers to real patient" (p. 22). Finally, the degree of authenticity perceived in medical simulations appears to have a positive effect on patient outcomes, which reflects on the quality of educational outcomes. For instance, according to McGaghie et al. (2011), "a growing body of evidence shows that clinical skills acquired in medical simulation laboratory settings transfer directly to improved patient care practices and better patient outcomes. Examples of improved patient care practices linked directly to SBME include studies of better management of difficult obstetrical deliveries (e.g., shoulder dystocia), laparoscopic surgery, and bronchoscopy. Better patient outcomes linked directly to SBME have been reported in several studies using historical control groups that address reductions in catheter-related bloodstream infections and postpartum outcomes (e.g., brachial palsy injury, neonatal hypoxic–ischemic encephalopathy) among newborn infants" (p. 708). Thus, skills learned from VPs that facilitate the students' perceptions of *authenticity* by using realistic cases that reflect the real clinical practice are expected to transfer to other clinical situations.

Models and theoretical frameworks of clinical reasoning also highlight the correlation between *authenticity* and the development and transfer of clinical reasoning expertise using VPs. For instance, Gruppen and Frohna's (2002) model for clinical reasoning posited that as the learner evaluates the patient situation and compares it to previous clinical experiences, students start generating detailed descriptions of the patient's conditions. According to Huwendiek et al. (2014), generating detailed descriptions of the patient situation, characteristics, and context behind clinical symptoms forms the basis for establishing the *authenticity of the patient encounter and the consultation.* Gruppen and Frohna (2002) also considered generating detailed descriptions essential for the development of clinical reasoning expertise.

In summary, the correlation between *authenticity of the patient encounter and the consultation* and the learning and transfer of skills learned from NERVE observed in this dissertation is well grounded in theory, research and practice.

Relationship between Cognitive Strategies, Learning, and Transfer

This section discusses *cognitive strategies in the consultation* in relation to student learning and transfer to include: (a) NERVE features that promoted the students' effective use of *cognitive strategies in the consultation* and how NERVE could be improved, and (b) relationship between *cognitive strategies* and the learning and transfer of clinical reasoning skills.

<u>NERVE Features that Promote Cognitive Strategies</u>

NERVE was designed with a series of features grounded on research and practice that facilitated students' employment of effective *cognitive strategies in the consultation.* For instance, in the NERVE Learning Center (LC), students can learn about and take quizzes to assess their knowledge and mastery of interconnected facts and skills. The LC provides information about the anatomy, physiology, symptoms and pathology associated with 12 CN palsies. The LC also provides information about the corresponding array of patient examination tools doctors normally use to assess CN conditions. Learning outcomes are provided to emphasize key knowledge and skills students should focus on. Feedback is also provided immediately after completing a quiz to include what answers were incorrect and why but the right answer is not provided. In addition, immediate feedback is provided in the NERVE exam room as students discover the facts most relevant to diagnosing the student condition. This type of immediate, tailored, specific feedback allows students to constantly reevaluate their mental representations of the patient's condition, (Figure 26).

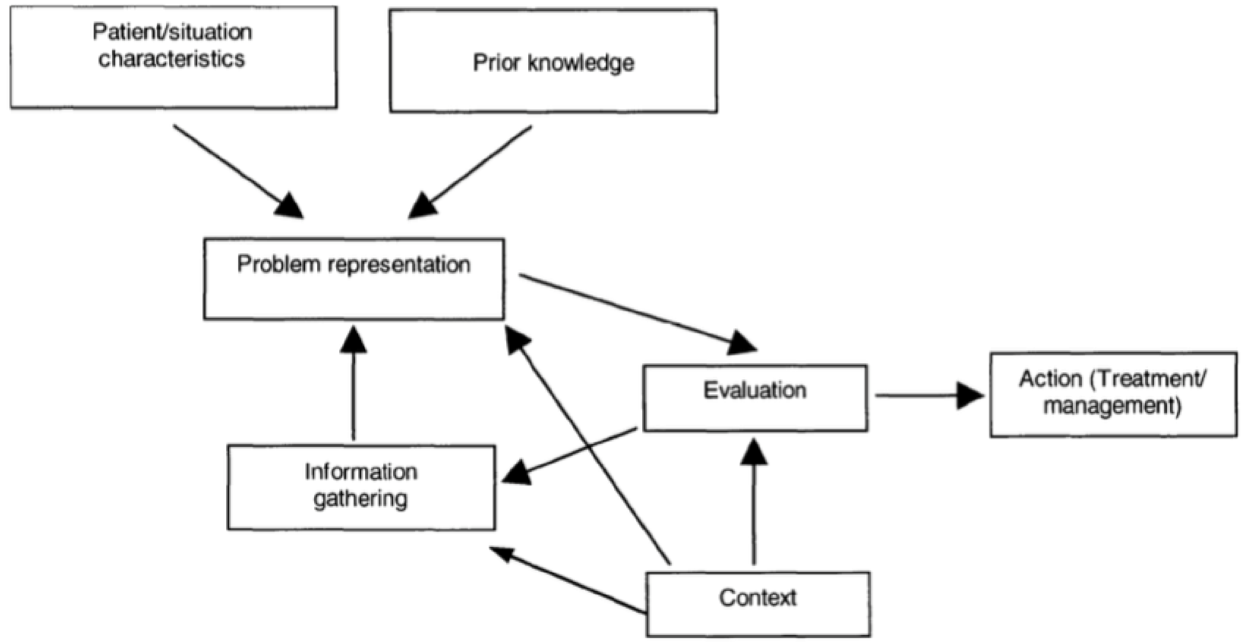


Figure 26.A model of clinical reasoning[a]

Note.[a]Adapted from: Gruppen, L. D., & Frohna, A. Z. (2002). Clinical reasoning *International handbook of research in medical education* (pp. 205-230): Springer.

The interplay between evaluating mental representations of the patient's problem is also facilitated in NERVE by the presentation of interactive CN animations and high-definition graphical depictions of CN anatomy, physiology, and pathology. Research indicates that, "apparently, students also thought the visual elements presented in NERVE were particularly useful, such as labeled diagrams, simulated manifestations of cranial nerve palsies, and interactive animations of eye movements" (Hirumi, et al., 2016, p. 15). Altogether, the learning objectives, quizzes, VP cases, feedback, medical instruments and multimedia tools in NERVE facilitated the students' employment of effective *cognitive strategies in the consultation* but there is room for improvement.

NERVE could be improved by adding prompt questions at certain stages during the consultation where students are asked for: (a) single-sentence summaries of patient problems in abstract terms, (b) the discriminating features of a set of diagnostic hypotheses, and (c) comparing and contrasting diagnostic hypotheses, as recommended by Bowen (2006). In addition, NERVE could be enhanced by applying some of Huwendiek et al.'s (2009a) VP design principles: (a) containing questions and explanations tailored to the clinical reasoning process, and (b) offer recapitulation of key learning points. NERVE provides questions and explanations while taking quizzes but not during the examination. Moreover, while NERVE provides feedback after the diagnosis phase, it does not offer a recapitulation of key learning points related to a specific VP case. Perhaps presenting a brief video at the end of each case where a physician succinctly goes over the symptoms, key patterns identified, key discoveries made, diagnosis and recommended treatment may help students internalize better the experience, promoting retention and transfer.

Cognitive Strategies, Learning, and Transfer

Research suggests there is a correlation between *cognitive strategies in the consultation* and skills learning and transfer from VPs. For instance, the VP design principles that appear most relevant to facilitate the employment of *cognitive strategies in the consultation* are: (a) relevancy, (b) appropriate level of difficulty, (c) offering specific feedback, (d) helping students focus on relevant learning points, (e) recapitulating key learning points, and (f) providing questions and explanations tailored to the clinical reasoning process (Huwendiek et al., 2009a). In particular, *helping the student focus on key learning points* trains the learner to focus on the aspects of patient history, symptoms, context behind the symptoms, diagnostic results, and response to treatment that are significant to reach a diagnosis. Similarly, offering specific feedback and providing questions with explanations helps students follow analogical and analytical reasoning processes to develop clinical reasoning expertise (Croskerry, 2009).

According to Issenberg et al.'s (2005) the characteristics of high-fidelity medical simulations (such as VPs) that appear most relevant to facilitate effective employment of *cognitive strategies in the consultation* are: (a) providing feedback, (b) repetitive practice, (c) range of difficulty level, (d) capture clinical variation, (e) individualized learning, and (f) learning outcomes. Others studies have supported these findings. Norman (2005), who conducted a systematic review of the literature on clinical reasoning studies published during the previous 30 years found that the development of clinical reasoning expertise requires "deliberate practice with multiple examples and feedback, both to facilitate effective transfer of basic concepts and to ensure an adequate experiential knowledge base" (p. 425). Consorti et al. (2012) indicated "clinical reasoning is … strongly dependent on the mental database of cases owned by healthcare professionals" (p. 1006) and that, "if the assessment of clinical skills after the exposure to an even limited set of cases on a specific topic is highly consistent with the training cases themselves, we may expect a good performance" (P. 1006). Findings highlighted the importance of clinical variation in promoting schema creation, activation, and modification in clinical reasoning learning as posited by Eva (2005). The importance of providing clinical variation cannot be overstated. Researchers suggest VP simulations should be designed to provide clinical variation through implementation across "all major clinical specialties" and should include "frequent diseases" and "topics not included in the study plan" (Botezatu et al., 2010a, p. 4). Botezatu et al. (2010a) thus stressed *clinical variation* as an important design element of VP simulations. Similarly, Cook et al. (2010) found that providing "enhanced feedback, and explicitly contrasting cases improved learning outcomes in randomized trials, with ESs ranging 0.29 to 1.47" (p. 1599). Findings underscored the importance of contrasting cases and providing feedback in eliciting the use of effective *cognitive strategies in the consultation* for the development of clinical reasoning skills. Other approaches include leveraging analogical and analytical approaches for developing clinical reasoning expertise. Gawronski and Creighton (2013), for example, discussed several dual process theories, “that are assumed to underlie intuition versus reasoning” (p. 297). Similarly, Croskerry (2009) proposed a universal model for clinical reasoning based on pattern recognition and dual-process theory, which is centered on using both analogical and analytical reasoning to develop clinical reasoning expertise. Moreover, Eva (2005) proposed a combined model for clinical reasoning that illustrates the interplay between the patient symptoms, evolving mental representations of the patient’s condition, and hypotheses being tested, requiring learners to employ diverse analogical and analytical strategies to reach a diagnosis. Altogether, research suggests there is a relationship between employing effective *cognitive strategies in the consultation* and the learning and transfer of clinical reasoning and decision-making skills.

Other models and theoretical frameworks of clinical reasoning underscore the correlation between *cognitive strategies during the consultation* and the development of clinical reasoning expertise using VPs. The elements of Bowen’s (2006) instructional strategies that appear to be most relevant to enable effective employment of *cognitive strategies in the consultation* are that teachers asks students for: (a) single-sentence summaries of patient problems in abstract terms, (b) the discriminating features of a set of diagnostic hypotheses, (c) comparing and contrasting diagnostic hypotheses, and (d) demonstrating presentations of diagnostic hypotheses and the relative probability of different diagnoses. Gruppen and Frohna’s (2002) model of clinical reasoning (Figure 26) appears to be very similar to Bowen’s (2006) educational strategies in eliciting the use of *cognitive strategies in the consultation*. For instance, Bowen’s (2006) recommendations require students to go through steps three to five of Gruppen and Frohna’s (2002) model: (a) information gathering, (b) problem representation, and (c) evaluation. Students then employ *cognitive strategies in the consultation* by making continuous comparisons between existing mental representations of the patient's problems and the actual patient condition. As a result, students make informed decisions regarding the type of information needed to effectively diagnose the patient's condition (Figure 26).

Research also suggests there is a correlation between *cognitive strategies in the consultation* and the transfer of skills learned from VPs to other clinical situations. For instance, research suggests case-based reasoning and problem solving promotes metacognition and conceptual knowledge (i.e., skills transfer) through deliberate practice with multiple examples (Adkins, 2005; Kolodner et al., 2003; Norman, 2005). This is important since working with VP cases is a form of case-based reasoning. According to Salminen et al. (2014), "for a student's ability to apply concepts to solve new problems, active learning with multiple examples can have major effects" (P. 7). Moreover, Posel et al. (2014) indicated, "errors made during the course of a case should be aligned with consequences. VP case authors can create several different endings depending upon the choices made by the learners. Each ending represents teaching moments that provide learners with clear and easily transferable exemplars to improve performance in real practice" (p. 4). Issenberg et al. (2005) found that repetitive practice was the primary factor in 43 studies showing skills transferring to real patients. According to Issenberg et al. (2005), "outcomes of repetitive practice include skill acquisition in shorter time periods than exposure to routine ward work and transfer of skilled behavior from simulator settings to patient care settings" (p. 23). Repetitive practice with multiple examples helps students develop a growing database of mental cases that can be retrieved upon demand depending on the situation. Moreover, the types of problems should be varied to enhance transfer of knowledge to new problems (Huwendiek et al., 2013). Once the student has developed a problem representation (i.e., steps 3-5 in Gruppen and Frohna's (2002) model of clinical reasoning), the sixth and last step in Gruppen and Frohna's (2002) model requires learners to *compare to prior knowledge to reach diagnosis or returning to prior steps* if no matches are made. The model is iterative so various cycles may be needed to reach a successful diagnosis. Thus, repetitive practice with multiple examples is intrinsically related to developing the *prior knowledge* needed to effectively employ *cognitive strategies in the consultation*. In addition, finding a match between prior knowledge and the learner's representation of the patient's condition activates the faster analogical thinking pathway that helps students find patterns between contrasted cases to reach a successful diagnosis (Croskerry, 2009). Otherwise, if no match is found, then the learner activates the analytical pathway to clinical reasoning. Alternating between analogical and analytical thinking to find a diagnosis is a *cognitive strategy in the consultation* that has been related to the development of clinical reasoning expertise (Croskerry, 2009). Students then index the new case into their mental database of cases and transfer the skills learned with the VP once presented with a similar case during their clinical rotation. According to McGaghie et al. (2011), "a growing body of evidence shows that clinical skills acquired in medical simulation laboratory settings transfer directly to improved patient care practices and better patient outcomes" (p. 708). Thus, skills learned from VPs that facilitate the effective use of *cognitive strategies in the consultation* through the provision of repetitive practice with multiple examples and feedback are expected to transfer to other clinical situations.

In summary, the correlation between *cognitive strategies in the consultation* and the learning and transfer of skills learned from NERVE observed in this dissertation is well grounded in theory, research and practice.

Relationship between Coaching, Learning, and Transfer

This section discusses *coaching during the consultation* in relation to student learning and transfer to include: (a) NERVE features that promoted effective *coaching during the consultation* and how NERVE could be improved, and (b) relationship between *coaching* and the learning and transfer of clinical reasoning skills.

NERVE Features that Promote Coaching

Most of the NERVE features that facilitate effective *coaching during the consultation* such as the provision of feedback after quizzes, and during and after VP cases have already been covered in the prior section. However, some other features worth mentioning here are the color-coded feedback schemes and a progress meter (see Figure 10). The color-coded feedback schemes make distinctions regarding the importance of discoveries made during the consultation. These design features replace the teacher in that it: (a) helps students identify important discoveries versus distractors, (b) provide reassurance that students are on the right path, and (c) gives them an idea of how far they are from reaching a successful differential diagnosis. However, NERVE goes beyond the teacher because it is available around-the-clock so students do not have to compete for attention or availability. Another important feature of *coaching during the consultation* in NERVE is the presence of learning outcomes and how the educational experience in NERVE was modeled after the universal principles of experiential learning: (a) framing the experience, (b) activating the experience, and (c) reflecting on the experience (Lindsey & Berger, 2009). Moreover, in NERVE, students are coached through the feedback they receive as they make important discoveries about the patient's condition and after the consultation. There are also a series of videos illustrating how to use the main features of NERVE to include the learning center, the exam room, and the tools used to perform the physical examination, among others. Altogether, these NERVE characteristics facilitated effective *coaching during the consultation* but there is room for improvement.

Much of the students' critical reflection happened during the AAR. During the AAR, faculty conducted a guided-reflective session geared towards strengthening students' understanding of CN palsies. Faculty also clarified misconceptions regarding the diagnosis of CN palsies presented through the VP cases in NERVE. Although the AAR occurred before students applied the skills they learned from NERVE, NERVE was not designed to elicit a similar process of reflection (compared to the AAR) after each clinical case. Thus, including additional components to promote critical reflection during and after the consultation could enhance NERVE's potential.

Coaching, Learning, and Transfer

Research supports the importance of providing effective *coaching during the consultation* to promote development of clinical reasoning expertise. For instance, research suggests that the provision of *feedback* is one of the key characteristics of high-fidelity medical simulations that lead to effective learning (Issenberg et al., 2005). Similarly, Cook et al. (2010) found that providing enhanced feedback and advance organizers, "improved learning outcomes in randomized trials, with ESs ranging 0.29 to 1.47" (p. 1599). Norman (2005) found that *feedback* is required for the development of clinical reasoning expertise. Moreover, *feedback* should be specific, immediate, and continuous (Huwendiek et al., 2009a; Posel et al., 2014; Salminen et al., 2014). Another important aspect of *coaching during the consultation* is facilitating reflection. Recognizing the importance of providing feedback, mentorship, and specific guidance, some scholars and medical educators have made a series of recommendations to improve the quality of coaching in medical education. For instance, Eva et al. (1998) recommended educators become active participants, providing *immediate feedback and guidance* through the range of meaningful comparisons that could potentially be drawn. Moreover, Eva (2005)'s teaching strategies for the development of clinical reasoning expertise include directing educators to provide *structure and guidance* as students consider clinical problems. Another important aspect of *coaching during the consultation* is facilitating reflection. Wilson (2012), for example, found VPs encourage reflection through the provision of feedback. According to Wilson (2012), "providing feedback supports the reflective process. Reflective elements built into the VP that prompted reflection were: allowing the user to review choices and providing feedback. Feedback allowed the user to consider the consequences of their actions. Guided reflection by a mentor is particularly important in professional practice. The mentor facilitates reflection, while supporting and challenging the learner's beliefs and assumptions" (p. 121). Gormley et al. (2011) asked students to reflect about what they learned and how it will affect their clinical practice in the future, which also emphasized the importance of *facilitating reflection* in *coaching during the consultation*. However, research suggests that coaching is not exclusive to teachers but could also involve peers as mentors in order to facilitate learning. For instance, Wilson (2012) found that "learning was viewed as a social activity with the opportunity to talk and share diverse perspectives. There was general consensus that there was much to be learned from one's peers" (p. 118). Finally, Cendan and Lok (2012) found that students in small groups learned better from NERVE than individual users. Altogether, research suggests there is a relationship between providing effective *coaching during the consultation* and the development of clinical reasoning skills.

Models, theoretical frameworks of clinical reasoning, and VP design principles also underscore the correlation between *coaching during the consultation* and the development of clinical reasoning expertise using VPs. Kim et al. (2006)'s recommendations for developing teaching cases appear to be very similar to Huwendiek et al.'s (2009a) VP design principles. For instance, Kim et al. (2006) recommends VP cases be *instructional*, which resembles Huwendiek et al.'s (2009a) design principles of *making optimal use of media*, *using questions and explanations*, *offering specific feedback,* and *focusing* on *and summarizing key learning points*. VPs designed based on these principles and recommendations will optimize learning and transfer through enhanced *coaching during the consultation.* Altogether, research suggests that providing specific, immediate feedback, mentoring, working in groups, and facilitating reflection are essential characteristics of effective *coaching during the consultation* and its correlation to improved learning outcomes. Moreover, "a growing body of evidence shows that clinical skills acquired in medical simulation laboratory settings transfer directly to improved patient care practices and better patient outcomes" (McGaghie, et al., 2011, p. 708). However, much of the existing literature on VP design does not specifically address the relationship between *coaching* and skills transfer so more research is needed in this area.

In summary, the correlation between *coaching during the consultation* and the development of clinical reasoning skills using NERVE observed in this dissertation is well grounded in theory, research and practice.

Limitations

Despite the measures taken throughout research design to maximize validity and reliability, there are still some limitations (i.e., threats to internal and external validity) that must be taken into consideration. For instance, the potential for measurement error in outcome measures (i.e., quizzes, VP cases, VP/SP assessment, and SP checklists) may have increased the probability of committing Type I or Type II errors. Moreover, the sample size (n=110) and correlational design may have limited the inferences that can be made in this dissertation regarding generalizability and causation (Shadish, Cook, & Campbell, 2002). Finally, Podsakoff, MacKenzie, Lee, and Podsakoff (2003), who conducted a review of the literature about method biases in behavioral research, found that some of the most common sources of bias in self-report measures are: (a) consistency motif, (b) social desirability, (c) acquiescence, (d) mood, and (e) transient mood, among others. Therefore, there may have been some sources of bias in this experiment related to the participants reporting beliefs using self-report measures.

Implications & Recommendations for Future Research

Potential implications for practitioners (i.e., medical educators, instructional designers) include providing empirical evidence supporting the use and integration of virtual patients in the medical curricula. This dissertation supported prior research that associated VPs to higher learning outcomes. Medical educators should consider using VPs to enhance students' development of clinical reasoning skills. In particular, VPs may be advantageous to expose students to rare conditions like CN palsies they would not have otherwise seen or experience during their clinical rotations. This dissertation also provided empirical evidence supporting the design of VPs based on the design features and principles evaluated by Huwendiek et al.'s (2014) instrument. Findings suggest that VPs whose design is modeled after the design factors evaluated Huwendiek et al.'s (2014) instrument may lead to higher learning and transfer outcomes. Therefore, medical educators are encouraged to use Huwendiek et al.'s (2014) instrument as a tool to evaluate the effectiveness of the design of their VPs. This dissertation also provided empirical evidence supporting the design of the intervention to include the integration of instructional design theory (InterPLAY), the principles of experiential learning and the AAR.

Findings support greater collaboration between medical educators and instructional designers in designing and integrating VPs that are effective for improving learning and transfer outcomes. Researchers are also encouraged to continue enhancing NERVE by applying some of the NERVE-specific design recommendations offered in this dissertation. Finally, it is hoped that increased collaboration, as it relates to VP design and integration, will help bridge the gap between the fields of instructional design and medical education (Reyes, R.J., Hirumi, A., 2016).

Potential implications for researchers include providing empirical evidence supporting the validity and reliability of Huwendiek et al.'s (2014) instrument for evaluating VP design. Future research should examine additional sources of validity for Huwendiek et al.'s (2014) VP design effectiveness instrument using larger samples and from other socio-cultural backgrounds. In addition, this dissertation provided empirical evidence supporting the predictive validity of Huwendiek et al.'s (2014) VP design effectiveness evaluation instrument. Future research should continue to examine the predictive validity of Huwendiek et al.'s (2014) instrument at Level 2 (Learning) and Level 3 (Application) of Kirkpatrick (1975)'s four-level model of training evaluation, whether it is using NERVE or other VPs, and other outcome measures. Finally, further research is recommended to investigate the potential for cause-and-effect relationships between the independent and outcomes variables examined in this dissertation.

Note: The research reported in this dissertation was supported by the National Institutes of Health (NIH) under award number 1R01LM010813-01. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health.

APPENDIX A
VIRTUAL PATIENT SIMULATION (VPS) DESIGN QUESTIONNAIRE

Virtual Patient Simulation (VPs)
Design Questionnaire

Instructions

1. There are 7 statements in this questionnaire. Please think about each statement in relation to your use of NERVE and indicate how much you agree with the statements using the scale provided below. Give the answer that truly applies to you and not what you would like to be true, or what you think others want to hear.

2. Think about each statement by itself and indicate how much you agree with it. Do not be influenced by your answers to other statements.

3. Select the option that best indicates your response.

Scale for Your Responses
1 = Strongly Disagree
2 = Disagree
3 = Neutral
4 = Agree
5 = Strongly Agree

1-------------------------2-------------------------3-------------------------4-------------------------5
Strongly Disagree Disagree Neutral Agree Strongly Agree

1. While working on the cases, I felt I had to make the same decisions a doctor would make in real life.

2. While working on the cases, I felt as if I were the doctor caring for the patients.

3. While working through the cases, I was actively engaged in revising my initial image of the patients' problems as new information became available.

4. While working through the cases, I was actively engaged in thinking about which findings supported or refuted each diagnosis in my differential diagnosis.

5. The questions that were given to me while working through the cases were helpful in enhancing my diagnostic reasoning in these cases.

6. The feedback I received was helpful in enhancing my diagnostic reasoning in the cases.
7. Overall, working through the cases was a worthwhile learning experience.

APPENDIX B

STANDARDIZED PATIENT CHECKLIST (COMPLETED IN EMS)

Table 43

Standardized Patient Checklist (Completed in EMS)

Professionalism and Ability to Foster a Relationship
1. Greeted the patient warmly and verified patient's identity by confirming name and date of birth (address patient as Mrs. /Miss/Mr., smile, shake hands or appropriate touch).
2. Washed hands before patient contact and maintained clean technique throughout encounter.
3. Introduced him/herself to the patient (first and last name, full title e.g. medical student).
4. Explained purpose of encounter and the role of the student within the first 1-2 minutes.
5. Treated the patient with respect.
6. Listened attentively.
7. Demonstrated genuineness, care, concern, empathy.
8. Expressed interest in the patient as a person.
Information Gathering Skills
9. Used open-ended techniques that encouraged the patient to tell story in his/her own words.
10. Explored the patient's perspective of illness or worries/fears about cause(s)/implications.
11. Explored how health issues have affected the patient.
Information Sharing Skills
12. Communicated clearly, avoided medical jargon or explained terms when used.
13. Provided information related to the working diagnosis and/or next steps.
14. Encouraged patient to develop and demonstrate a full and accurate understanding of key messages.
15. Asked if the patient has any other questions or concerns prior to leaving room.

APPENDIX C
INFORMED CONSENT

We would like to request your consent for participation in the research component of the activity (i.e., completion of the demographic survey requiring less than 5 minutes of your time; use of data for analysis and reporting). Your participation will assist us in establishing best practices related to the use of simulation in the medical curriculum.

Your name will not appear on this consent form or on any of the educational instruments associated with this activity. Initially, your responses will be associated with your email address so that Dr. Teresa Johnson is able to merge your responses together across the study activities. Once a complete data set has been assembled by Dr. Johnson, she will remove your email address from the database to de-identify all responses.

Thank you for your consideration--your important contribution to this field of study is greatly appreciated.

Click the "Done" button at the bottom of this page when you have responded to the item below and are ready to proceed to the next page of this instrument.

Student Consent:

- ☐ I **consent** to participate in the research component of the educational activity.
- ☐ I **do not consent** to participate in the research component of the educational activity.

[Students who consent to participate in the research move on to page 2 of the instrument; students who do not consent to participate in the research move on to a thank-you page that terminates their access to the remainder of the instrument.]

Thank-You Page [for participants who **do not consent**]

Thank you for completing the consent form.

APPENDIX D
DEMOGRAPHIC SURVEY

This instrument requests some basic information about you, including select demographic details. Gender and race are being requested specifically to satisfy reporting requirements of the research-funding agency.
Click the "Done" button at the bottom of this page when you have responded to the items below.
What is your gender?

- Female
- Male

What is your age?

- [drop-down menu of ages will be provided to participants]

What is your race? (select all that apply)

- American Indian or Alaska Native
- Asian
- Black or African-American
- Hispanic or Latino
- Native Hawaiian or Other Pacific Islander
- White

Thank-You Page [for participants who **consent**]
Thank you for completing the consent form and demographic survey.

APPENDIX E
VP/SP ASSESSMENT POST-ENCOUNTER NOTE

Please provide a patient assessment according to:

1. Localization of problem
2. Differential diagnosis
3. Evaluation plan
4. Management & counseling plan

APPENDIX F
EXPLANATION OF RESEARCH FORM

Principal Investigator: Juan C. Cendán, M.D.
Co-Investigators: Teresa R. Johnson, Ph.D. and Atsusi Hirumi, Ph.D.

You are being invited to take part in a research study. Whether you take part is up to you.

- The purpose of this study is to evaluate a computer-based simulation developed for practicing the physical examination and diagnosis of patients with cranial nerve palsies, and to improve our understanding of how the system can be most effectively integrated into the medical education curriculum. Thus, this study will examine how you interact with virtual patients.

- If you choose to participate in this study, you will be asked to complete a brief demographic survey. You will also be asked to allow access to your data from the course simulation activity. Specifically, access to the following data from the course simulation activity is requested: performance data from the simulation exercises and your responses to survey items related to your experience with the virtual patient. If you choose not to participate in this study, you will still be asked to complete these materials, with the exception of the brief demographic survey, but your data will be collected only as course feedback and will not be used for research purposes.

- The additional time required for your participation in this study, beyond the scheduled course activities, is estimated to be less than five minutes.

- Regardless of whether or not you choose to participate in this study, your performance on the simulation exercises will not be graded.

You must be 18 years of age or older to take part in this research study.

There are no direct benefits, risks, or compensation to you for participating in the study.

Study contact for questions about the study or to report a problem: If you have questions, concerns, or complaints, contact Juan C. Cendán, M.D., College of Medicine, UCF, (407) 266-1153 or by email at juan.cendan@ucf.edu.

IRB contact about your rights in the study or to report a complaint: Research at the University of Central Florida involving human participants is carried out under the oversight of the Institutional Review Board (UCF IRB). This research has been reviewed and approved by the IRB. For information about the rights of people who take part in research, please contact: Institutional Review Board, University of Central Florida, Office of Research & Commercialization, 12201 Research Parkway, Suite 501, Orlando, FL 32826-3246 or by telephone at (407) 823-2901.

APPENDIX G
INSTITUTIONAL REVIEW BOARD APPROVAL

University of Central Florida

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Exempt Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Juan C. Cendan** and Co-PIs: **Atsusi Hirumi, Teresa R. Johnson**

Date: **January 21, 2015**

Dear Researcher:

On 01/21/2015, the IRB approved the following activity as human participant research that is exempt from regulation:

Type of Review:	Exempt Determination
Project Title:	Investigating the Use of Simulation-Based Training in Cranial Nerve Palsies with Medical Educators and Second-Year Medical Students
Investigator:	Juan C Cendan
IRB Number:	SBE-15-10931
Funding Agency:	National Institutes of Health
Grant Title:	Neurological Exams Teaching and Evaluation Using Virtual Patients
Research ID:	1050133

This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made and there are questions about whether these changes affect the exempt status of the human research, please contact the IRB. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Joanne Muratori

Signature applied by Joanne Muratori on 01/21/2015 03:27:36 PM EST

IRB Manager

REFERENCES

Adams, E., Rodgers, C., Harrington, R., Young, M., & Sieber, V. (2011). How we created virtual patient cases for primary care-based learning. *Medical Teacher, 33*(4), 273-278.

Adkins, J. (2005). Metacognition: Designing for transfer. *Educational Technology Research and Development, 62*(3), 82-99.

Bateman, J., Allen, M., Samani, D., Kidd, J., & Davies, D. (2013). Virtual patient design: exploring what works and why. A grounded theory study. *Medical Education, 47*(6), 595-606.

Bateman, J., Allen, M. E., Kidd, J., Parsons, N., & Davies, D. (2012). Virtual patients design and its effect on clinical reasoning and student experience: a protocol for a randomised factorial multi-centre study. *Bmc Medical Education, 12*(1), 62.

Botezatu, M., Hult, H., & Fors, U. G. (2010a). Virtual patient simulation: what do students make of it? A focus group study. *Bmc Medical Education, 10*(1), 91.

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

Boud, D., & Prosser, M. (2002). Appraising new technologies for learning: A framework for development. *Educational Media International, 39*(3-4), 237-245.

Bowen, J. L. (2006). Educational Strategies to Promote Clinical Diagnostic Reasoning. *N Engl J Med, 355*, 2217-2225.

Bradley, P. (2006). The history of simulation in medical education and possible future directions. *Medical Education, 40*(3), 254-262.

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry, 189*(194), 4-7.

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

Chodos, D., Stroulia, E., Boechler, P., King, S., Kuras, P., Carbonaro, M., & de Jong, E. (2010). *Healthcare education with virtual-world simulations.* Paper presented at the Proceedings of the 2010 ICSE Workshop on Software Engineering in Health Care.

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., Erwin, P. J., & Triola, M. M. (2010). Computerized virtual patients in health professions education: a systematic review and meta-analysis. *Academic Medicine, 85*(10), 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.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *psychometrika, 16*(3), 297-334.

Croskerry, P. (2009). A universal model of diagnostic reasoning. *Academic Medicine, 84*(8), 1022-1028.

Dick, W., Carey, L., & Carey, J. O. (2009). *The systematic design of instruction*. Upper Saddle River, N.J.: Merrill/Pearson.

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.

Ellaway, R. H., & Davies, D. (2011). Design for learning: deconstructing virtual patient activities. *Medical Teacher, 33*(4), 303-310.

Ellaway, R. H., Poulton, T., Fors, U., McGee, J. B., & Albright, S. (2008). Building a virtual patient commons. *Medical Teacher, 30*(2), 170-174.

Eva, K. W. (2005). What every teacher needs to know about clinical reasoning. *Medical Education, 39*(1), 98-106.

Eva, K. W., Neville, A. J., & Norman, G. R. (1998). Exploring the etiology of content specificity: factors influencing analogic transfer and problem solving. *Academic Medicine, 73*(10), S1-5.

Fischer, M. R., 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.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (1993). *How to design and evaluate research in education* (Vol. 7): McGraw-Hill New York.

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research: An introduction*: Longman Publishing.

Gawronski, B., & Creighton, L. A. (2013). Dual-process theories. *The Oxford handbook of social cognition*, 282-312.

Georg, C., & Zary, N. (2014). Web-based virtual patients in nursing education: development and validation of theory-anchored design and activity models. *Journal of Medical Internet Research, 16*(4).

Gormley, G. J., Mcglade, K., Thomson, C., Mcgill, M., & Sun, J. (2011). A virtual surgery in general practice: Evaluation of a novel undergraduate virtual patient learning package. *Medical Teacher, 33*(10), e522-e527.

Gruppen, L. D., & Frohna, A. Z. (2002). Clinical reasoning *International handbook of research in medical education* (pp. 205-230): Springer.

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.

Hirumi, A. (2006). Analysing and designing e-learning interactions. *Interactions in online education: Implications for theory and practice*, 46-71.

Hirumi, A. (2014). *Advancing Virtual Patient Simulations with InterPLAY: Examining the Role of ID Principles and Practices*. Concurrent Session presented the annual Association for Educational Communication and Technology conference, Jacksonville, FL. Nov. 5-9.

Hirumi, A., Johnson, T., Reyes, R., Lok, B., Johnson, K., Bogert, K., … Cendan, J. (in press). Advancing virtual patient simulations through design research and Inter*PLAY*: Part II–Integration and field test. *Educational Technology, Research & Development*.

Hirumi, A., Kleinsmith, A., Johnsen, K., Kubovec, S., Eakins, M., Bogert, K., . . . Cendan, J. (2016). Advancing virtual patient simulations through design research and interPLAY: part I: design and development. *Educational Technology Research and Development*, 1-23. doi:10.1007/s11423-016-9429-6

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal, 6*(1), 1-55.

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

Huwendiek, S., De Leng, B. A., Kononowicz, A. A., Kunzmann, R., Muijtjens, A. M., van der Vleuten, C. P., . . . Dolmans, D. (2014). Exploring the validity and reliability of a questionnaire for evaluating virtual patient design with a special emphasis on fostering clinical reasoning. *Medical Teacher*(0), 1-8.

Huwendiek, S., De Leng, B. A., Zary, N., Fischer, M. R., Ruiz, J. G., & Ellaway, R. (2009b). Towards a typology of virtual patients. *Medical Teacher, 31*(8), 743-748.

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

Issenberg, B. S., McGaghie, W. C., Petrusa, E. R., Lee Gordon, D., & 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.

Jäger, F., Riemer, M., Abendroth, M., Sehner, S., & Harendza, S. (2014). Virtual patients: the influence of case design and teamwork on students' perception and knowledge-a pilot study. *Bmc Medical Education, 14*(1), 137.

Kim, S., Phillips, W. R., Pinsky, L., Brock, D., Phillips, K., & Keary, J. (2006). A conceptual framework for developing teaching cases: A review and synthesis of the literature across disciplines. *Medical Education, 40*(9), 867-876.

Kirkpatrick, D. L. (1975). *Evaluating training programs*: Tata McGraw-Hill Education.

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development* (Vol. 1): Englewood Cliffs, NJ: Prentice-Hall. .

Kolodner, J. L., Gray, J., & Fasse, B. B. (2003). Promoting transfer through case-based reasoning: Rituals and practices in learning by design classrooms. *Cognitive Science Quarterly, 3*(2), 119-170.

Lang, V. J., Kogan, J., Berman, N., & Torre, D. (2013). The Evolving Role of Online Virtual Patients in Internal Medicine Clerkship Education Nationally. *Academic Medicine, 88*(11), 1713-1718.

Lindsey, L., & Berger, N. (2009). Experiential approach to instruction. *Instructional-design theories and models, 3*, 117-142.

McGaghie, W. C., Issenberg, B. S., Cohen, M. E. R., Barsuk, J. H., & Wayne, D. B. (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: journal of the Association of American Medical Colleges, 86*(6), 706.

McGaghie, W. C., Issenberg, B. S., Petrusa, E. R., & Scalese, R. J. (2010). A critical review of simulation-based medical education research: 2003–2009. *Medical Education, 44*(1), 50-63.

Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741.

Moher, D., Cook, D. J., Eastwood, S., Olkin, I., Rennie, D., Stroup, D. F., & Group, Q. (1999). Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *The Lancet, 354*(9193), 1896-1900.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine, 151*(4), 264-269.

Moher, D., Schulz, K. F., & Altman, D. G. (2001). The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *BMC Medical Research Methodology, 1*(1), 2.

Norman, G. R. (2005). Research in clinical reasoning: past history and current trends. *Medical Education, 39*(4), 418-427.

Norman, G. R. (2006). Editorial–Less is more. *Advances in health sciences education, 11*(2), 111-114.

Nunally, J. (1978). Psychometric methods. *McGraw-Hill, New York*.

Ong, L. M., De Haes, J. C., Hoos, A. M., & Lammes, F. B. (1995). Doctor-patient communication: a review of the literature. *Social science & medicine, 40*(7), 903-918.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of applied psychology, 88*(5), 879.

Posel, N., Mcgee, J. B., & Fleiszer, D. M. (2014). Twelve tips to support the development of clinical reasoning skills using virtual patient cases. *Medical Teacher*(0), 1-6.

Poulton, T., Conradi, E., Fors, U., Fischer, M. R., Huwendiek, S., Ellaway, R., . . . Laidler, P. (2007). *eViP: an EC-funded project to create a bank of multilingual, multicultural, virtual Patients.* Paper presented at the AMEE Conference.

Rahm, E., & Do, H. H. (2000). Data cleaning: Problems and current approaches. *IEEE Data Eng. Bull., 23*(4), 3-13.

Reyes, R. J. & Hirumi, A. (2016). Examining The Pedagogical Foundations Of Virtual Patient Simulations Used For Medical Education: A Review of the Research. Presented at the University of Central Florida Graduate Research Forum, Orlando, FL, April 5.

Salas, E., & Gregory, M.E. (2011). Simulation-based training: Beyond the bells and whistles! *CyberTherapy Magazine*, *4*(4), 18-19.

Saleh, N. (2010). The value of virtual patients in medical education. *Annals of Behavioral Science and Medical Education, 16*(2), 29-31.

Salminen, H., Zary, N., Björklund, K., Toth-Pal, E., & Leanderson, C. (2014). Virtual patients in primary care: developing a reusable model that fosters reflective practice and clinical reasoning. *Journal of Medical Internet Research, 16*(1).

Sematech, N. (2006). Engineering statistics handbook. *NIST SEMATECH.*

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Statistical conclusion validity and internal validity. *Experimental and quasi-experimental designs for generalized causal inference*, 45-48.

Shapiro, J., & Hunt, L. (2003). All the world's a stage: the use of theatrical performance in medical education. *Medical Education, 37*(10), 922-927.

Sivo, S. A., Fan, X., Witta, E. L., & Willse, J. T. (2006). The search for" optimal" cutoff properties: Fit index criteria in structural equation modeling. *The Journal of Experimental Education, 74*(3), 267-288.

Szumilas, M. (2010). Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry, 19*(3), 227.

Wilson, J. I. (2012). A qualitative study of postgraduate students' perceptions of virtual patients in emergency medicine tutorials. *International Journal of Medical Education, 3*, 115-123.