7-4-2016

Examinees' Perceptions of the Physical Aspects of the Testing Environment During the National Physical Therapy Examination

Ellen Kroog Donald
University of South Florida, ekwill@fgcu.edu

Examinees' Perceptions of the Physical Aspects of the Testing Environment

During the National Physical Therapist Examination

by

Ellen Kroog Donald

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Curriculum and Instruction
with a concentration in Measurement and Evaluation
Department of Educational and Psychological Studies
College of Education
University of South Florida

Major Professor: Robert Dedrick, Ph.D.
John Ferron, Ph.D.
Jenni Wolgemuth, Ph.D.
Waynne James, Ed.D.

Date of Approval:
July 6, 2016

Keywords: testing environment, human factors, computer-based testing

DEDICATION

It is with genuine gratefulness that I dedicate this work to the faculty, family and friends that endured this process. My sincere gratitude to Dr. Robert Dedrick, who supported me through the whole process of my interview, coursework, and finally, the completion of my dissertation. Our respective families have grown and life has gone on, yet you always showed your support and confidence for me to complete this process. I truly thank you.

Paisleigh, Chandler and Graeme – you have been born, grown up and endured so much in your life during my doctoral years. You are my joy and so much of the reason that I keep pressing on. I love you and I am excited to see where *your* life leads you.

Thank you Mom and Dad for a lifetime of love and support. Those many years of coursework would not have been possible without grandchild "drop offs" halfway to Tampa. A special thanks to my "three mom's" for constantly asking if I was done yet. Finally, my answer is yes. My love and thanks to my husband, Thomas. You have heard about this dissertation for the entirety of our relationship, and I always knew that you were rooting for me. I love you forever. Lastly, I know that it is only by God's grace that everything in my life is possible.

Philippians 4:13: I can do all this through Him who gives me strength.

ACKNOWLEDGMENTS

I want to thank the faculty of the College of Education at University of South Florida for your expertise and the dedication that you have shown to your role as faculty. My particular appreciation is shown to Dr. John Ferron and the other faculty of the Educational Measurement faculty for the rigor of your teaching and the caring you show for your students. Thank you to the staff of the Federation of State Boards of Physical Therapy for your wiliness to provide information as it was needed for this research. Lastly, I appreciate the support provided by staff and faculty at Florida Gulf Coast University for sharing your expertise in the review of the survey used in this research.

TABLE OF CONTENTS

List of Tables

List of Figures

ABSTRACT

Despite the increasing number of individuals taking computer-based tests, little is known about how examinees perceive computer-based testing environments and the extent to which these testing environments are perceived to affect test performance. The purpose of the present study was to assess the testing environment as perceived by individuals taking the National Physical Therapy Examination (NPTE), a high-stakes licensure examination. Perceptions of the testing environments were assessed using an examinee self-report questionnaire. The questionnaire included items that measured individuals' preference and perception of specific characteristics of the environment, along with demographic information and one open-ended item. Questionnaires were distributed by email to the 210 accredited physical therapy programs at the time, encouraging programs to forward the instrument by email to the most recent class of physical therapy graduates. Two hundred and sixteen respondents completed the study, representing 101 testing centers in 31 states.

Data from these 216 examinees were used to answer four research questions. The first research question focused on the examinees' environmental preferences for the NPTE testing environment and the relation between these preferences and examinees' background characteristics (e.g., sex, program GPA, age, online experience, online testing experience, comfort level with online testing, and preferred testing time). A clear preference toward one end of the scale was observed for preferring a quiet room and a desktop area that had a great deal of adjustability.

Examinees' preferences and their demographic characteristics were not strongly related to the seven demographic variables accounting for < 7% of the variability in examinees' environmental preferences.

The second research question used the data from multiple examinees nested within the same testing center to examine the within- and between-center variability in examinees' perceptions of the testing environment and their satisfaction with the environment. Results indicated that the majority of the variance in these variables was within testing centers with average between-center variability equal to .032 for the perception ratings and .078 for the satisfaction ratings. Research questions (RQ) three and four explored whether examinees' background characteristics (RQ 3) and center characteristics (RQ 4) were significantly related to the 12 environmental perception ratings, 12 satisfaction ratings, and two items representing examinees' perceptions of the effect of the testing environment on their performance and the likelihood they would choose the same center again. In terms of examinee characteristics, age, online testing experience, and comfort with online testing were the most consistent predictors of the various examinee ratings. The most consistent predictors for the satisfaction ratings were examinee online test comfort, online test experience, and age. For center characteristics, the newness of the center and the room density of the center were the most consistent predictors of examinee ratings. For satisfaction ratings, the most consistent predictor was the newness of the center. Center newness was significantly related to the outcome variables related to the size, lighting and sound of the center which may reflect changes in building standards and materials.

The results of the study suggest the need for further exploration of the environmental and human factors that may impact individuals taking high stakes examinations in testing centers. Although there may not be an effect on all examinees, there may be subsets of individuals who are more sensitive to the effects of the testing environment on performance. Further exploration of the uniformity of testing environments is also needed to minimize error and maximize potential threats to test security.

CHAPTER ONE:

INTRODUCTION

Licensure and certification examinations are commonly used in health profession fields to measure a minimum standard of knowledge in a specific field of practice. Through the use of these exams, state regulatory bodies can determine, in part, if an individual is qualified to provide health care to the citizens of the state. Licensure examinations vary in format, length, and administration methods. The settings in which these exams are offered vary, although most are now administered in a computer lab environment. Licensure and certification exams by their nature carry high-stakes for the examinee, often serving as the only portal through which an individual must pass to enter a career for which he or she has often dedicated many years of study. Interpretation of test scores and the decisions that may follow have critical and direct consequences on individuals, the educational programs from which they graduate, the profession, and health care in general (Association of Test Publishers, 2002).

Because licensure exams are critically important to individuals, the profession, and for ensuring public safety, considerable effort has been put into the design and development of the exams to ensure high measurement quality. Test developers and publishers generally follow the guidelines set forth by the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 2014), International Guidelines on Computer-Based and Internet-Delivered Testing (ITC) (2005), and the Guidelines for Computer-Based Testing defined by the Association of Test Publishers (ATP) to establish their practices for test administration (Association of Test Publishers, 2002). The ATP guidelines were specifically written to supplement and elaborate on standards unique to high-stakes, computer-based exams. Although the guidelines presented by the Association of Test Publishers are extremely comprehensive for guiding best practice in planning, designing, developing and administering computer-based tests, only one of the 59 guidelines/criteria addresses the testing environment. Criterion 3.2 states:

> Variability across testing environments should not have meaningful impact on test scores. In addition to factors such as test-taker comfort, noise level, amount of workspace, and lighting, appropriate steps should be taken to ensure that the test environments meet the specified hardware and software requirements. (p. 21)

Similarly, Standard 5.4 of the Standards for Educational and Psychological Testing (AERA, 2014) states that "the testing environment should furnish reasonable comfort with minimal distractions" (p. 63). The authors of the standard comment that the physical environment should have minimal noise, avoid extremes in temperature, minimize distractions, provide adequate workspace, and use legible materials. The International Test Commission's guidelines (ITC, 2005) give attention to human factors in screen and software design, but do not address specifically the physical environment of the examinee, except for those with disabilities. The guidelines promote hardware/software features that facilitate participation for individuals with special needs, but warn that these adaptations should meet the individual needs without adversely affecting the examinee's score validity (ITC, 2005).

Although these standards recognize the testing environment as a potential source of test score invalidity, the testing environment has not received the same level of research attention as other factors that may impact test scores (e.g., bias, test anxiety). As a result of the lack of research, very little is known about how much variability there is across testing environments and how the testing environment may contribute to test score invalidity. Much of what is known about testing environments has come from anecdotal accounts from test takers. Test takers have described situations that are filled with visual and auditory distractions and environmental conditions of low/high temperature, poor air quality, physical crowding, and inadequate workspace.

The need for research that systematically examines the testing environments experienced by examinees is particularly critical at this time given a number of current trends in testing. One of these trends is the increased amount of testing for licensing and certification, which has led to more and more testing agencies outsourcing the administration of exams to testing centers managed by for-profit companies. These centers are handling a greater volume of test takers from more diverse professional areas (e.g., education, medicine, business). For example, Prometric™ contracts with many different testing organizations (e.g., American Dental Association, United States Medical Licensing Exam, National Board of Veterinary Medicine) to administer their organizations' exams. At any given time, examinees within a center may be participating in one of many different examinations. This introduces numerous factors that may affect the testing environment. For example, noise level in its basic, ambient form, is a characteristic of an exam facility's physical attributes. On any given day, noise levels may vary depending on the nature of the exams being taken, the number of examinees, or foot traffic during test periods by persons taking other tests at the same time in the same room. Other factors that may affect noise level are the number of writing intensive exams versus multiple choice exams, the variety of short versus long exams being administered, the number of scheduled breaks, and whether paper and pencil tests are being administered in the same space as computer-based tests. Examinee behaviors can also affect the noise level (e.g., fidgeting, shuffling paper). No known assessment data are available to identify the degree of variation between testing centers on these factors. This variation may be a source of score invalidity.

Another trend faced by testing agencies and test centers that has implications for the testing environment is the focus on test security. Testing agencies have implemented sign in/out procedures, photographing, finger printing and test room monitoring processes in order to minimize test security breaches. Test agencies spend monetary and personnel resources to monitor on-line test question sharing and to prosecute those caught sharing information. With the focus on maintaining a secure test, resources and attention must be diverted from other test administration issues. This diversion of resources and attention may have an impact on the level of standardization of test environments across centers.

Test administrators provide accommodation to those with qualifying, documented disabilities and who request a specialized testing environment, including distraction-free environments. It could be questioned if those examinees without documented disability are affected by the physical environment in which high-stakes examinations are offered. It is not known what preferences examinees have for testing environments, nor which components of these environments affect those taking tests.

Purpose of the Study

In view of the importance of the testing environment for test score validity, the purpose of the present study was to assess the testing environment as perceived by individuals taking the National Physical Therapy Examination (NPTE), a high-stakes licensure examination. The NPTE is a five-hour, computer-based multiple choice examination that allows candidates who successfully pass the examination to seek licensure as a physical therapist in the state in which they apply. Typically, 10,000-11,000 persons sit for the NPTE annually, of which approximately 7,000 are U.S.-trained candidates (C. Searcy, personal communication, July 16, 2009).

The testing environments in which the NPTE is administered were assessed using an examinee self-report questionnaire. The questionnaire was developed using multiple frameworks grounded in environmental psychology and education. The field of environmental psychology has a long history and has contributed to understanding how ambient conditions of physical spaces such as temperature, sound, lighting, and air quality can influence the human activities that occur in homes, workplaces, health care settings, schools, and prisons. Both physiologic and psychological preferences for the environment are observed within the research (Banbury & Berry, 1998; Charles & Veitch, 2002). Age and gender differences have been looked at for different sensitivities to environmental factors (Charles & Veitch, 2002; Yildirim, Akalin-Baskaya, & Celebi, 2007). It could be questioned whether certain groups of people are more sensitive to environmental factors than others.

Within the educational theory literature, researchers have discussed the theory of productivity style both in the classroom and in work settings. This theory is based on the premise that productivity improves when "corporate organization and instruction are provided in a manner that capitalizes on each individual's learning strengths" (Gordon, 1996, p. 5). One component of the theory addresses the physical environment and working conditions that maximize individual output. It is believed that individual preferences exist and that, when in preferred environments, productivity is maximized (Price, Dunn, & Dunn, 1991).

Application of the concepts from environmental psychology and educational research was used to investigate the human factors related to the testing environment. In order to understand examinees' perceptions of the environment, individual preferences need to be understood. Therefore, this study investigated both individual preferences for and perceptions of the testing environment.

Research Questions

The research study focused on examinees who had taken the National Physical Therapist Examination (NPTE) and addressed the following research questions:

Research Question 1:

1a. What are the environmental preferences for the NPTE testing environment of examinees who have taken the NPTE?

1b. What is the relationship between examinees' background characteristics (e.g., sex, program GPA, age, online experience, online testing experience, comfort level with online testing, and preferred testing time) and environmental preferences (e.g., room size/layout, climate, lighting, sound, workstation chair, and desktop design)?

Research Question 2:

2a. On the dimensions of the test environment identified (e.g., room size/layout, climate, lighting, sound, workstation chair and desktop design), how much variability of examinees' perceptions of the testing environment exists between testing centers and how much is within testing centers administering the NPTE?

2b. On the dimensions of the test environment identified (e.g., room size/layout, climate, lighting, sound, workstation chair and desktop design), how much variability in the absolute difference scores (i.e., difference in examinees' perceptions and preferences of the testing environment) exists between testing centers and how much is within testing centers administering the NPTE?

2c. How much variability in there in examinees' perceptions that the testing environment had an effect on their performance and their likelihood that they would use the center again exists between testing centers and how much is within testing centers administering the NPTE?

Research Question 3:

3a. What is the relationship between examinee characteristics (sex, program GPA, age, online experience, online testing experience, comfort level with online testing, preferred time) and how examinees perceive dimensions of the test environment (room size/layout, climate, lighting, sound, workstation chair and desktop design)?

3b. What is the relationship between examinee characteristics (sex, program GPA, age, online experience, online testing experience, comfort level with online testing, preferred time) and absolute difference scores of examinees' perceptions and preferences of the testing environment (room size/layout, climate, lighting, sound, workstation chair and desktop design)?

3c. What is the relationship between examinee characteristics (sex, program GPA, age, online experience, online testing experience, comfort level with online testing, preferred time) and examinees' perceptions of the effect the testing environment had on their performance and their likelihood that they would use the center again?

Research Question 4:

4a. What is the relationship between center characteristics (room size, room density, center newness, presence of ambient light, break space, and access to food/drink) and examinees' perceptions of the testing environment (room size/layout, climate, lighting, sound, workstation chair and desktop design)?

4b. What is the relationship between center characteristics (room size, room density, center newness, presence of ambient light, break space, and access to food/drink) and the absolute difference between examinees' preferences and perceptions of the testing environment (room size/layout, climate, lighting, sound, workstation chair and desktop design)?

4c. What is the relationship between center characteristics (room size, room density, center newness, presence of ambient light, break space, and access to food/drink) and examinees' perceptions of the effect the testing environment had on their performance and their likelihood that they would use the center again?

To address these research questions, individuals who had taken the NPTE were surveyed using an on-line survey. The survey measured examinees' preferences for certain characteristics of components of the physical environment, as well as their perception of these components in their most recent test experience. The survey also collected self-reported responses of academic achievement, age, sex, and descriptors of the Physical Therapy (PT) program from which they graduated. Open-ended questions allowed participants to provide descriptions of their experience with the physical environment while taking the NPTE. The study examined the variability in examinees' perception between testing centers and within centers.

Significance of the Study

Although testing has long been an important means of measuring professional competence, the advent of computer-based testing has become increasingly used to standardize and streamline the evaluation of persons wishing to enter or continue practice in certain fields. While much effort has been taken to make test media fair and reflective of the expected common knowledgebase of a competent professional in a given field, far less attention has been given to the environment of testing venues. This may be of critical importance because certain aspects of the test-taking environment may be disadvantageous to achieving optimal performance on an otherwise well-constructed and psychometrically sound test. Given the gatekeeper role of standardized professional credentialing exams, it is important that all aspects of the testing process be examined and evaluated. This study provides insight into the human and environmental factors involved in testing.

Physical therapists are educated in the structural and sensory components of the environment that may enhance or impede patient functional performance. Environmental conditions may be a more sensitive and salient issue for physical therapist candidates and new licensees due to their education. Through the participation of this trained group of professionals, greater insights into the testing environment may be obtained.

Definition of Terms

Several terms must be defined for this study. *Academic achievement* was defined as an individual's performance during the professional phase of their education (PT program GPA) and performance on standardized graduate examinations (verbal and quantitative GRE score). Multiple measures are taken into account to determine program grade point average (GPA) in PT programs (test scores, practical testing, profesional writing). GRE scores reflect a content knowledge and capacity for learning, as well as test taking ability. In addition to the GRE scores being used as a measure of test taking ability, individuals' performance on the NPTE was also used as a measure of test taking ability. Characteristics of the individual (e.g. academic achievement, sex, age, PT program experience) were level-1 variables within the two-level multilevel data structure (i.e., examinees were nested within testing centers).

Terms related to the *physical environment* of the testing center include: room size, room density, room newness, room layout, room light, room sound, workstation area and workstation chair. Since judging the physical dimensions of a room would be difficult for an average person, *room size* was defined as the number of total workstations present in the testing room. *Room density* is a measure of how crowded a room is and was defined by the number of workstations in use at the time of testing. *Room newness* is the degree to which the space has been updated and kept new (clean, fresh). *Room layout* was described by the spaciousness and openness of a testing room. *Room light* included both the brightness and intensity of the lighting. *Room sound* was defined by both the loudness and clarity of the sound within the testing room. The *workstation desktop* includes the desk and computer system that is assigned to a test taker. The workstation desktop was defined by both its size and level of adjustability for the individual test taker. The *workstation chair* was defined as the hardness of the chair's surfaces as well as the amount of adjustability the chair has to fit the test taker's physical dimensions.

In this study, the physical environment was evaluated based on participants' environmental preferences and perceptions. *Environmental preference* is defined as the favoring of certain environmental characteristics over others. An example of this is a person preferring absolute quietness in a space versus another person who prefers low level sounds while concentrating on a task. *Environmental perception* is the recognition and interpretation of sensory stimuli from the actual physical environment that a person experiences. For example, an individual may perceive a workstation desktop to feel small, or a room to feel cold. To analyze the differences between participants' environmental factors and those that they perceived in the center in which they took the NPTE, difference scores were created. The *difference score* demonstrated the degree to which participants preferences for their environment and what they actually experienced differed. Since the direction of this difference (- or +) did not have relevance, and *absolute difference score* was generated.

Limitations

Obstacles exist to directly acquiring information about the test centers from the agencies who administer commercially developed tests and the testing centers themselves. Information sharing has the potential to expose an exam to potential security breach, which can affect the exam's integrity. Large amounts of variability in the environmental conditions between testing centers would raise questions about the validity of the test scores emerging from these centers. In-depth study of test centers may highlight non-standardization of the facilities, which may raise concern by the agencies that contract with test centers such as Prometric™ and open test centers up for potential legal issues. Ultimately this might have an impact on test center profit viability.

Since data cannot be collected from the testing agencies and centers through direct observation or assessment of the testing centers, an indirect method using examinees' perceptions of the centers must be used. The use of a survey to measure examinees' perceptions was employed. To survey a representative sample of examinees, surveys must be broadly distributed. A national distribution to all physical therapist education programs in the United States was needed since recruitment of study participants cannot be accomplished through test registration or licensure records. Participant recruitment was dependent on physical therapy program directors forwarding research study materials to recent alumni, and in turn, alumni who were recent examinees willing to volunteer to participate. This dependency on program directors and then on the willingness of alumni to participate was a study limitation. Too few respondents taking the exam at the same location gave a small cell size for a given test center and decreased the reliability of the center's rating.

Some center-level (Level-2) variables that were studied may have been subject to variability based on the particular day that an exam was taken. For instance, noise levels may have fluctuated depending on how many individuals were scheduled to take an exam in the center on a given day. Certain participants may have taken a test during a day where the center was highly populated, while others may have taken the test on a low-volume day. Also, during a testing period, other test takers for other exams may have started and stopped at varying times. Since each participant in the study took the exam in a specific center at some point over the period of several months, these day-to-day fluctuations in noise level should have averaged out over time. To examine this potential fluctuation, a measure of the variability within the center was computed. To be aware of fluctuations in the number of examinees present during each participant's exam period, the survey included an item asking for an estimate of the number of workstations in the testing room, as well as the percentage of workstations that were occupied at the time of the exam. Another variable that may have demonstrated fluctuation was lighting if the center had external windows. Ambient light in centers with external windows may have changed depending on the weather on a given day. Although centers are generally found in climate-controlled buildings, small fluctuations in room temperature were possible. All other variables of interest related to the physical environment were stable. These include the room size, workstation desktop size and design, workstation desktop adjustability, chair firmness and adjustability, workstation density, center newness, level of visual distractions, presence of windows and break space, and access to food/drink.

CHAPTER TWO:

REVIEW OF THE LITERATURE

This chapter provides a review of the literature related to the development of licensure exams, the National Physical Therapy Examination, the unique aspects of computer-based testing, the interaction of examinees with their physical environment, the Environmental Load Approach theory, and the physical components of the test environment. Review of the literature related to the development of licensure examinations provides a foundation on which issues related to the administration of these tests can be explored. Once accepted guidelines of test development are reviewed, a review of the literature related to the National Physical Therapist Examination provides an overview of what is known about this examination and those who take the NPTE. Since computer-based tests have test administration issues unique to this mode of testing, the literature related to computer-based test administration is explored. To best address the research questions, a separate body of literature related to the physical environment must be included. Research studies that explore the interaction between humans and their physical environment are reviewed. Discussion of literature related to the Environmental Load Theory provides structure to the exploration of test takers' preferences and perception of their physical environment. Lastly, literature that discusses components of the physical environment found in computer-based testing environment assists in defining the specific areas of interest in this study.

Test Design and Development of Licensure Exams

The goal for test developers is to create a test from which the interpretations and inferences made from the test scores are valid. By defining the purpose, clearly documenting the test specifications, thoroughly assessing the content domain, and constructing test items that meet those specifications, the test developer is much more likely to develop an instrument that produces valid scores.

Test developers and publishers generally follow the guidelines set forth by the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014). These standards outline a process for developing a test that is grounded in the stated purpose of the test. By first identifying the test's purpose and defining the content domain that it will measure, test developers can then move through the process of development and evaluation with a clear focus. The next step is to define the test framework, or content outline. The framework outlines the components of the domain that the test will measure. For licensure exams, this is often done through a role delineation study or practice analysis (job task analysis). Once a framework is established--or at times simultaneously--the test specifications are developed that define how the aspects of the domain being tested will be measured. Test specifications describe the item type, response format, length, scoring procedures, and desired psychometric properties of items. They also define the overall test difficulty, reliability, procedures for how the test will be administered, as well as the population of test takers for which the test is intended (AERA, APA, & NCME, 2014).

Once test specifications are clearly defined, the process of developing and evaluating items begins. This participatory process should include a representative group of experts from the profession for which the exam is intended, including a group or groups external to the item developers to provide external review. For computer-based tests, it is recommended that the individuals who are involved in the item development process are familiarized with the software format in which the items will be displayed to the examinee (ATP, 2002). Items must pass through a process of evaluation, which should include pre-testing to determine if the items meet the planned psychometric properties detailed in the test specifications (AERA, APA, & NCME, 2014).

The final step in the process of exam development includes the construction and evaluation of the exam/exam form. During this step, it must be determined if the test specifications have been met through the use of internal and external review of the exam form. In addition, when multiple test forms are created, the equating of test forms must be completed.

Since the inferences and decisions made from a licensure exam are of high importance, strict adherence to this process of development and evaluation is essential. Although such a rigorous process addresses the appropriate development of an examination, it stops short of addressing the potential sources of invalidity of test scores that may be present in the administration of computer-based examinations.

Consistency across computer-based test administration sites is critical to the maintenance of test score validity (Parshall, Spray, Kalohn, & Davey, 2002). Sources of invalidity in the test administration of computer-based tests are often present because of exam administration costs, maximizing test security, and examinee issues. For example, differences in software and hardware platforms may create inconsistencies in test administration between testing centers, and therefore may have an effect on test score validity. Another example would be the consistency in which test security measures are implemented. Security methods used during the administration of an exam may create inconsistencies in the testing process, and therefore may challenge the validity of test scores. A third example of sources of test score invalidity comes from the examinees themselves. Examinees who lack computer experience may react to more complex, innovative item types differently than those with extensive experience navigating these forms of computer-based items, thus creating a source of measurement error and potentially score invalidity. What is not known is the degree to which the testing environment varies across testing centers and the role that the testing environment has on test score validity.

The National Physical Therapist Examination (NPTE)

There are many licensure examinations in the health professions. The National Physical Therapy Examination (NPTE) is the only licensing examination for physical therapists acknowledged by the state regulatory boards in each of the 50 states. Under the oversight of the Federation of State Boards of Physical Therapy (FSBPT), the NPTE has undergone significant change over the last several years, including a transition from paper-based to computer-based administration. The FSBPT follows the guidelines for test development and administration set forth in the Standards for Educational and Psychological Testing (AERA, 2014). In recent years, the FSBPT has been challenged by issues of test security and a sudden decline in pass-rates. Consequently, physical therapy programs have felt pressure to re-evaluate curriculum, teaching methods, and methods of student assessment to address the increase of failing candidates. Although pass rates have steadily increased, issues of test security have become a growing problem requiring greater attention and resources.

Each exam form, as well as each section, of the NPTE follows a content outline, representing the knowledge areas identified as entry-level to practice as a physical therapist in the United States. This content outline is a result of a practice analysis that occurs every five years. The NPTE consists of 250 multiple choice items, divided into five item blocks. Each item block contains 50 items that represent the content outline of the full exam. Fifty items on each exam are pre-test items. These items are randomly distributed throughout the exam and are not included in the candidate’s score. Five hours are allowed for the examinee to complete the NPTE. Additional time is provided for the pre-exam tutorial, one 15-minute scheduled voluntary break, and a post-exam survey. The 15-minute scheduled break is offered to the examinee after the first two sections of the exam are completed. Three additional unscheduled breaks are offered during the exam; however, the time taken for these breaks is taken from the time one could work on exam questions (Candidate Handbook, FSBPT, 2014).

The NPTE is administered in designated Prometric™ testing centers. Prometric™ centers, which administer exams from a wide variety of test developers and professional organizations, offer the NPTE at approximately 320 centers nationwide. Centers at which the NPTE is offered are located mostly in the cities where a PT/PTA program exists, with some large metropolitan areas having more than one center available to test takers. At the time of this study, there were not standardized hours during which the NPTE was offered. Prometric centers were open during the days and hours "which are appropriate in the area's business environment" (C. Searcy, personal communication, July 16, 2009)., and contracted with FSBPT to meet the scheduling needs of examinees within 30 days of application and within 50 miles of the examinees choice site. Specific information on quality control measures and policies for the testing centers is limited. Prometric™, like other established test administration companies, follows the Guidelines for Computer-Based Testing defined by the Association of Test Publishers (ATP) to establish their practices for test administration (ATP, 2002). These test administration guidelines include specific recommendations for hardware and software requirements, the testing environment, testing interface and function, information for testing personnel, information for the test taker, providing reasonable accommodations, and procedures to address test event irregularities. Standard 3.2 (ATP, 2002) specifically addresses test environment by stating:

> Variability across testing environments should not have meaningful impact on test scores. In addition to factors such as test-taker comfort, noise level, amount of workspace, and lighting, appropriate steps should be taken to ensure that the test environments meet the specified hardware and software requirements. (ATP, 2002, p. 21)

To a degree, Prometric™ centers differ in size, number of computer stations, lighting, break space, noise level, and the amount of workspace for each station. There is no documentation available to the public that describes these components of each center.

Previous research on the NPTE is limited, and has focused mostly on pre-admission factors, academic measures, and clinical performance measures that predict NPTE pass rate (program-level) and examinee performance on the NPTE (examinee-level). Roehrig's (1988) well-cited study was an early look at pre-admission factors and their ability to predict NPTE outcome. This study occurred well before PT education was at the graduate level and the NPTE was computer-based. In her study of the relationship of preadmission factors and success on the NPTE, as well as academic success, Dockter (2001) found only one factor having a significant relationship with NPTE results. First-year program GPA showed a moderate correlation with NPTE performance (r = .648, p < .05). Mohr, Ingram, Hayes, and Du (2005) surveyed program directors of all CAPTE accredited physical therapist education programs in the United States to examine the effect of program-level characteristics on program pass rates for the NPTE. Surveys were received from 132 programs, representing a 75% response rate. Using stepwise regression analysis, the results of this study found that 30.2% of the variance in pass rates was accounted for by: 1) program accreditation status, 2) number of PhD and EdD prepared faculty, and 3) years of pre-professional and professional coursework combined.

Vendrely (2007) studied measures of performance of physical therapy students once in the program and their relationship to success on the NPTE. Vendrely included students' performance in clinical education experiences, their academic (didactic) performance, critical thinking skills and scores on the NPTE. Data from 42 graduates of one physical therapy program were analyzed. Statistically significant relationships were found between the critical thinking scores and success on the NPTE, and between final grade point average in the program and success on the NPTE (Vendrely, 2007). Similarly, Riddle, Utzman, Jewell, Pearson and Kong (2009) found that academic difficulty during a physical therapy education program increased the odds of failing the NPTE by almost six fold when compared to those who had no academic difficulties.

Utzman, Riddle, and Jewel (2007) studied pre-admission measures to determine whether any of these commonly used measures predict whether students fail the NPTE at least one time. The data of 3,585 students admitted to 20 physical therapist education programs were analyzed. Using hierarchical regression models, variables were entered into the model in a predetermined order. Both within-program and between-program analyses were conducted. The results of the study found undergraduate GPA, verbal GRE, and quantitative GRE were predictive of NPTE failure. Race also contributed significantly to the model; however, the contributions of ethnic groups other than white/non-Hispanic were very small. While looking at the addition of personal interviews to aid admissions decisions, researchers found behavioral interviews and the verbal GRE subscale predictive of first-time performance on the NPTE (Hollman, Rindflesch, Youdas, Krause, Hellyer, & Kinlaw, 2008). Data from 89 interviewees and 141 graduates of a PT program were used to analyze which variables used in admissions decisions distinguished graduates who passed and did not pass the exam on the first attempt.

Lastly, non-cognitive variables of graduates of a physical therapy program were explored by Guffey, Farris, Aldridge and Thomas (2002) to evaluate their role in predicting scores on the NPTE. Correlational analysis was used to explore the relationship between the eight domains of the Non-Cognitive Questionnaire-Revised (NCQ-R) and self-reported scores on the NPTE. Regression models were generated and four of the eight domains (long-range goals, leadership, community ties, and academic familiarity) accounted for 21.3% of the variance in NPTE scores. However, some of these domains were related inversely to scores. The authors concluded that although non-cognitive traits may contribute to predicting NPTE success, the NCQ-R may not be helpful in admission decisions in its current form.

Computer-based Testing in Licensing Examinations

The shift in regulatory exams from paper-pencil administered exams to those administered with the use of the computer has brought a mixture of benefits and challenges to both testing organizations and examinees. The benefits include more frequent opportunities for administration, more convenient locations for the examinee, and greater ease in scheduling a test administration time. In general the computer interfaces are easy to use and the environments allow adaptability for those with disabilities (Jones, 2000). When proper procedures are in place, test security can be enhanced while data collection and scoring are simplified. Computer-based test software not only collects data related to examinee responses, but also gathers information about individuals' test taking behaviors and strategies (time on item, item skipping, returning to an item). Test developers may be able to broaden the scope of measurement through testing more types of cognitive processes and skills through the use of innovative item types (Parshall, Spray, Kalohn, & Davey, 2002). This flexibility in item types may allow those developing tests in the health professions to simulate real-life situations within the computer-based examination.

With these benefits come some drawbacks. Initially, there was concern related to the mode effect that computer-based administration of tests may have on examinee performance. Mode effect is defined as the observed performance differences between paper and pencil and computer-based test administrations when other factors are controlled. In particular, researchers have been interested in the differences between paper-based and computer-based tests. Many studies since 1990 have attempted to establish whether a test mode effect is observed. Conclusions drawn from the body of literature related to mode effects are inconclusive. Although mode effect is not consistently observed in the research, it could be questioned whether mode effect could be magnified if the environmental conditions are not ideal. For example, environmental conditions unique to computer-based testing potentially include dimmer lighting to accommodate for monitor use, temperature effects from computer heat, and the need for computer workstation room configuration.

Computer proficiency has been an initial concern for some test takers (Parshall et al., 2002; Wallace & Clariana, 2005). Computer skill level becomes a greater issue when innovative item types are used. Test developers may include innovation through various item formats, response actions, amount of interactivity, methods of scoring, and the inclusion of graphics and other media. In addition, prior experience with computer-based testing can serve as a factor for the individual's level of anxiety. With assessment instruments that utilize simple multiple choice items, the shift to computer-based assessment has had little effect on most examinees. Multiple-choice items that do not require multiple screens or scrolling to be viewed, and do not have complex forms to respond to the options, require little computer skill to answer (Parshall et al., 2002). Mode effects can be addressed by understanding the skill level of the examinees and preparing test takers through thorough instructions and practice items (Parshall et al., 2002).

The cost for computer-based testing has also been a criticism, driving up costs for testing agencies and the examinee. These increased costs are due to increased item development time, the need for more sophisticated testing sites, and software/hardware expenses. Additionally, because of the increased number of test administrations, the item pool must be increased significantly to decrease item exposure. This may be even more of a concern when innovative item types are used. Since much of the focus for computer-based administration has been on maintaining a secure environment, some aspects of the physical environment may need to be less than optimal for the examinee (Jones, 2002; Parshall et al., 2002). Security measures such as video cameras, fingerprinting or retina scanning may create distractions for examinees (Parshall et al., 2002). Concentrated efforts to develop more sophisticated testing sites and security procedures may decrease the focus on quality control monitoring of sites for consistency of the overall environment.

In order to minimize issues related to computer-based test administration, The Guidelines for Computer-based Testing were developed by the Association of Test Publishers (ATP, 2002) to address seven aspects of test administration. First, test agencies and test centers are expected to establish a standard and provide acceptable hardware and software to meet the demands of the test being administered. This allows for appropriate processor speed, visual display, and platform stability for test administration. Second, test takers should be provided the time to become familiar with the test interface and navigation. This is often done through providing access to practice questions just prior to the test commencing. Third, well trained test administration staff should be present to provide a secure test environment, ensure access to technical support, and handle problems as they arise. Fourth, when the test is non-proctored, the test taker must be made aware of the hardware, software, and security requirements, and how to access help if there is a problem. Fifth, reasonable accommodations should be available for examinees who have been approved for accommodations. These accommodations may include a larger monitor for those with visual impairment, extended test time, or modification of the process used by an examinee to respond to test items. This allows for a test experience that is equitable for all examinees. Sixth, procedures must be in place to address and report irregularities of the test event (e.g., computer malfunction) or anomalies such as unusual behaviors by examinees (e.g., making noise during the test session). Lastly, test administrators should ensure that examinees' needs for comfort, light, sound, and space are addressed, allowing for a test environment that has "no meaningful impact on test scores" (p. 21).

Similarly, the International Test Commission's International Guidelines on Computer-based and Internet Delivered Testing (ITC, 2005) include a standard to "consider human factors issues in the presentation of material via computer or the Internet" (p. 4). The standards focus on the importance of screen resolution, color, page design, page colors, text style, prompts, and error message alerts. It is also encouraged in these standards that appropriate levels of control are provided during test administration. Specifically, test publishers are to address the "health and safety" (p. 17) of test users through notifying them of the testing conditions, such as availability of break space. Test users are to be comfortable with the workstation and the worksurface (proper sitting posture, reach the keyboard, sufficient leg room, and room to shift positions during testing). Lastly, these standards stress the importance of ensuring that the facilities meet national health and safety standards for time spent at a computer, and adequate lighting, heating, and ventilation. (ITC, 2005).

Examinees and Their Testing Environment

Through addressing the need for test administrators to examine the test environment, the testing guidelines begin to acknowledge that the testing experience includes the individual's interaction with and reaction to the physical environment in which the test is being administered. In general, individuals are shown to have preferences for the lighting, sound, room temperature and other physical environment factors and seek preferred environments to maximize their confidence and competence (De Young, 1999). Much of the research related to environmental factors has been focused on office environments. Although these studies are not found within the context of testing, they can be generalized, to some extent, to the testing environment. Both work and testing performance requires individuals' attention and concentration, and are conducted in workspaces within a larger room.

The work of Rita and Kenneth Dunn (1978) identified five dimensions that describe the differences in how individuals differ in their learning style. One of the five dimensions identified was the environment in which individuals learn, including the elements of sound, light, temperature and seating design. The Dunn's observed that some may prefer warm environments where others prefer a cooler room, and others prefer more subdued lighting over very brightly lit environments. Based on this early work, Price, Dunn, and Dunn (1991) identified variables that describe preferred components of the learning or work environment for adults. It is theorized that productivity for learning and work improves if adults are provided situations that capitalize on their learning and cognitive strengths (Gordon, 1996; Price, Dunn, & Dunn, 1991). Twenty areas of adult preference for the work/learning environment have been identified, of which nine relate to the physical or temporal aspects of the situation (Price, Dunn, & Dunn, 1991). These nine areas of preference include: sound (level), light (illumination, type of lighting), room warmth (temperature, ability to adapt with clothing, room color), seating design (formal/structured, informal/casual), opportunity for intake (opportunity for breaks with food or at workstation), time of day (evening/morning, late morning, afternoon), and the need for mobility (breaks, ability to move at workstation). Although not specific to the testing environment, these components address individual needs that persons performing a task at a work station may require to work optimally. Because of the similarity between the cognitive demands of work/learning tasks and those needed during testing, it may be critical to understand examinees' reactions to the computer-based testing environment.

Support for the importance of these environmental components and an understanding of the testing environment is also provided by the field of environmental psychology. Environmental Psychology as a discipline seeks to identify the dimensions of environments that influence human activities and then uses this information to match the physical environment to the people using it (Sweet, 1989). Research in the field of environmental psychology examines the "interrelationship between environments and human behavior" (De Young, 1999, p. 1) and includes empirical studies that examine ambient conditions of physical spaces, and variables such as temperature, sound, lighting and air quality (Sundstrom et al., 1996). Theoretical models that are based in the environmental psychology literature may contribute to understanding the effects that testing environments have on adults who take examinations. De Young (1999) discusses how the field of environmental psychology considers the effect that the physical environment may have on humans. First, environmental psychology addresses the issue of how people, both voluntarily and involuntarily, notice their physical environment.

Voluntary observations are those that are mentally acknowledged by an individual; involuntary observations are those that are made in the subconscious. Attention given to voluntary observations of the environment can often be re-directed; however, involuntary observations can serve as a source of distraction. For example, if a person recognizes (voluntary observation) that the sound of a radio is a source of distraction, it can be turned down or off. However, the low noise produced by office equipment may not be recognized, yet still serves as a source of distraction. The latter would be an involuntary observation of the noise in the environment. Secondly, individuals' perception of and attitude toward an environment can be associated with recall of past experiences. Examples of this would be a physical space that is similar to a positive school environment where the person experienced success, or one that is similar to an environment where a person failed a course. Third, people have preferred environments and tend to seek out these environments to feel the most confident and competent. The characteristics of these spaces would be different for each person. Lastly, environmental stressors can be linked to stimulus overload and serve as a cause of attentional fatigue. As with many forms of stress, components of the environment that cause mental or physiological stress detract from a person's performance (De Young, 1999).

Environmental Load Approach

Theorists working in the field of environmental psychology have developed models to explain the observed phenomena between humans and their physical environment. Cohen (1977) described the Environmental Load Approach which is based on four assumptions:

1. Individuals have a limit to their ability to process stimuli and can only focus on a limited number of these stimuli at one time.
2. When environmental stimuli exceed an individual's capacity to attend to each of them, attention is given to the most relevant stimuli. Stimuli that are important to a task are given attention, and those not central to a task are ignored.

3. If a stimulus is at a level of intensity or unpredictable, an individual may need to give it attention to produce an adaptive response.
4. Individuals' level of attention is not constant and can be depleted over time. With this depletion of attention, an individual's capacity for attention may reach an overload, where performance may deteriorate.

Figure 2.1 depicts a Model of Environment-Behavior Relationships, consistent with the Environmental Load Approach, as interpreted by Bell, Fisher, and Loomis (1976). This dynamic process demonstrates the differences in individuals' adaptation to the environment as well as how the objective physical conditions lead to how an individual perceives the environment. This perception can either create a state of homeostasis or stress. If a stressful state is created, the resolution of the stress is based on an individual's ability to cope. The outcome may be one of adjustment or adaptation, or one of increased stress or distraction, which can lead to a decrement of performance. Either of these responses to stress can create a cumulative effect, which may affect individuals' adaptation skills the next time they respond to the physical conditions in which they are placed.

When this theory is applied to an environment in which a cognitive task such as test taking is occurring, you would not expect the stimuli in the environment to be central to the task and would be ignored. However, based on this theory, if the environmental stimulus is great enough or unpredictable, it may cause individuals to divert their attention to a factor in the environment. This would trigger an adaptive response. A response may be to turn attention away from the stimulus. If the stimulus interferes with a person's attention too much, overload may occur. Over time, stimulus overload may increase frustration for a task and create errors in mental functioning (Bell, Fisher, & Loomis, 1978).

Within the testing environment, the most relevant task for an individual is to give his or her attention to the computer-based test items. However, it is not known if environmental conditions can reach a level of distraction that could contribute to the invalidity of an individual's test score, especially as fatigue increases over the period of a long testing session.

Limited research in the area of background noise has been conducted. Errett, Bowden, Choiniere, and Wang (2006) discuss whether individuals become increasingly more aggravated by background noise the longer it persists or whether individuals habituate to the sound in their environment. Their study indicated that the length of exposure to background noise did not have a significant impact on performance on various types of tests, but that the perception on the background noise did impact performance. Participants' scores on math, typing and verbal reasoning tests tended to decrease if the participant was more annoyed by the noise. Bowden and Wang (2005), studying architectural acoustics, found no significant correlations between noise that caused annoyance and productivity on typing and proofreading tasks. They discussed that the small subject population, short exposure time (12 minutes), and minimal changes in performance levels may have contributed to this lack of significance. However, they observed in individual's data, that some subject were more able to "tune out" the noise in the background than others. They questioned whether noise that causes annoyance may have a greater effect on tasks requiring more cognitive thinking than monotonous tasks performed in an office environment. Research that included longer exposure times (60 minutes) demonstrated effects on performance of cognitive tasks and the effects developed over time (Persson Waye et al., 2001). Landstrom (2004) also observed that background noise, ventilation noise in this study, is more easily identified by individuals and therefore has more of an effect on these cognitive tasks in a relatively silent environment.

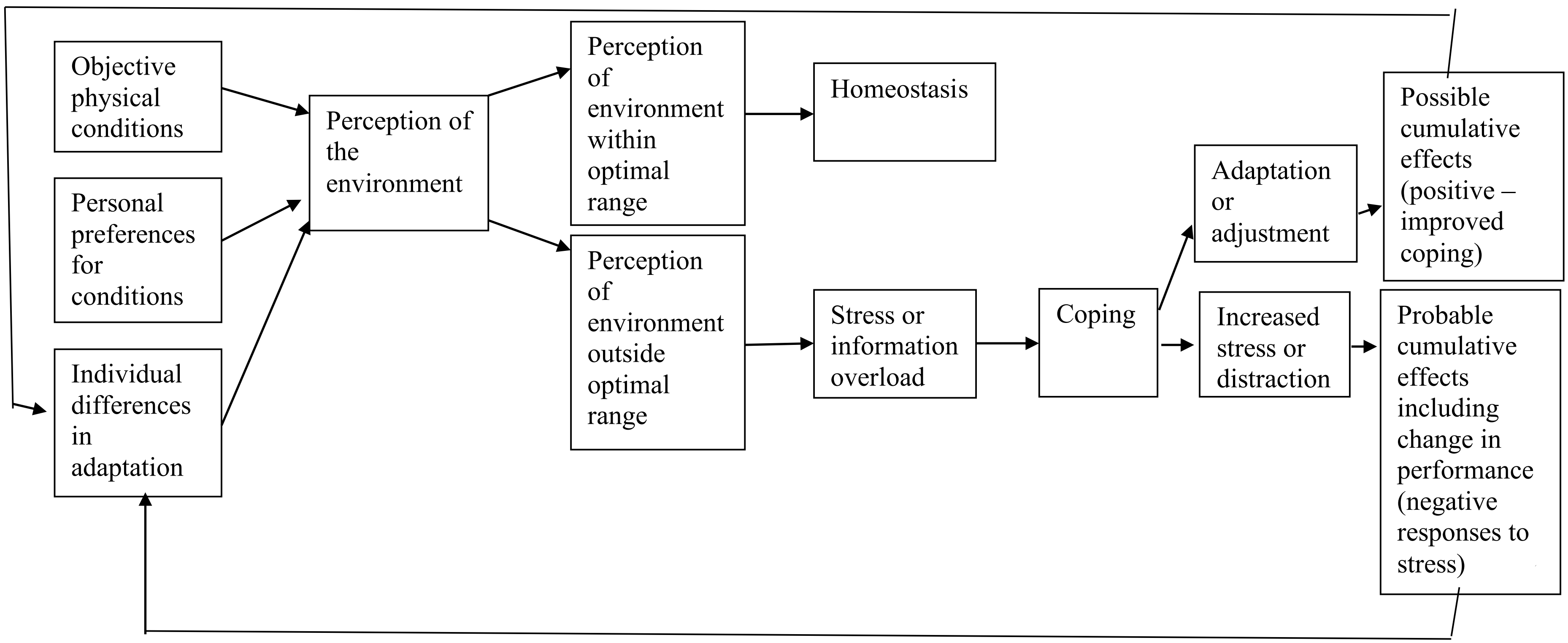

Figure 2.1. Model of Environmental-Behavior Relationships (Adapted from Bell, Fisher, & Loomis, 1976)

Measurement Issues Related to Measuring the Environment

There are many strategies used to measure the physical environment. Light can be measured directly using digital light meters. Sound level is typically measured using digital sound meters. Temperature and humidity are directly measured using a hygro-thermometer (or thermo-hygrometer). However, in certain settings, direct measurement is not possible due to the ethical and practical issues that using these measurement devices may create. Within the testing environment, there are two major barriers to direct measurement of environmental factors. Ethically, there is the risk of creating an interruption or distraction to test takers, disadvantaging them during a high-stakes exam. Practically, testing center personnel restrict access to the test environment due to test security issues, as well as protecting the welfare of the test taker, their customer. Many studies of the environment include self-report measures as a source of data on environmental factors (Fang, Clausen, & Fanger, 1998; Lee & Brand, 2005; Sailer & Hassenzahl, 2000). Unlike interviewing or surveying individuals about their educational or work space, care must be taken when using these self-report measures about the testing environment. Sensitizing participants to the content of an interview or survey prior to the testing event may also create a source of distraction for some test takers. Being more aware of the physical environment may interfere with their testing, as well as potentially hypersensitizing some to components of the environment.

Physical Components of the Test Environment

The evolution of professional standards in testing has broadly addressed the physical environmental needs of examinees. Standards address examinees' need for physical comfort and minimal distractions (AERA, 2014; ATP, 2002; ITC, 2005). These standards address a general need by individuals for a physical environment that does not add distractions and stressors.

In addition to these general guidelines, the role that individual preference has on how people perform must also be considered. The physical environment is composed of not only the material components of the space, but also of individuals' perceptions of those components of the space (Fulton, 1991). Common components that are considered include light, noise level, temperature, and ergonomic design.

Light. Although individual preferences exist for room lighting, there are some common features of light that apply to most people. Well-lit environments enhance an individual's reading performance. A mixture of natural and artificial light balances the positives and negatives of each of the lighting sources. Although brightly lit areas can increase a person's attention, it is important for the light to be high-quality (Bechtel & Churchman, 2002; Sweet, 1989). Space used for reading both horizontal, paper documents and reading from a vertical computer screen may make ideal lighting difficult. Lighting would need to be offered at different angles to the reading surface in order to accommodate the orientation of the hard copy document and electronic document. Preferences for low light are reported when reading from a computer screen, where brighter light is necessary when reading from print documents (Bechetel & Churchman, 2002; Bernecker et al., 1994). Glare can cause eyestrain, headaches, and increased stress (Veitch, 1998). Rea, Oulette, and Kennedy (1995) report that individuals modify their seating position to adapt to situations of improper lighting, resulting in awkward or poor body posture.

Butler and Biner (1987) studied individual preferences of college students for lighting levels during a variety of tasks in a variety of settings. Using a 129-item questionnaire, participants were asked to rate their preference for lighting in 11 settings while performing various activities. These scenarios included activities that were rated in more than one setting (e.g., reading in the bedroom, reading in a library). Ratings were given on a four-point scale where "very dark" was at one pole and "very bright" was at the other. Their study concluded that although there are significant individual preferences for lighting, lighting levels are chosen based on the attention required for a task. These researchers did observe differences between men's and women's preferences for lighting for select tasks (e.g., doing dishes in the kitchen). Tasks requiring reading and concentration were preferred to be done in brighter lit areas and showed little variability in preference. In addition to studying lighting level preference, Butler and Biner looked at the importance of the lighting conditions for each task and how much the individual would like to be able to control these levels. In situations where participants preferred very bright or very dark lighting, the importance and control ratings were high. The importance of the lighting level and control over that lighting level had a strong linear relationship ($r = .94$, $p < .001$). Further research that includes *in situ* evaluation of lighting and preferences was recommended (Butler & Biner, 1987).

A summary of research related to lighting indicates that direct measures of ambient characteristics of work environments may help predict individuals' performance, satisfaction, and psychosocial issues related to work performance (Bechel & Churchman, 2002). Although most of the research has focused on the daily work environment of individuals, there is little known about the environments in which individuals are asked to perform focused, high stress tasks that occur over a prolonged period (e.g., half-day to day-long sessions). Also, individuals in a day-to-day work environment may be more likely, or able, to adapt to their work environments or make small changes to the work environment itself. This is in contrast to persons visiting a testing center, where the environment is both unfamiliar and unchangeable.

Noise. Noise is simply defined as sound that is unwanted. It is not only physically perceived by the ear and higher brain structures but also must be psychologically perceived as an unwanted stimulus (Bell, Fisher, & Loomis, 1978). As is the case with issues related to lighting, no studies of noise within the testing environment and its effects on the test takers' experience appear to have been published. However, concepts can be drawn from the literature on noise in work environments. Sailer and Hassenzahl (2000) acknowledge that although noise levels in office workspaces are rarely detrimental, low level noise present in office spaces may lead to decreased concentration, productivity, and working capacity. However, they believe that individuals have the ability to cope with noise annoyances in the workplace. Coping levels are specific to individuals and may help to moderate the effects of noise on work performance. As noise that cannot be controlled by an individual becomes relevant to him or her, the noise is considered a distraction and cause for stress, and therefore detrimental to performance of a task (Brill et al., 2001; Cohen et al., 1991). The most notable sources of noise annoyance in the office workplace are talking by others, and computer and other office machines (Banbury & Berry, 1998; Sailer & Hassenzahl, 2000).

The effects that noise can have on work performance are based on the nature of the task, characteristics of the noise, and individual differences or preferences (Bechtel & Churchman, 2002). If a task is unfamiliar, noise can detract from a person's performance in a learning task. As the complexity of the task increases, noise has a larger effect on performance (Sweet, 1989).

Unpredictable or high-intensity noise results in greater frustration and decreased performance. Periodic bursts of noise can be associated with greater concentration, leading to improved performance when the task is routine, repetitive, or boring (Bechtel & Churchman, 2002; Sweet, 1989).

Some individuals are more skilled at screening out noise when performing a task (Toplyn, 1988). Highly intelligent people demonstrate decreased performance even on routine tasks when in noisy environments (Sweet, 1989). As in the case with lighting, the effects of noise on performance are diminished if individuals have control over the noise (Cohen et al., 1991).

Ambient noise levels in testing environments, like those in an office environment, are an area of concern. Although talking and telephones would likely not be present in the testing environment, sounds by other test takers as well as ambient sound found in the space may be a source of distraction for some examinees. These sounds may be produced by building machinery (air conditioning, plumbing), computer use, sounds from outside the testing room, and sounds from other test takers. Typically ambient noise is measured using a sound level meter and is measured in decibels (dB). This reflects the physical component of sound/noise. Standards have been set by the American National Standards Institute (ANSI) for acoustic performance criteria in schools (e.g., 35 dB for maximum background noise and 0.6-0.7 seconds for reverberation time); however, no standards are set specifically for testing environments (ANSI, 2002).

Temperature/Climate. The perception of ambient temperature involves both physical and psychological components. Physical components involve the body's ability to regulate differences between the core temperature (typically 98.6 degrees) and the ambient temperature. Ambient temperature is the temperature of the surrounding environment. The psychological component of the perception of temperature is how individuals perceive and respond to this difference in temperature. Temperature can also have an impact on the perception of indoor air quality. Humidity and air flow influences the perception of ambient temperature, and therefore air quality. The quality of room air is perceived to be better when the temperature is held constant in the range of 69-73°F (21-23°C) and humidity is relatively low (30-60% relative humidity) (CCOHS, 2007; USEPA, 2003; Fang, Clausen, & Fanger, 1998). Colder environments may cause discomfort resulting in fidgeting and lack of concentration, as well as reduced manual dexterity and speed. Similarly, overly warm environments may cause a more rapid onset of fatigue or sleepiness and a sense of poor air quality, again resulting in a loss of concentration (Bell, Fisher, & Loomis, 1978; CCOHS, 2007, USEPA, 2003). Inadequate ventilation has also been found to decrease student and teacher performance within a classroom (USEPA, 2003). Organizations such as the Environmental Protection Agency and the Canadian Centre for Occupational Health and Safety offer publications to help guide the management of indoor air quality in a variety of settings (CCHOS, 2007; USEPA, 2003).

Ergonomic design. The ergonomic design of physical space is concerned with the anatomy and physiology of humans, and how humans use components of an environment to complete a task. Since each person is different in their size and shape, it is important to make ergonomic design as individualized as possible to decrease stressors and maximize performance (Smellie, 2003; Zandvliet & Straker, 2001). Within a computer lab or space, multiple components of the work environment are commonly considered. In addition to noise, temperature, humidity, and lighting, the workstation itself must be considered. Components of the workstation that may affect an individual's comfort and productivity include: sitting position, work surface, keyboard and mouse position, computer monitor features, and the ability to move within and away from the space during breaks (Workers Compensation Fund, 2016). In addition to the workstation itself, the room in which a workstation exists can influence an individual's sense of comfort. The amount of space desired by individuals is strongly affected by each person's preferences. A room that feels crowded to one may not to another (Fulton, 1991). Also, factors such as temperature, air flow, odor, and cleanliness can affect an individual's perception of space (Sweet, 1989).

Workstation ergonomics has been an area of focus for health occupations to attempt to avoid injury and overuse conditions. Guidelines and checklists are available from many sources to use in the evaluation of workstations (OSHA, n.d.; UC Irvine, n.d., Workers Compensation Board, 1999). The components of workstations that are of greatest interest are those that cause postural strain, eye strain, and discomfort. Smellie (2003) analyzed a standard computer workstation at which computer use is the primary function. He broke the work area into three areas of fixed contact: the floor, the chair seat, and the desk surface. To accommodate most users, especially extreme user dimensions, at least two of these surfaces must be adjustable. Recommendations were made for providing reasonable adjustments for the workstation chair (seat height, seat depth, seat width, seat surface, backrest, seat angle, armrests) and the workstation desk (leg room, desk surface, desk height, computer screen position, wrist supports). Many of these recommendations can be carried out in a typical workstation with the use of chairs and desks that allow adjustability. Zandvliet and Straker (2001), when studying the use of workstations in schools, noted that not only is there a need for physical comfort and for limiting distractions, but when factors of physical environment are not optimal, satisfaction, learning, and productivity are negatively affected. Components of the environment that were included in these researcher's recommendations were adequate workspace, adjustable chair height, variable screen height, air quality, lighting, and the spatial orientation of the computer workstations.

Multilevel Modeling Analysis

Consideration of the physical environment in which people take exams is naturally suited for multilevel analysis. Examinees take tests within the context of a particular testing center and thus the perceptions of examinees clustered within a testing environment are likely to be statistically dependent and be a function of individual characteristics of the examinee (e.g., age of the examinee) and characteristics of the testing centers (number of work stations). The variability that exists in the physiological responses to the physical environment between testing center and within testing centers is best analyzed using a two-level model of analysis.

Multilevel modeling continues to grow in application, especially in the areas of education, public health and health care. Much of the data collected in the social and health sciences are inherently hierarchical in nature (Paterson & Goldstein, 1991). Often research questions are focused upon the relationship between a set of variables and a particular social, educational, or health outcome. Using aggregate data to predict these outcomes can lead to missing an explanation or cause for a particular individual outcome. Using multilevel analysis allows for the variability at the individual level (or the individual unit of analysis) to be explained, while being able to generalize the findings across the group (or level-2) level of analysis by learning the source of variability at this level (Paterson & Goldstein, 1991). Within the area of test administration there appears to be no studies where multilevel analyses have been conducted. In the field of physical therapy, the application of multilevel analysis has been limited to a few studies focused on the care of back pain, and mostly conducted in the Netherlands (Bekkering et al., April 2005; Bekkering et al., June 2005; Engers et al., 2005; Hendriks et al., 2003; Kerssens et al., 1999; Swinkels et al., 2005).

Summary

Despite all that is known about the importance of the physical environment to individuals' satisfaction, productivity, concentration and endurance for a task, there has been limited research on the testing environment. Although the fields of environmental psychology, public health, and ergonomics have studied the physical environment both at the micro and macro level, there is no apparent literature describing aspects of the environment in testing, and the effects of these components on the test taker.

Across disciplines, the components that are consistently evaluated include lighting, sound/noise, air quality/temperature, space, and ergonomic design. Various studies have attempted to analyze these physical characteristics and the individual and generalized effects on satisfaction, productivity, and psychological and physical stress.

The Model of Environmental-Behavior Relationships (Figure 2.1) provides a loose structure for exploring the dynamic between individual preferences and experiences, and the variation between one testing center and another. Although the effects that the physical environment may have on test administration are complex, beginning to understand these relationships is important to increasing the validity of high-stakes testing.

CHAPTER THREE:

METHODS

This study explored examinees' environmental preferences and perceptions of the environment in which they were administered the NPTE. Data related to examinees' demographic information, academic ability, previous test taking experiences, program characteristics, test administration information, as well as test center characteristics were also collected. These data were collected through the use of an on-line survey instrument that included open and closed-item formats.

This chapter is divided into four sections and describes the development of the survey instrument, data collection, sample, and data analysis. In the first section on instrumentation, the development of the survey instrument, including pilot testing and instrument revision, is discussed. Section two details the process of the distribution and collection of the surveys, and data management. The third section provides a description of the participants as well as the sampling and participant recruitment process. The final section provides an overview of the data analysis plan. Data analysis procedures included both descriptive and multilevel modeling procedures.

Instrumentation

Instrument Design

This study used an on-line survey instrument developed and pilot tested prior to data being collected from a national sample. The instrument was developed using Checkbox® 4.1

software, which is supported by Florida Gulf Coast University Office of Planning and Assessment. A systematic approach to survey instrument development was used that is consistent with the construction of instruments for subject-centered measurement (Crocker & Algina, 2006).

Instrument development was initiated by identifying the primary purpose of the survey. This survey instrument's primary purpose was to collect data on the preferences and perceptions of the environment in which participants were administered the NPTE. It was also used to collect self-report data related to academic achievement, on-line experience, program characteristics, test administration information, test center characteristics, and examinee characteristics.

After identifying the survey's purpose, the constructs that were of interest were analyzed to determine measures and behaviors that may represent each construct. Measures and behaviors related to each construct were identified through literature review, discussion with experts in testing and physical therapy education, and professional experience of critical incidents that characterize extremes within the constructs. This process served to ground the instrument in the realities of practice. The review of the literature included the fields of physical therapy education, environmental psychology, adult education and assessment, and measurement. The major components of the survey included: examinee background characteristics, NPTE testing experience, examinee environmental preferences, examinee perceptions of the testing environment where the NPTE was taken, and testing center characteristics.

Examinee background characteristics. Areas of interest were the examinee characteristics related to academic achievement, time-of-day preference for taking tests, sex, and age. Examinee is defined as an individual who has taken the NPTE at least once. The measures of academic achievement in this study included program GPA, highest verbal GRE score, highest quantitative GRE score, and NPTE score. The score from the analytical portion of the GRE was not included in the survey due to changes in the scoring and format of this portion of the exam in 2002. These demographic and academic achievement variables were considered level-1 (examinee) variables in this multilevel study.

The examinees' background characteristics also included the characteristics of the program from which they graduated. The characteristics of the examinees' PT program curriculum and testing style were relevant to this study since this would reflect the examinees' most recent learning and testing experiences. These program-related background characteristics included the degree level (masters, doctoral), type of curriculum (categories used by the Commission for Accreditation of Physical Therapy Educational Programs), and the amount of coursework and testing that was administered online/computer-based (percentage of coursework/exams delivered on-line). The amount of on-line/computer-based learning and testing opportunities represents the level of experience with computer-based applications. These program-related background characteristics were also considered level-1 variables in this study.

Examinee NPTE experience. Examinees' experience with taking the NPTE was another area of interest. This was defined at the individual level by the number of times examinees took the NPTE, their most recent performance on the NPTE, and the time of day when they took the NPTE. It was also represented by whether or not they used headphones, the number of scheduled and unscheduled breaks they took, and whether they left the testing area for those breaks. Again, the examinee's NPTE experience variables were considered level-1 variables in this study.

Examinee testing environment preferences. Examinees' preferences for the physical environment were constructs of interest. The survey instrument included semantic differential scales that participants used to rate their preferences for the testing environment. The components of the environment that were analyzed included room size/layout, climate, lighting, sound, workstation desktop size, chair firmness, and the adjustability of both the workstation desktop and chair. Two items were included in the survey for each of these environmental components, measuring two separate aspects of each component. Anchor words used on each scale provided opposite ends of a spectrum without judging the response. For instance, when preferences and experience for lighting were scaled, neutral words such as dim and bright were used. It is only personal preference as to what is considered “good” or “best”. Once again, these environmental preference variables were considered level-1 variables in this study.

Examinee testing environment experiences*.* In addition to examinees’ preferences for components of the testing environment, examinees’ center-specific experience was also an area of interest. As with the preference rating scale, the survey instrument included semantic differential scales for participants to rate the testing environment experience for the most recent testing experience based on their perceptions of the environment. The components of the environment, as well as the anchor words for each scale, were identical to those used in the examinee preference survey items. The examinee’s perceptions of the environment in which the NPTE was taken were considered level-1 variables in this study. Examinee perceptions of the environment within a center when aggregated created a new level-2 (center-level) variable.

Center characteristics*.* Lastly, other aspects of the environment of the center were an area of interest. Centers were identified by the city and state of their location, and described by the number of workstations, the number of workstations in use, the presence of exterior windows and natural light, the availability of a break room, access to food and drink, the level of visual distractions present, and the overall newness and cleanliness of the center. These were also considered center-level variables and treated as level-2 variables in this study.

Creation and Review of Items

The survey was constructed using a framework reflecting the constructs of interest as well as the related measures and behaviors. Items were drafted representing each of the areas of the framework. Item formats were determined based on the nature of the question. A mixture of short response, radio-button, and pull-down menu items was constructed.

Items related to the participants' preference and perception of specific characteristics of the environment were included, using a format similar to a semantic differential scale. Semantic differential (SD) scales can be used to map individuals' connotations for a given word or phrase to scale attitudes on a particular concept. In this study, the bipolar adjectives were contrasting and directional, but were not judgmental (e.g., good-bad) as with most semantic differential scales. Preferences for environmental factors are specific to a person (e.g., some like it warmer and some like it colder) with the potential for large amounts of variability. Items were organized so that similar components of the environment (e.g., room temperature) were sequential. Response options were all in the same direction where the low degree of a measure (e.g., cold) was on the left of the scale and the high degree of a measure (e.g., hot) was placed on the right. See Figure 3.1 for a sample of the preference/perception items.

	What is your preference?		What was your actual experience?	
Room	Quiet	Loud	Quiet	Loud
Sound	◦ ◦ ◦ ◦ ◦ ◦ ◦		◦ ◦ ◦ ◦ ◦ ◦ ◦	

Figure 3.1: Sample Semantic Differential Item.

Osgood and colleagues (1957) developed the semantic differential method based on the hypothesis that words include both literal and affective meanings. Bipolar adjectives are used to judge multiple impressions of materials, experiences, and behaviors. Adjectives such as strong-weak, bright-dark, or hard-soft are what Osgood called the potency factor. These *potency* factors are not associated with clear positive or negative effects and are associated with information received through sensory modalities. Instead they differentiate attitudinal intensity based on individuals' subjective response to the connotative meaning of each word. The seven-point scale can be coded identifying a neutral (zero) point and ranging from negative to positive values, or can be coded from one to seven along a continuum (Al-Hindawe, 2006). The assignment of numeric values is typically done after the scale is completed so the respondent focuses on the adjectives and not the numeric value assigned, as seen with Likert scales.

The strength of semantic differential scales is that they assist the measurement of abstract concepts that are not easily measured through other methods. It has also been shown that the use of adjective pairs has demonstrated meaning across various cultures in measuring attitudes and abstract concepts (Heise, 2010; Osgood, May & Miron,1975). Although developed long ago, these scales continue to be used in research in various fields including: speech/language (Altenberg & Ferrand, 2006; Lallh & Rochet, 2000; Swartz, Gabel, & Irani, 2009), education (Christensen & Knezek, 2009; Ribich, Barone, & Agostino, 1998; Zevin & Corbin, 1998), social work (Zugazaga, Surette, Mendez, & Otto, 2006), psychology (Mehrabian & Russell, 1974; Sakuta & Gyoba, 2006; Short & Magana, 2002), medicine/health (Stillman, Braitman, & Grant, 1998; Tracey, Arroll, Richmond, & Barham, 1997), and marketing/advertising (Van Auken, Barry, & Anderson, 1993).

Semantic differential scales have certain limitations. When semantic differential scales are used in research where potentially controversial or socially undesirable topics are being explored, it has been questioned whether respondents may self-sensor and provide only socially desirable responses (Swartz, Gabel, & Irani, 2009). Yu, Albaum, and Swenson (2003) raised the question of whether there was more central tendency errors in the use of semantic differential scales, especially in certain cultural groups. Their findings were inconclusive and not supported in other research. There has also been some debate about the level of measurement that is reflected in these scales. Arguments have been made for these scales to be ordinal measures, while others believe that the neutral center point allows for these to be used as interval scales (Himmelfarb, 1993). Semantic differential scales that are more evaluative in nature are used in research, but are not included in this study.

One open-ended multiple-line text item was included to conclude the survey and to gather more loosely structured, descriptive data. Examinees were asked to describe the room and space in which they took the exam. Examinees were prompted to consider room size, lighting, temperature, sound, workstation characteristics, break space, and furniture and workstation comfort and adjustability. Individuals who took the NPTE more than once were asked to describe each time they took the exam and describe the difference in centers if taken at a different location. Physical therapist program graduates are sensitized to the physical environment by the nature of their education to become a physical therapist. These participants have skill in describing the physical environment as it is a component of practice with their patients.

The initial draft instrument was reviewed and revised by an expert in the field of measurement, a physical therapy educator, and an expert in the use of the survey software. Feedback was provided on survey content, wording of items, item layout, general usability and navigation. After multiple reviews and revisions, a sixth draft instrument was created.

Once the instrument was developed, the Institutional Review Boards (IRB) at the University of South Florida and Florida Gulf Coast University were contacted to gain approval for pilot testing of the survey instrument. Both IRBs notified the researcher that no application was needed to earn approval for this initial instrument testing. Pilot testing was conducted with the sixth version of the instrument.

Twenty graduate students in a physical therapy program, who had recently taken the GRE, were asked to complete the on-line draft survey and participate in interviewing to investigate word meaning for the anchor words used in the differential semantic scales. This population of individuals was identified to participate in the pilot as the GRE is administered in testing centers similar to those being studied and serves as a similar experience to those persons who served as participants in the study. Since this group of students was scheduled to graduate in two years from the time of the study, it was known that they were not to be included in the group of invited participants.

Numeric codes were assigned to each item's options for scoring and data analysis. Reliability estimates were generated from data collected from piloting the instrument with 20 first-year Physical Therapy students. This provided a way to check for correlations between items and to identify whether there was need for modifying the instrument. The amount of correlation between items allowed for decisions to be made about combining items, particularly the semantic differential items. Cronbach's alpha values for the six environmental characteristic preferences and six environmental characteristic perceptions, each measured by two items, were calculated and found to range from .01 to .78. Since only weak to moderate relationships existed between the items for each characteristic, it was decided to treat each of the items in this section of the instrument as separate items.

In addition to completing the survey, the pilot test participants were asked to provide comments and suggestions on wording, readability, options, and to answer the open-ended question that concludes the survey. Comments were used to make decisions on modification of wording items and word choice in the semantic differentials.

Additionally, faculty in physical therapy programs and measurement experts reviewed the draft survey as part of the content validation process. Expert review included the evaluation of item quality, relevance, and completeness. All participants in the pilot testing were asked to review the survey for accuracy, grammar, language bias, readability, and online navigation clarity and ease. Adjectives used within the semantic differential scales were evaluated for meaning and to determine if they were appropriate anchor words for each scale. Anchor words were evaluated to verify if words within each word pair were polar opposites and whether the words were non-evaluative in nature. This comprehensive review of the draft survey served as the content validation process for the instrument. The results of the pilot testing were used to revise the survey instrument. Further evaluation of the anchor words to determine if they were polar opposites and non-evaluative in nature was conducted using doctoral students in the Department of Educational Measurement and Research. Lastly, robust technical testing of the on-line functioning of the survey application was conducted. This testing was conducted by an expert in the use of the survey software and construction of online surveys and included an evaluation of the ease of online navigation, the clarity and format of response options, the length and survey layout, and readability of questions. The results of these evaluations were used to further revise the instrument.

Once complete pilot data were collected for the study, the issue of combining variables was revisited. Observation of mean scores for the 12 preference items and 12 perception items indicated differences between test takers preferences for environmental characteristics and the experiences they had in testing centers.

In summary, Table 3.1 contains the finalized items that represent the environmental constructs that were used as outcome variables.

Table 3.1. Environmental Characteristics

Items	Construct
Perceptions:	
1. Intimate/Spacious	Room size
2. Enclosed/Open	Room size
3. Cold/Hot	Room climate
4. Dry/Humid	Room climate
5. Dim/Bright	Room light
6. Soft/Intense	Room light
7. Quiet/Loud	Room sound
8. Muffled/Clear	Room sound
9. Soft/Hard	Workstation chair
10. Not Adjustable/Highly Adjustable	Workstation chair
11. Small/Large	Workstation desktop
12. Not Adjustable/Highly Adjustable	Workstation desktop
Environmental Effects	
1. Prevented From Performing Best/ Allowed Me to Perform at Best	Environmental Effect on Performance
2. Not Likely/Highly Likely	Environmental Effect on Center Use
Observed Center Characteristics	
1. Number of Workstations	Room Size
2. Number of Workstations in Use	Room Density
3. Degree (rating) of room renovation	Newness of Center
4. Presence of Exterior Windows	Access to Ambient light
5. Presence of Break Space	Access to Other Room for Breaks
6. Access to Food/Drink	Access to Refreshment
7. Level of Visual Distractions	Visual Distraction

The final open ended question provided descriptive narrative, reinforcing the idea that there may be a great deal of variability between testing centers.

Procedures/Data Collection

The development, revision and validation of the survey instrument constituted the first stage of this research. During the second stage of this research, an invitation e-mail and link to the online survey was emailed to all cooperating U.S. Physical Therapy program directors, who then forwarded it to potential participants. E-mail invitations were sent to program directors in November, 2008 with follow up e-mails sent during the months of November and December. In January, 2009 it was determined that the first round of data collection did not represent enough testing centers for the planned analysis. Therefore, a second round of data collection was conducted through individualized addressed e-mail invitations to the program directors who had not replied either that they would participate or were unable to participate. Phone calls were also made to some programs where the e-mail address may have been inaccurate. This second round of data collection was concluded in April, 2009.

Once participants provided consent to their participation, they viewed the first page of the survey. They were directed to proceed through the nine pages of the on-line survey and submit their responses to each item. Responses were electronically transferred into a database for data analysis. Open-ended item responses were collected. Participants were de-identified and provided a participant code for use during data analysis.

Sample

A two-level, multilevel design was used to collect data on survey participants as well as the centers in which they took the NPTE. Level-1 variables consisted of the individual characteristics of participants. Level-2 variables consisted of the characteristics of the testing centers as reported by survey participants.

Survey Participants

Invited participants included persons who had graduated from U.S. physical therapy programs accredited by the Commission for Accreditation of Physical Therapy Education (CAPTE) and who had taken the NPTE in the most recent 12-month period prior to the initiation of the study. This population was used because of its unique sensitivity to issues related to the physical environment. Physical therapist education programs teach students to be aware of physical and sensory components of living spaces (home, work, recreation) and how these components of the environment may enhance or hinder function. This education may have allowed participants in this study to be more aware of components of the environment, as well as their environmental preferences, and provide valuable insight into this topic.

The American Physical Therapy Association (APTA) 2007-08 Fact Sheet of Physical Therapist Education Programs (APTA, 2008) projected that 5715 persons graduated from accredited PT programs in 2007 in the United States. It is typical for physical therapy programs to graduate students in cohort groups. In 2007, the average graduating cohort was 29, with a class range of 0-145. For those who graduated in 2008, 24.7% graduated from a Master's level program, with the remainder (75.3%) graduating with the entry-level degree of Doctor of Physical Therapy (E. Price, personal communication, July 20, 2009). Of those examinees who took the NPTE in 2008 who were graduates from U.S. programs, 21.4% graduated with a master's degree and 76.7% graduated with a clinical doctorate (DPT) (C. Searcy, personal communication, July 16, 2009). Eighty-one percent of physical therapy graduates are non-Hispanic white, with the remaining 19% representing each of the other ethnic groups (American Indian, Asian, African American, Hispanic and other). Seventy-one percent of students enrolled in PT programs are female. The typical physical therapy graduate is 25-45 years of age, with a national mean of enrolled students being 23.8 years (APTA, 2007). However, the age of first-time test takers who graduated from CAPTE-accredited institutions in 2007 or 2008 who sat for the NPTE in 2008 was a mean of 34 years, with a median of 27 years and mode of 26 years (C. Searcy, personal communication, July 16, 2009).

The majority of programs (60.5%) graduate students at the end of the spring semester (late April - early June) each year, with 18% graduating students at the end of the fall semester

(December - January 1) (E. Price, personal communication, July 20, 2007). First time test-takers are most likely to take the NPTE in the months of June, July, and August. In a typical year, 48% of first time test takers sit for the NPTE during these months, with an additional 9.8% taking the exam in September (C. Searcy, personal communication, January 10, 2008).

Participants for the study were recruited through e-mails being sent to all 210 accredited programs that graduate physical therapist candidates in the United States. A current list of all program directors (PD) and related contact information was available electronically through the American Physical Therapy Association, Department of Accreditation at no charge. After IRB approval was received, an e-mail was sent to PDs asking if they were willing to assist with the study. Program directors were asked to respond to the e-mail, notifying the researcher that they were willing to forward the survey to their recent alumni who had graduated from the program in the previous 12 months. Cooperating program directors were sent a second e-mail and asked to forward it to each of their most recent class of graduates to solicit participation in the study. All program directors of CAPTE accredited PT programs were contacted to request their cooperation in forwarding the study survey to recent graduates. Ultimately, 70 program directors (35%) agreed to forward the invitation e-mail and survey link to their recent graduates. The majority of the twenty-five program directors who responded to the requests for assistance but were unable to participate cited the following reasons for not forwarding the survey: they had no graduates during the transitional year between MSPT and DPT; no access to group e-mail for graduates; or they lacked up-to-date e-mails on graduates. Four programs that were CAPTE accredited were located outside United States. Only a few responded that they were too busy, didn't want to bother graduates, or felt their graduates were over surveyed. One program reported that I would need to have IRB approval at their institution to survey their graduates.

The e-mail included a brief invitation to participate in the study and a link to the online survey, which included the Letter of Informed Consent. Recent alumni who selected "agree" after reading the letter of informed consent proceeded to the survey. By selecting “agree” and completing the survey, participants gave their consent to participation. Participants met the following inclusion criteria: (a) Graduate from a physical therapy program in the United States accredited by the Commission for Accreditation of Physical Therapy Education (CAPTE); (b) took the NPTE within the 12-months preceding participation in the study; (c) had current access to the Internet; and, (d) selected “agree” to participate after reading the letter of informed consent. Data collected from those participants who did not meet the second criterion (exam taken within the 12-months preceding the survey) were retained for future analyses. Although participants would most likely be more sensitized to components of the testing environment if recruited to the study prior to taking the NPTE, it was determined that surveying participants after taking the examination would avoid potential distractions for the test-taker during this high-stakes exam.

A response rate of 25% was anticipated. Response rates for Internet-based surveys have been studied with mixed results. Response rates for online surveys with electronic follow-up are 32.5% on average, with more than 50% of surveys having a 26% response rate (Hamilton, 2003). Age of the respondent appears to be a consistent factor that increases the response rate for on-line surveys (Lusk et al., 2007; Mc Cabe et al., 2002). Younger respondents are more likely to respond to an on-line survey than their older counterparts, who in general prefer hard copy, mailed surveys. Certain populations are also noted to have low response rates to surveys, no matter the mode. Health professionals, particularly physicians, are noted as having poor response rates, and may be dependent on the mode of delivery, purpose of the survey, and the particular sub-group that is being targeted (Lusk et al., 2007). Since the survey in this study was designed for physical therapy graduates within the first year of practice, it was expected that the population would have a high frequency of internet access and would be skilled at on-line applications. For this study, response rates were dependent on a representative from each of the 210 accredited PT programs in the United States forwarding the survey to recent alumni, and on the willingness for the recent alumni to participate in the research. Therefore it was difficult to predict an expected rate of return. Thirty respondents from each of 30 testing centers would provide more than adequate data for the planned analyses. At any given time, there are approximately 330 testing sites for the NPTE, with more than one at a given location at select centers in large metropolitan areas.

Methods that were used to increase the response rate were to keep the survey as brief as possible, make the survey access and navigation easy, and to use a four-contact method for follow-up with program directors. This method has been associated with higher response rates (Puleo et al., 2002). The first e-mail contact to program directors was an encouraging letter asking them to assist in this research. The second e-mail was sent one week after the first and thanked program directors who had agreed to cooperate, and to re-invite non-respondents to assist with the research. A third contact to non-participating program directors was sent three to four weeks after the initial e-mail invitation. After the decision was made that there was a need for additional data, a fourth contact was made to those programs where no response had been received. A combination of phone calls and individually addressed e-mail invitations was made at that time. Any program directors who notified the researcher that they did not want to or could not participate did not receive any follow-up contacts.

Survey data were collected from 216 participants. Eight participants had taken the NPTE greater than 12 months prior to answering the survey. It was decided to include seven participants who participated in the survey within 15 months of taking the NPTE. Demographic data on these seven participants is described below. Only one participant, who took the NPTE 22 months prior to taking the survey, was excluded.

The sample of participants was fairly representative of the population of graduates from CAPTE accredited physical therapy programs in 2007 and 2008 who tested for the first time

(3/1/08 through 2/28/09) (FSBPT, 2009). Participants ranged in age from 19-52 and had a mean age of 26.3 (*SD* = 4.33) and a median age of 25*.* This is somewhat consistent with the distribution of age of graduates of CAPTE accredited PT programs (range 25-45 years, *M* = 23.8 years), but quite representative of the first-time test takers in 2008 (range 24-71 years, median = 27 years, mode = 26 years). The sample of participants consisted of 19.5% male and 80.5% female, which represents slightly more women than first-time NPTE examinees from 2008 (72.6% female, 27.4% male). The participants differed from the population of first-time test takers (U.S. graduates) in 2008 in that 41.2% graduated with a Master's degree in Physical Therapy and 58.8% graduated with a DPT (national statistics are 21.4% masters, 76.7% clinical doctorate). Greater than ninety percent (90.7%) of respondents had received their score on their most recent attempt at the NPTE. Additional demographic data related to these participants is found in Tables 3.2 and 3.3.

The seven cases that represent individuals who took the exam between 12 and 15 months or less from the time of the survey reflected similar demographic data. Seventy-one percent were female with 57% earning an entry-level DPT. The mean age of this subgroup was 29 years old and they graduated from a variety of program curriculum types, mostly hybrid. All but one was a first time test taker. Overall, the participant group included in the analysis is representative of graduates from CAPTE accredited programs in 2007 and 2008 when the study data were collected (CAPTE, 2013).

Table 3.2. Descriptive Statistics for Participant Sex, Degree Earned, Curriculum Type, and Number of Times Taken

Variable	Percentage	*n*
Sex		
Male	19.5	41
Female	80.5	169
PT degree earned		
Masters	41.2	75
Doctorate	58.8	107
Curriculum type		
Case-based	4.6	10
Guide-based	3.7	8
Hybrid	45.8	99
Lifespan-based	<.1	1
Modified PBL	15.3	33
PBL	4.2	9
Systems-based	16.7	36
Traditional	6.9	15
# of times NPTE taken		
Table 3.2 (Continued)		
1	90.8	187
2	6.3	13
3	1.9	4
4	1.0	2

Table 3.3. Descriptive Statistics for Participant Program GPA, GREv and GREq

Variable	*M*	Median	SD	Skew	Kurtosis	Range (low)	Range (high)
Program GPA	3.60	3.60	0.26	-0.43	-0.55	2.92	4.00
GREv	495.94	480.00	84.70	0.81	0.59	330	780
GREq	589.03	600.00	100.50	-0.22	-0.31	300	800
NPTE score	636.07	654.00	122.20	-3.66	14.26	71	800

Note. Both GREv and GREq statistics were based on data from 94 and 88 valid cases, respectively.

Testing Centers

To gain insight into the characteristics of testing centers, survey participants reported on the center(s) which they took the NPTE. Data were collected on 101 centers in a total of 31 states. Table 3.4 describes the distribution of survey respondents within the centers represented, as well as the sample size of each unit of analysis.

Table 3.4. Sample Size for Examinees Nested Within Centers in a 2-Level Design

Number of Examinees Per Center	Number of Centers with Specified Number of Examinees	Cumulative Frequency of Examinees
1	63	63
2	14	91
3	13	130
4	5	150
5	3	165
8	2	181
16	1	197

Data Analysis

The data were visually and graphically inspected for outliers and missing data using SPSS Version 17.0. For this study, an outlier was defined as a score or other data point that was greater than two standard deviations from the mean. Data were visually inspected to identify if there was an identifiable cause for each outlier. Those data which appeared to be arbitrary responses to items were treated as random missing data and were removed through listwise deletion. For instance, five participants provided GRE scores for both verbal and quantitative sections as zero. Since it is unlikely that the participant received a zero score, but rather chose not to provide the score or did not take the GRE, these data were removed from the analysis. Outliers for which there was no apparent cause were included in the analysis. Results were reported both including and excluding these outliers.

Before addressing each research question, preliminary descriptive statistics were generated for each variable to provide a description of the participants, as well as the centers in which they took the NPTE. The participants were described by age, sex, academic ability measures (pGPA, GRE-V, GRE-Q), degree earned, type of curriculum, on-line experience, comfort with on-line testing, preference for testing time, NPTE performance, time of day the exam was taken, use of break time, and use of headphones during testing. For continuous variables, means, standard deviations, and values for skewness and kurtosis were generated. These continuous variables included program GPA, highest verbal GRE score, highest quantitative GRE score, participant age, percent of coursework taken on-line, percent of exams taken on-line, comfort level with on-line testing, and most recent NPTE score. Additional continuous variables included the number of: times the NPTE was taken, scheduled breaks used, unscheduled break used, scheduled breaks taken in break space, and unscheduled breaks taken in break space. Categorical variables were analyzed and frequency counts, percentages, and modal scores were reported. Categorical variables included sex, curriculum type, preferred test taking time, and testing time. Dummy variables were created for all categorical variables with more than two levels for later regression analyses. Two dichotomous variables, sex and degree earned, were analyzed and frequency counts and percentages were reported.

Correlational statistics were generated to analyze the interrelationships between participant characteristics and examine potential multicollinearity of the variables. Pearson correlation coefficients were generated for each independent variable measured on an interval scale or higher (examinee characteristics) to identify the magnitude and direction of the relationship between the variables. Analyses that took into account the scale of measurement of the variables (e.g., independent *t*-tests to examine the relation between participant sex and age, and chi-square analyses for sex and type of curriculum) were conducted to examine the relations between variables.

A comparison of survey participants versus overall program graduates/examinees was made based on demographic characteristics. Available data on recent graduates from all accredited physical therapy programs (examinee candidate) in the United States were used to describe the overall number of test takers during the survey period, percent of examinees who pass on first attempt, number of overall graduates, average age, sex of the graduates, and program type. These data were obtained from the American Physical Therapy Association and the Federation of State Boards of Physical Therapy. This comparison provides further descriptive information on the representativeness of the survey participants to the total population of test takers.

Information about center characteristics (e.g., number of workstations) was collected from examinees. Data for each characteristic were aggregated across examinees for each center (e.g., mean number of workstations). Variability in the examinees' responses within a center on these center characteristics was examined. Descriptive statistics were generated for each center characteristic to provide a description of the centers in which participants took the NPTE. The centers were described by room size (# of workstations), room density (# of workstations in use), newness of the center, access to ambient light (presence of exterior windows), access to a separate break space, access to refreshments (food/drink), and the level of visual distractions. Correlational statistics were generated to analyze the interrelationships between center characteristics.

The following section organizes the discussion of the data analysis around each of the four research questions.

Research Question 1:

1a. What are the environmental preferences for the NPTE testing environment of examinees who have taken the NPTE?

1b. What is the relationship between examinees' background characteristics (e.g., sex, program GPA, age, online experience, online testing experience, comfort level with online testing, and preferred testing time) and environmental preferences (e.g., room size/layout, climate, lighting, sound, workstation chair and desktop design)?

The variables of interest are the environmental preference ratings of the participants for each of the six dimensions of the testing environment being measured. These include: (a) room size/layout, (b) room climate, (c) lighting, (d) sound, (e) workstation chair design, and (f) workstation desktop design. These six environmental dimensions were measured by 12 individual variables, which were analyzed separately. Measures of central tendency (mode, median, mean) and standard deviation, along with skewness and kurtosis of the distribution were generated. Since the adjectives used in the semantic differential scales are neutral, interpretation of the data describe the preferences of examinees and not judgment as to what is 'better' or 'higher' on a scale. Therefore, composite scores across items were not appropriate.

A series of multiple regression equations were generated, one for each of the 12 environmental preference variables (dependent variables). Given the number of dependent variables examined, an alpha level of .001 was used in tests of significance. Effect sizes (f^2) were also reported for the multiple regressions. Cohen's (1992) guidelines of .02, .15, and .35, for small, moderate, and large effects were used. The predictor variables were the examinee characteristics. Since a series of multiple regression equations were generated it was important to examine the data for violations of the assumptions of multiple regression analysis: 1) normality of the residuals, 2) linear relationship between independent and dependent variables,

3) homoscedasticity of the residuals, and 4) independence of the residuals. Predictor variables were entered into the equation simultaneously.

Research Question 2:

2a. On the dimensions of the test environment identified (e.g., room size/layout, climate, lighting, sound, workstation chair and desktop design), how much variability of examinees' perceptions of the testing environment exists between testing centers and how much is within testing centers administering the NPTE?

2b. On the dimensions of the test environment identified (e.g., room size/layout, climate, lighting, sound, workstation chair and desktop design), how much variability in the absolute difference scores (i.e., difference in examinees' perceptions and preferences of the testing environment) exists between testing centers and how much is within testing centers administering the NPTE?

2c. How much variability in the "effect on performance" and "use center again" variables exists between testing centers and how much is within testing centers administering the NPTE?

Multilevel modeling was used for this second question, which focused on how the environment of testing centers was perceived by examinees. The problem is well suited for this approach as the data of the examinees were nested in the centers in which they took the NPTE. Two level models were built using HLM 6. Analysis began with an unconditional model (no level-1 or level-2 predictors are included) to look at within and between center variability on the perceptions of each of the 12 environmental characteristics. These characteristics were: room size/layout (intimate-spacious, enclosed-open), room climate (cold-hot, dry-humid), room light (dim-bright, soft-intense), room sound (quiet-loud, muffled-clear), workstation chair (soft-hard, not adjustable-highly adjustable), and workstation desktop area (small-large, not adjustable-highly adjustable). These center characteristics, as well as the 'experiences' of the examinees, were measured through the examinees' perceptions of the test environment in which they took the exam.

An intraclass correlation coefficient (ICC) was calculated for each of the environmental characteristic variables. This baseline model allowed for partitioning of the total variability, and identified the amount of variability that is explained between centers and the amount explained within centers.

$$ICC = \frac{\text{Between center variability}}{\text{Between center variability + within center variability}}$$

ICCs were generated to partition the total variability of perceptions of environmental characteristics, the variability of absolute differences between perceptions and preferences, and the perception of the effect on performance and likelihood of choosing the same center variables.

Research Question 3:

3a. What is the relationship between examinee characteristics (sex, program GPA, age, online experience, online testing experience, comfort level with online testing, preferred time) and how examinees perceive dimensions of the test environment (room size/layout, climate, lighting, sound, workstation chair and desktop design)?

3b. What is the relationship between examinee characteristics (sex, , program GPA, age, online experience, online testing experience, comfort level with online testing, preferred time) and absolute difference scores of examinees' perceptions and preferences of the testing environment (room size/layout, climate, lighting, sound, workstation chair and desktop design)?

3c. What is the relationship between examinee characteristics (sex, program GPA, age, online experience, online testing experience, comfort level with online testing, preferred time) and perception of effect on performance and likelihood to use the center again?

Two-level multilevel modeling with examinees (level-1) nested within centers (level-2) was used for this third question. Since individuals experience environments differently, characteristics of the examinee may be related to their perceptions of the test environment. Characteristics of the examinees (level-1) that were examined included age, sex, academic ability measures (pGPA, vGRE, qGRE), degree earned, type of curriculum, on-line experience (coursework, testing), comfort with on-line testing, preference for testing time, NPTE performance (recent score, number of times taken), time of day NPTE taken, and use of break time.

A level-1 model was built for each of the 12 environmental perception ratings (dependent variables). Regression coefficients were used to examine the relation of each of the independent (predictor) variables with the dependent variables. Predictor variables were examined one at a time. Level-1 non-dichotomous predictors were grand-mean centered. The variability in the regression coefficients across centers (i.e., random effects) was evaluated to determine if the variability was significantly different from zero. Variables that have statistically significant fixed effects were kept in the Level-1 model. There is no evidence in the literature of interactions between the level-1 independent (predictor) variables (e.g., age and sex), however this is not fully known. Level-1 residuals were examined for normality of the distribution and homoscedasticity, assumptions of regression analysis. Level-2 residuals for the intercepts were also examined for normality and homoscedasticity. This analysis was repeated to look at the relationship between examinee characteristics and absolute difference scores of examinees' perceptions and preferences of the testing environment (Research question 3b) and the relationship between examinee characteristics and perception of effect on performance and likelihood to use the center again variables.

Research Question 4:

4a.What is the relationship between center characteristics (room size, room density, center 'newness', presence of ambient light, break space, and access to food/drink) and examinees' perceptions of the testing environment (room size/layout, climate, lighting, sound, workstation chair and desktop design)?

4b. What is the relationship between center characteristics (room size, room density, center 'newness', presence of ambient light, break space, and access to food/drink) and the absolute difference between examinees' preferences and perceptions of the testing environment (room size/layout, climate, lighting, sound, workstation chair and desktop design)?

4c. What is the relationship between center characteristics (room size, room density, center 'newness', presence of ambient light, break space, and access to food/drink) and examinees' perceptions of the effect on performance?

The analysis of data for Research Question 4 focused on the center level (Level-2). A level-2 model for each of the 12 dependent variables was built. These models contained level-1 predictors that were identified as statistically significant in the previous analyses (see research question 3). Regression analysis was conducted to estimate the relation of each of the independent variables (center characteristics) with each of the dependent variables (examinee environmental perceptions). Each level-2 predictor variable was centered around the grand mean. For the level-2 models, only variables with complete data on the level-2 variables were included in the model. Variables that were statistically significant remained in the level-2 model. Although there is no evidence in the literature that there are interactions between the level-2 independent variables, exploratory analyses were conducted to examine potential two-way interactions. Level-1 and level-2 residuals for the final model were examined for normality and homoscedasticity using graphical (boxplots, scatterplots) and statistical procedures (e, g., measures of skewness and kurtosis). This analysis was repeated to look at the relationship between examinee characteristics and absolute difference scores of examinees' perceptions and preferences of the testing environment (Research question 4b) and the relationship between examinee characteristics and perception of effect on performance.

Open-Ended Question Responses

Lastly, participants were asked an open ended question at the end of the survey. The open-ended question asked:

"Using your own words, please describe the testing center at which you took the exam. Think about the testing room itself (size, temperature, lighting, sound, etc.), the workstation, the chair, the computer set up, the break space. If you took the exam more than once, and used a different center, please describe the differences in the centers."

One hundred and eighteen of the participants provided a response to the open-ended question (55%). Most of these participants wrote several sentence responses addressing more than one aspect of the environment. Responses from the open ended question were grouped by environmental characteristic (e.g., room size, window, temperature, etc.). Frequency counts of responses categorized under each environmental characteristic were reported. Common topics that emerged were summarized and used to enhance the survey data findings. Responses were also coded by whether the category was "not mentioned", "mentioned–negative evaluative statement", "mentioned–neither negative nor positive evaluative statement", or "mentioned–positive evaluative comment". A reliability check was performed by a person familiar with coding assessment data as well as with licensure/certification testing. Twenty percent of the open-ended responses were categorized using the same coding system. Agreement for categorization of these responses was 91%.

CHAPTER FOUR:

RESULTS

The purpose of this study was to assess the testing environment as perceived by individuals taking the National Physical Therapy Examination (NPTE). Both individual preferences for and perceptions of the testing environment were explored. This chapter is organized in terms of the four research questions posed in Chapter 1. Prior to examining the research questions, descriptive statistics of the participants' previous online experience, as well as their experience and behaviors related to taking the NPTE, are provided. Descriptive information about the testing centers also are presented. Then the results of the analysis of data related to each of the research questions are reported.

Participants

Details of the participants are provided in Chapter 3. To summarize, survey data were collected from 216 participants. Eight participants had not taken the NPTE within 12 months of taking the survey. Seven of these eight participants' data were included in the analysis as they were within 15 months of taking the NPTE and it was believed that memory for the testing experience was still current. One participant was excluded from the analysis as this person was almost two years from taking the NPTE.

Participants were 80.5% female, with 41.2% earning an entry-level master's degree and 58.8% earning an entry level DPT. The participants' ages ranged from 19 to 52 years, with a mean age of 26.3 years. They graduated from programs that used a wide variety of curriculum designs. The vast majority of participants were first time test takers (90.8%), with 6.3% of the participants being second time test takers. Additionally, the seven cases that represent individuals who took the exam between 12 and 15 months or less from the time of the survey reflected similar demographic data. Seventy-one percent were female with 57% earning an entry-level DPT. The mean age of this subgroup was 29 years old and they graduated from a variety of program curriculum types, mostly hybrid. All but one was a first-time test taker. Overall, the participant group included in the analysis appears to be representative of graduates from CAPTE accredited programs in 2007 and 2008 when the study data were collected (CAPTE, 2013).

Examinees' Previous Online Experiences

Since the NPTE is a computer-based exam, the participants' experiences with learning and testing using a computer were of interest. In higher education, coursework using computers is most typically delivered in an online format. Therefore, participants were asked to provide an estimate of the percentage of their Physical Therapy (PT) coursework that was delivered online, the percentage of their testing that was conducted online, and their overall comfort with taking exams online. Approximately 90% of the participants had relatively low levels of experience with online coursework and testing, reporting that only 0-25% of their coursework and testing in their professional program was delivered online. As a whole, the participants had a relatively high level of discomfort in taking exams in a fully online format. Comfort level with online testing was rated on a 1-5 scale where 1 was "*very comfortable*" and 5 "*very uncomfortable*". One quarter of the participants rated themselves in the middle of the scale (3) and another 11% rated themselves in the comfortable to very comfortable range. Sixty-four percent of the respondents rated themselves in the uncomfortable to very uncomfortable range. Appendix B displays additional details for these three variables.

Participants were also asked about their behavior related to taking breaks during the five-hour testing period. The frequency of breaks taken was of interest to analyze the participants' need for breaks during this long exam period. Participants were asked the number of scheduled and unscheduled breaks that they took, and the number of times that they left the testing room for these breaks. The vast majority of participants elected to take the one allowed 15-minute break period outside the testing room (96.1%) during which the test clock stops. Forty-two percent of the participants elected to take additional unscheduled breaks during which the time clock does not stop. The vast majority (88%) of those taking these additional breaks left the testing room during the break. Additional descriptive statistics for these variables are displayed in Appendix C. Responses to the open-ended question asking about taking breaks during the exam and the space provided are found in Appendix D: Responses to open-ended question related to "breaks".

Fifty-four percent of participants were able to take the test at their preferred time of day. Greater than 90% of participants preferred a morning start time, with better than half of those preferring an early morning start time. Sixty-five percent of the participants reported that their exam was scheduled for an early morning start time. However, only 40% of those having an early morning start time would prefer to take a test at that time of day. Most appointments for the NPTE are scheduled early in the day because of the length of the exam. Only a very small number of participants who preferred testing during the afternoon actually took the exam in the afternoon (Table 4.1: Crosstabulation).

Table 4.1. Crosstabulation for Participant Preferred Test Time vs. Actual Test Time

	Preferred Test Time				
	Early Morning	Late Morning	Early Afternoon	Late Afternoon	
Actual test time	*n*	*n*	*n*	*n*	
Early morning	82	9	8	4	
Late morning	47	20	16	3	
Early afternoon	3	2	6	0	
Late afternoon	1	1	1	1	
	132	32	31	8	204

Use of Headphones

Examinees are not permitted to bring earplugs or any other device to dampen or block out noise. The Prometric centers are expected to offer examinees headphones to use during the exam. An equal number of participants reported using headphones as those not using headphones during the exam. As part of the open-ended question asked of participants, many comments referred to the use of headphones. Comments included that the headphones made them feel like they were in a tunnel and that they became uncomfortable over time, but that for some, it reduced the distraction of other test takers' typing and moving about. All comments related to the use of headphones are in Appendix E: Responses to open-ended question related to use of headphones.

Center Characteristics

Preliminary descriptive statistics were generated for each center characteristic. Information on center characteristics could only be obtained by self-report by test takers. Prometric does not provide descriptive data on the centers so participant self-report was the only method to gather these data. Data were aggregated by center and represented 103 centers across the United States. Sixty-two centers were rated by only one participant. For centers where more than one participant rated the center, one-way ANOVAs were run by center to determine the mean rating and standard deviation for estimation of room size, room density, center newness, presence of ambient light, the level of visual distractions, presence of a break room, and access to food/drink.

Participants rated room size through estimating the number of workstations that were in the testing room. A rating of one estimated 1-10 workstations, two estimated 11-20 workstations, three estimated 21-30 workstations, and four estimated >30 workstations. Mean ratings for room size for each center ranged from 1.00 to 4.00, with a mean of 2.13 and a median of 2.00. The majority of the centers (56.6%) were estimated to have 11-20 workstations in the room. Table 4.2 displays the mean rating for room size across centers. The level of agreement in the perception of room size by individuals in the multi-participant centers varied. Of the 41 multi-participant centers, 14 demonstrated perfect agreement by participants in their estimate of size (SD = 0.00). Twenty-six centers lacked perfect agreement by the participants who rated room size (SDs ranged 0.45 to 1.16).

Participants rated room density through estimating the number of workstations that were in use in the testing room. A rating of one estimated 1-10 workstations in use, two estimated 11-20 workstations in use, three estimated 21-30 workstations in use, and four estimated >30 workstations in use. Mean ratings for center density for each center range from 1.00 to 4.00 with a mean of 1.71 and median of 2.00. Participants estimated that the vast majority of the centers

(88.3%) had 0-20 other test takers in the room. Table 4.2 displays the mean rating for room density across centers. The level of agreement in the perception of room density by individuals in the multi-participant centers varied. Of the 41 multi-participant centers, 13 demonstrated perfect agreement by participants in their estimate of room density (*SD* = 0.00). Twenty-eight centers lacked perfect agreement by the participants who rated room density (*SD*s ranged 0.50 to 1.41). Agreement may be affected not only by the participant's perception of the room, but also center usage which may vary by the day of the wek, time of the year, and time of testing.

Participants also rated center "newness" on a scale of 1= *new/newly renovated* and 5 = *worn/outdated*. Data represented the same 103 centers across the United States. Mean ratings for center newness for each center ranged from 1.00 to 5.00 with a mean of 3.46 and a median of 4.0. The centers were mostly rated toward the '*new/newly renovated'* end of the scale. Table 4.2 displays the mean rating for this variable. The level of agreement in the perception of room newness by individuals in the multi-participant centers varied. Of the 41 multi-participant centers, 10 demonstrated perfect agreement by participants in their rating of newness (*SD* = 0.00). Thirty-one centers lacked perfect agreement by the participants who rated center newness (*SD*s ranged 0.35 to 1.53).

The presence of ambient light was rated on a dichotomous scale where 0 was when there was no window and 1 was when there was a window present. Mean ratings for the presence of ambient light for each center ranged from 0.00 to 1.00 with a mean of .39 (*SD* = 0.42). The level of agreement in the perception that a window was present or not in multi-participant centers varied. Of the 41 multi-participant centers, 29 demonstrated perfect agreement by participants in their perception of the presence of a window (*SD* = 0.00). Twelve centers lacked perfect agreement by the participants who rated the presence of a window (*SD*s ranged .032 to 0.86). There was no apparent reason that the participants would rate this center characteristic differently. When asked if there were exterior windows in the testing room that provided natural light, only 38.6% of all participants reported that this was present.

Participants rated visual distractions on a scale of 1-5 where 1 equaled *no distractions* and 5 equaled *constant distractions*. Data represented the same 103 centers across the United States. Mean ratings for center visual distractions for each center ranged from 1.00 to 5.00 with a mean of 1.94 and a median of 2.0. Centers were perceived to be on the lower end of the scale for visual distractions. Table 4.2 displays the mean rating for this variable. The level of agreement in the perception of visual distractions by individuals in the multi-participant centers varied. Of the 41 multi-participant centers, only eight demonstrated perfect agreement by participants in their rating of visual distractions (*SD* = 0.00). Thirty-three centers lacked perfect agreement by the participants who rated center visual distractions (*SD*s ranged 0.50 to 1.41). Similar to the presence of ambient light, agreement may be affected not only by the participant's perception of the room, but also by center usage which may vary by the day of the week, time of the year, and time of testing.

The presence of a break room was rated as 0 for no break room and 1 for one being present. Data represented the same 103 centers across the United States. Mean ratings for the presence of a break room for each center ranged from 0.00 to 1.00 with a mean of 0.81 and a median of 1.000 (*SD* = 0.39). Centers were perceived to be on the higher end of the scale for presence of a break room. Table 4.2 displays the mean rating for this variable. The level of agreement in the perception of the presence of a break room by individuals in the multi-participant centers varied. Of the 41 multi-participant centers, only 21 demonstrated perfect agreement by participants in their rating of presence of a break room (*SD* = 0.00). Twenty centers lacked perfect agreement by the participants who rated the presence of a break room (*SD* ranged 0.45 to 1.26). Similar to the perception of the presence of ambient light, there is no apparent reason that the participants would rate this center characteristic differently. There is the possibility that the participant was not informed of this center feature. It was reported by only 19.1% of the participants overall that a separate break space was provided in the center in which they took their exam.

Access to food/drink was rated on a 3-point scale where 0 was when no access was provided, 1 was when there was only access to drink, and 2 was when both food and drink were accessible. Data represented the same 103 centers across the United States. Mean ratings for access to food/drink for each center ranged from 0.00 to 1.00 with a mean of 0.53 and a median of 1.0 (*SD* = 0.44). Table 4.2 displays the mean rating for this variable. The level of agreement in the perception of the access to food/drink by individuals in the multi-participant centers varied. Of the 41 multi-participant centers, only 15 demonstrated perfect agreement by participants in their rating of visual distractions (*SD* = 0.00). Thirty-six centers lacked perfect agreement by the participants who rated center visual distractions (*SD* ranged 0.47 to 1.32). Similar to the perception of the presence of break space, there is no apparent reason that the participants would rate this center characteristic differently unless the participant was not informed of this center feature. It was reported by only 46.8% of the participants overall that a food and/or drink was accessible in the center in which they took their exam. Within the open ended question, some reported being able to store refreshments in lockers provided at the testing center.

Table 4.2 displays descriptive statistics for the participants' estimate of room size (# of workstations), room density (# of workstations in use), the newness of the center, presence of a window (ambient light), level of visual distractions, availability of a break room, and access to food/drink.

Table 4.2. Descriptive Statistics for Center Characteristics Aggregated by Center Code: Testing Room Size, Density, Newness, Ambient Light, Level of Visual Distraction, Access to Food/Drink, and Effect on Performance

Variable	*n*	Min	Max	*M*	*SD*
Room size	103	1.00	4.00	2.13	0.68
(Number of workstations)					
Room Fullness/density	103	1.00	4.00	1.71	0.66
(number of workstations in use)					
Center Newness	103	1.00	5.00	3.46	0.74
Window/ambient light	100	0.00	1.00	0.39	0.42
Visual distractions	103	1.00	4.00	1.94	0.65
Break Room	103	0.00	1.00	0.81	0.39
Access to food/drink	100	0.00	1.00	0.53	0.44
Effect on performance	100	2.00	7.00	4.83	1.17

Note: Room Size: 1 = 1-10 workstations, 2 = 11-20 workstations, 3 = 21-30 workstations, 4 = > 30 workstations. Room Density: 1 = 0-10 workstations, 2 = 11-20 workstations, 3 = 21-30 workstations, 4 = > 30 workstations. Center Newness: 1 = new/newly renovated; 5 = worn/outdated; Window: 0 = no, 1 = yes; Visual Distractions: 1 = no distractions; 5 = constant distractions; Break room: 0 = No, 1 = yes; access to food/drink: 0= no, 1= yes, drink only, 2= yes, food and drink. *n* = number of centers

Lastly, correlational statistics were generated to analyze the interrelationships between center characteristics. Significant positive correlations (p<.01) were generated for room size and center newness (r =.29), room size and room density (r =.65), center newness and presence of ambient light (r =.21) and the presence of a break room and access to food/drink (r =.26). A small significant negative correlation was observed between the center newness and the level of visual distractions (r =-.20).

Research Question 1

1a. What are the environmental preferences for the NPTE testing environment of examinees who have taken the NPTE?

1b. What is the relationship between participants' background characteristics (e.g., sex, , program GPA, age, online experience, online testing experience, comfort level with online testing, and preferred testing time) and environmental preferences (e.g., room size/layout, climate, lighting, sound, workstation chair and desktop design)?

Description of Preferences

Data from the 12 environmental preference items were analyzed and the results are provided in Table 4.3. Respondents rated their preference for six elements of the environment, two for each element. They rated each item on a 7-point scale where one descriptor was at one end and an opposite descriptor was on the other. Responses to items related to room size (intimate/spacious, enclosed/open), room temperature (cold/hot, dry/humid), room light (soft/intense), and room sound (muffled/clear) were clustered toward the middle of the scale. Eleven out of the 12 items had skewness and kurtosis values with -1.5 to 1.5 suggesting that these item responses exhibited no major departures from normality. The item with the greatest departure from normality was Room sound (quiet) (M = 1.53, SD = 0.92, Sk = 2.59, K = 9.14). The preference toward the workstation chair was slightly toward a softer (M = 5.11, SD = 1.34) and more adjustable chair (*M* = 4.51, *SD* = 1.79), for the room lighting to be slightly toward the bright end of the scale (*M* = 5.09, *SD* = 1.14) and for workstation desktop size (small/large) to be slightly on the smaller side (*M* = 3.06, *SD* = 1.26). There was only a clear preference toward one end of the scale on two of the 12 items. The vast majority preferred a quiet room (*M* = 1.53, *SD* = 0.92) and most respondents preferred a desktop area that had a great deal of adjustability (*M* = 5.91, *SD* = 1.26).

Correlational Analysis

Correlational statistics were generated for each of the 12 preference variables prior to multiple regression analysis. Weak relationships existed between variables ranging from -.33 to .43, but none were statistically significant (p>.001).

Multiple Regression for Preferences

Multiple regression analysis was conducted for each of the 12 environmental preference outcome variables. Seven predictor variables (sex, program GPA, age, online experience, online testing experience, comfort level with online testing, and preferred testing time) were entered into each model. These seven predictor variables accounted for < 7% of the variability found in each of the outcome variables (sample sizes for these models ranged from 186 to 191). R-square values ranged from .015 to .068, and none was statistically significant. R-square values, unstandardized coefficients and standard errors for each of the 12 models are provided in Table 4.4. None of the predictor variable unstandardized regression coefficients were statistically significant for any of the 12 models. These unstandardized coefficients were mostly very small to medium in size, ranging from .00 to 1.00.

Table 4.3. Descriptive Statistics for Examinee Environmental Preferences

						Ratings (percent of responses)						
Variable	*n*	*M*	*SD*	*Sk*	*K*	1	2	3	4	5	6	7
Room size												
Intimate[a]/Spacious[b]	201	4.34	1.19	-0.05	-1.04	8.5	10.4	14.4	23.4	13.4	8.0	21.9
Enclosed[a]/Open[b]	200	3.55	2.02	0.36	-1.16	19.0	20.5	13.5	13.5	12.0	9.5	12.0
Room temperature												
Cold[a]/Hot[b]	199	3.90	0.95	-0.35	0.64	1.5	5.0	22.1	47.7	20.1	3.5	0.0
Dry[a]/Humid[b]	198	3.36	1.04	-0.91	0.25	8.6	10.6	21.7	54.5	4.0	0.5	0.0
Room light												
Dim[a]/Bright[b]	196	5.09	1.14	-0.13	-0.48	0.0	1.0	6.1	24.5	31.1	26.0	11.2
Soft[a]/Intense[b]	194	3.59	1.22	-0.31	0.12	6.7	13.4	16.5	45.4	14.4	2.6	1.0
Room sound												
Quiet[a]/Loud[b]	196	1.53	0.92	2.59	9.14	64.3	26.0	4.6	4.1	0.0	0.5	0.5
Muffled[a]/Clear[b]	198	3.79	2.21	0.19	-1.42	20.7	17.7	10.1	14.1	7.1	11.6	18.7
Workstation chair												
Soft[a]/Hard[b]	197	5.11	1.34	0.01	-0.51	13.8	20.0	24.6	31.3	8.7	1.0	0.5

Table 4.3. (Continued)

No Adjust[a]/High Adjust[b]	196	4.51	1.79	-1.01	0.33	0.5	0.0	4.1	12.2	15.3	21.9	45.9
Workstation desktop												
Small[a]/Large[b]	195	3.06	1.26	-0.26	-0.46	0.5	2.5	6.6	25.4	25.4	20.3	19.3
No Adjust[a]/High Adjust[b]	196	5.91	1.26	-0.30	-0.68	8.2	6.6	10.7	23.5	21.4	11.2	18.4

Note. Superscript a for each of the first words (category 1) and b for last words (category 7)

Residuals from each of these 12 environmental preference models were examined to evaluate the assumptions underlying the multiple regression analyses. The distributions of the residuals were approximately normally distributed through visual examination of plots as well as descriptive statistics (skewness and kurtosis). Standardized residuals (y-axis) and predicted values (x-axis) were examined using scatterplots. Plots indicated no evidence of heterogeneity of variance of the residuals across predicted values and no signs of curvilinearity. Standardized residual values were between -4.0 to 3.5.

Since there were multiple respondents from the same testing center (i.e., nested data structure), the previous analyses were replicated using multilevel modeling. HLM 6 was used to examine the relationships between the seven predictor variables and each of the outcome measures (environmental preferences) using two-level models (respondents nested within centers). The ICCs for these 12 outcome variables ranged from .004 to .249 with a median of .069 and a mean of .087. This indicates that there was very little between center variability. No coefficients for the predictors were statistically significant. Predictor variables were entered simultaneously and were treated as fixed effects. Tables are presented in Appendix D *HLM Models for nested data for preference variables* presents contain B (unstandardized regression coefficient) and standard errors for nested data. No coefficients were statistically significant (p>.001). A more stringent significance level of .001 was used because of the number of comparisons that were done. No participant characteristics were related to participant preferences. The results from the analyses that took into account the nested structure versus the results that did not take into account the nested structure were virtually the same.

Table 4.4. Regression Model Summary of Participants' Environmental Preferences

Predictor Variable	Size (small-spacious)		Size (enclosed-open)		Temp (cold-hot)		Temp (dry-humid)	
	r	*B*	*r*	*B*	*r*	*B*	*r*	*B*
Sex[a]	-.01	.05 (.39)	.01	.10 (.37)	.09	.16 (.18)	.06	.16 (.19)
GPA	.01	.11 (.56)	.06	.53 (.59)	.05	.08 (.28)	.01	-.09 (.30)
Age	.04	.03 (.03)	-.00	.02 (.03)	-.01	-.01 (.02)	.13	.03 (.02)
Online[b]	.14	.66 (.31)	.17	.77 (.32)	-.08	-.20 (.15)	-.06	-.06 (.16)
Online-test[c]	-.06	-.43 (.36)	.01	-.15 (.38)	.08	.20 (.18)	-.07	-.28 (.19)
Comfort[d]	-.05	-.10 (.15)	-.05	-.13 (.15)	.13	.11 (.07)	.16	.15 (.08)
Preferred time[e]	-.06	-.14 (..20)	-.03	-.05 (.21)	-.03	.03 (.10)	.09	.16 (.11)
R^2		.04		.04		.04		.06
N		*191*		*190*		*189*		*188*

Dependent Variable (spans all outcome columns)

Note. None of the unstandardized regression coefficients (b) was statistically significant (all *p*-levels were > .001). Standard errors are in parentheses. [a] Sex (1 = Male, 2 = Female); [b] Online – percentage of coursework 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%; [c] Online test – percentage of testing 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%; [d] Comfort – online testing 1 = very comfortable, 5 = very uncomfortable; [e] preferred time for taking tests 1 = early morning (before 10 am), 2 = late morning (10 am – noon); 3 = early afternoon (after noon but before 3 pm); 4 = late afternoon (3 pm or after).

Table 4.4 (Continued) Regression Model Summary of Participants' Environmental Preferences

Predictor Variable	Dependent Variable: Light (dim-bright)		Light (soft-intense)		Sound (quiet-loud)		Sound (muff-clear)	
	r	*B*	*r*	*B*	*r*	*B*	*r*	*B*
Sex [a]	.12	.34 (.22)	.04	.20 (.23)	-.14	-.29 (.15)	-.08	-.41 (.42)
GPA	-.09	-.44 (.33)	-.09	-.18 (.36)	.10	.32 (.23)	-.11	-1.04 (.66)
Age	-.03	-.02 (.02)	.03	.02 (.02)	-.07	-.01 (.01)	-.07	-.06 (.04)
Online [b]	-.04	-.15 (.20)	-.12	-.25 (.20)	.02	.01 (.13)	-.02	-.11 (.36)
Online-test [c]	.12	.38 (.21)	-.10	-.28 (.23)	-.06	-.07 (.15)	-.02	-.05 (.42)
Comfort [d]	-.06	-.05 (.09)	-.12	-.15 (.09)	-.05	-.04 (.06)	-.08	-.07 (.17)
Preferred time [e]	-.02	-.09 (.12)	.09	.17 (.13)	-.03	-.01 (.08)	-.05	-.21 (.24)
R^2		.05		.05		.04		.03
N		*186*		*186*		*187*		*189*

Note. None of the coefficients was statistically significant (all *p*-levels were > .001). Standard errors are in parentheses.
[a] Sex (1 = Male, 2 = Female); [b] Online – percentage of coursework 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%; [c] Online test – percentage of testing 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%; [d] Comfort – online testing 1 = very comfortable, 5 = very uncomfortable; [e]preferred time for taking tests 1 = early morning (before 10 am), 2 = late morning (10 am – noon); 3 = early afternoon (after noon but before 3 pm); 4 = late afternoon (3 pm or after).

Table 4.4 (Continued) Regression Model Summary of Participants' Environmental Preferences

	Dependent Variable							
Predictor Variable	Chair (soft-hard)		Chair (not adj-high adj)		Station (sm-large)		Station (not adj-high adj)	
	r	*B*	*r*	*B*	*r*	*B*	*r*	*B*
Sex [a]	-.03	-.10 (.24)	.14	.49 (.24)	-.06	-.20 (.26)	.06	.33 (.34)
GPA	.00	-.04 (37)	.00	.23 (.37)	.02	.26 (.40)	-.02	-.08 (.53)
Age	-.08	-.02 (.02)	.11	.05 (.02)	-.04	-.00 (.02)	-.04	-.01 (.03)
Online [b]	.04	.11 (.20)	.07	.21 (.20)	.03	.06 (.22)	.01	-.05 (.29)
Online-test [c]	-.07	-.22 (.24)	.08	.17 (.24)	.00	.03 (.26)	-.02	-.05 (.34)
Comfort [d]	-.02	-.01 (.10)	-.12	-.21 (.10)	-.14	-.20 (.11)	-.12	-.21 (.14)
Preferred time [e]	.04	.07 (.14)	.00	.02 (.13)	.03	.09 (.14)	-.00	.00 (.20)
R^2		.02		.07		.03		.02
N		*186*		*187*		*188*		*187*

Note. None of the coefficients was statistically significant (all *p*-levels were > .001). Standard errors are in parentheses.
[a] Sex (1 = Male, 2 = Female); [b] Online – percentage of coursework 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%; [c] Online test – percentage of testing 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%; [d] Comfort – online testing 1 = very comfortable, 5 = very uncomfortable; [e]preferred time for taking tests 1 = early morning (before 10 am), 2 = late morning (10 am – noon); 3 = early afternoon (after noon but before 3 pm); 4 = late afternoon (3 pm or after).

Research Question 2

2a. On the dimensions of the test environment identified (e.g., room size/layout, climate, lighting, sound, workstation chair and desktop design), how much variability of examinees' perceptions of the testing environment exists between testing centers and how much is within testing centers administering the NPTE?

2b. On the dimensions of the test environment identified (e.g., room size/layout, climate, lighting, sound, workstation chair and desktop design), how much variability in the absolute difference scores (i.e., difference in examinees' perceptions and preferences of the testing environment) exists between testing centers and how much is within testing centers administering the NPTE?

2c. How much variability in the "effect on performance" and "use center again" variables exists between testing centers and how much is within testing centers administering the NPTE?

Research question two was examined using two-level multilevel models. This question focused on how the environment of testing centers was perceived by examinees. The problem is well suited for this approach as the data of the examinees were nested in the centers in which they took the NPTE. It was important to identify whether there was between-group and within-group variability prior to examining research question 3 and 4 where individual participant characteristics and center characteristics are analyzed. If there is within group variability, it is reasonable to further examine individual examinee characteristics in research question three. If between center variability was observed, further analysis of center characteristics may provide insight into the center characteristics that may account for this variability (research question four).

Description of Perceptions of Environmental Characteristics

Prior to these analyses, descriptive statistics were generated for each of the 12 perception variables (Table 4.5). Respondents rated their perception of their experience within the testing center on six elements, two for each element. They rated each item on a 7-point scale where one descriptor was at one end and an opposite descriptor was on the other. Since there is little known about the characteristics of these testing centers, descriptive statistics of these indirect measures of the centers' characteristics provide insight to the environment in these centers. Ratings were normally distributed and mean values were relatively central on the scale of 1 to 7, ranging from 2.81 to 3.89 (SD < 2), except for room light dim/bright (M = 4.57, SD = 1.11) and the adjustability of the workstation desktop (no adjust/high adjust) (M = 2.36, SD = 1.66).

Observations of the distribution of responses indicate that the majority of participant perceptions of the testing environment were that the centers were relatively intimate/small and enclosed, and room climate was fairly neutral for cold/hot and dry/humid. Most participants perceived the center in which they took the NPTE to have moderate or slightly toward brighter room lighting with the intensity being neutral or slightly toward the softer end of the scale. Room sound was perceived in the majority of centers to be toward the quiet end of the scale and somewhat muffled. Furniture (chair, desk) was also perceived to be clustered around the center of each scale, with the chairs being slightly toward the softer end of the scale and the desk being slightly toward the smaller end of the scale. The participants' perception of chair adjustability was relatively equally distributed across the scale. However, the vast majority of participants rated the adjustability of the desktop to not allow for any or very little adjustments.

Description of Absolute Differences Between Perceptions and Preferences

A correlational analysis was performed to investigate the relationship between environmental preferences and perception variables. Table 4.6 represents the correlation between the participants' preferences and perceptions of environmental testing conditions.

Table 4.5. Descriptive Statistics for Examinee Environmental Perceptions

Variable	*n*	*M*	*SD*	*Sk*	*K*	Ratings (percent of responses)						
						1	2	3	4	5	6	7
Room size												
Intimate[a]/Spacious[b]	200	3.50	1.45	0.26	-0.54	7.0	21.0	23.0	25.0	14.0	8.0	2.0
Enclosed[a]/Open[b]	198	2.81	1.51	0 .62	-0.51	21.7	29.8	17.7	13.6	12.1	4.0	1.0
Room temperature												
Cold[a]/Hot[b]	198	3.58	1.18	-0.14	0 .08	5.6	11.6	24.7	42.4	9.1	6.6	0.0
Dry[a]/Humid[b]	198	3.34	1.05	-0.86	-0.03	8.6	11.6	22.7	51.0	6.1	0.0	0.0
Room light												
Dim[a]/Bright[b]	196	4.57	1.11	0.10	-0.07	.5	1.0	13.8	35.2	29.1	1.8	4.6
Soft[a]/Intense[b]	193	3.88	1.07	-0.12	1.33	3.1	5.2	20.2	50.8	15.0	4.1	1.6
Room sound												
Quiet[a]/Loud[b]	195	2.82	1.44	0.77	0.27	18.5	29.2	24.1	15.9	7.2	3.1	2.1
Muffled[a]/Clear[b]	197	3.59	1.51	0.34	-0.30	7.6	16.8	24.4	27.9	11.2	7.6	4.6
Workstation chair												
Soft[a]/Hard[b]	195	3.89	1.34	0.20	0.36	4.6	8.2	22.1	40.0	14.4	5.6	5.1

Table 4.5. (Continued)

No Adjust[a]/High Adjust[b]	159	3.89	1.79	-0.06	-0.99	13.2	10.7	18.9	17.0	18.9	14.5	6.9
Workstation desktop												
Small[a]/Large[b]	196	3.48	1.29	-0.08	-0.25	8.2	13.8	25.5	31.6	16.8	3.1	1.0
No Adjust[a]/High Adjust[b]	175	2.36	1.66	0.87	-0.59	49.1	14.3	9.1	11.4	11.4	4.0	0.6

Note. Superscript a for each of the first words (category 1) and b for last words (category 7)

Table 4.6 Pearson Product Moment Correlations Between Participants' Preferences and Perceptions of Environmental Testing Conditions

Condition	*r*	*p*
Room Size: small/spacious	.092	.193
Room Size: enclosed/open	.183	.010*
Room Climate: cold/hot	-.018	.798
Room Climate: dry/humid	.719	.000*
Room Light: dim/bright	.363	.000*
Room Light: soft/intense	.329	.000*
Room Sound: quiet/loud	.020	.786
Room Sound: muffled/clear	.533	.000*
Chair: soft/hard	.288	.000*
Chair: not adjustable/adjustable	-.086	.279
Desk: small/large	.111	.121
Desk: not adjustable/adjustable	-.051	.501

Note. * $p < .01$

To analyze the differences between participants' preferred environmental factors and those that they perceived in the center in which they took the NPTE, difference scores were created. The difference score demonstrated the degree to which the preference score and the perceived score differed, and the direction of that difference (- or +). Since the direction of the difference does not have meaning in this study, an absolute difference score was generated. This score eliminated the direction of the difference and just focused on the size of the difference between the preference (what an individual wants) and perception score (what the individual got). These difference scores allowed for the analysis of how different individuals' preference for an environmental characteristic was from what they perceived in the testing environment. For instance, if a person preferred a spacious room (e.g., score of 7) and the person took their exam in a room perceived to be spacious (e.g., score of 7), the absolute difference would be zero. If a person preferred a more intimate environment (e.g., score of 2) and tested in a very spacious room (e.g., score of 7), the absolute difference score would be 5. Descriptive statistics were generated for each of these 12 absolute difference variables (Table 4.7). The same 7-point scale was used when looking at absolute differences between participants' preferences and perceived factors (0 = no difference between preference and perceived rating; 6 = maximal difference between preference and perceived rating).The greatest agreement (least absolute difference) and strongest grouping of absolute differences was found in the characteristic of room temperature dry/humid) in which 78.7% of the participants had no difference in ratings (K = 6.10). Fifty-six percent of the participants had full-agreement with their perception of the room's lighting (soft/intense) and their preference (K=3.66). All other distributions of the absolute difference variable were approximately normally distributed. Approximately 30% of the participants reported that their perception of the testing room was the same as their preference in the areas of room size (intimate/spacious), room sound (quiet/loud), and workstation desktop (small/large). A moderate amount of agreement (moderate absolute difference) was seen in the characteristics of room size (enclosed/open), room temperature (cold/hot), room light (dim/bright), room sound (muffled/clear), and workstation chair (soft/hard) with approximately 40-45% of participants in full agreement and only 5-10% differing in ratings by 4 or more.

Observations of the distribution of difference scores indicate that the areas of greatest difference in the preferences and perceptions of the testing environment were in the area of adjustability of the furniture (desk, chair). Most participants preferred the center in which they took the NPTE to have highly adjustable furniture (desk, chair) but the perception of the centers was that the adjustability was low to moderate. The scores of absolute difference between preference and perception for the adjustability of both the chair and desktop were distributed evenly from 0 (full agreement) to 6 (maximal disagreement). The mean absolute difference between preferences and perception of workstation adjustability was 2.38 (chair) and 2.64 (desktop), with only 18.9 and 17.7 percent of participants having 0 absolute difference, respectively.

Description of effect on performance and using the center again

Participants were asked to rate their likelihood of choosing the same center if they needed to take the NPTE again. The 7-point scale was anchored with “not likely” at the low end of the scale (1) and “highly likely” at the high end of the scale (7). They were also asked to rate the effect that the center may have had on their performance on the NPTE. The scale for this question ranged from “it prevented me from performing at my best” at the low end of the scale (1), to “it allowed me to perform at my best” at the high end of the 7-point scale (7). These questions were asked to gain an overall perspective of the participants’ attitude about the center and whether they perceived a cause/effect between the center characteristics and performance. These data were then entered into

Table 4.7. Descriptive Statistics for Absolute Differences in Examinee Environmental Preferences and Perceptions

Variable	*n*	*M*	*SD*	*Sk*	*K*		Ratings (percent of responses) 0	1	2	3	4	5	6
Room size													
Intimate[a]/Spacious[b]	200	1.76	1.68	0.76	-0.45		29.5	24.0	17.5	10.55	9.0	7.5	2.0
Enclosed[a]/Open[b]	198	1.65	1.68	0.88	-0.32		35.4	23.7	11.6	10.6	10.1	5.1	3.5
Room temperature													
Cold[a]/Hot[b]	198	1.03	1.16	-1.02	0.34		42.9	27.8	16.7	8.6	3.5	0.5	0
Dry[a]/Humid[b]	197	0.32	0.71	2.44	6.10		78.7	12.7	6.6	1.5	0.5	0	0
Room light													
Dim[a]/Bright[b]	195	0.93	1.00	1.11	1.24		41.5	33.3	18.5	4.6	1.5	.5	0
Soft[a]/Intense[b]	192	0.78	1.12	1.81	3.66		55.7	22.9	15.1	3.1	.5	2.6	0
Room sound													
Quiet[a]/Loud[b]	195	1.53	1.47	1.05	0 .71		28.2	30.3	19.5	11.8	5.1	3.1	2.1
Muffled[a]/Clear[b] 197	1.21	1.45	1.13	0.59		45.7	18.8	16.8	10.2	5.1	2.5	1.0	
Workstation chair													
Soft[a]/Hard[b]	194	1.53	1.47	1.05	0.71		46.4	21.6	19.1	8.2	1.0	1.0	2.6
No Adjust[a]/High Adjust[b] 159	2.38	1.92	.443	-1.01		18.9	23.3	15.1	13.2	10.7	10.7	8.2	
Workstation desktop													
Small[a]/Large[b]	196	1.76	1.61	.748	-.135		28.6	22.4	16.8	18.9	6.1	4.1	3.1
No Adjust[a]/High Adjust[b] 175	2.64	1.98	.269	-1.12		17.7	17.7	15.4	13.7	15.4	15.4	12.6	

Note. Superscript a notes each of the first words (category 1) and b for last words (category 7).

a correlation analysis with the 12 absolute difference variables created from the difference between preferences and perceptions. The absolute difference scores were new measures that were created. Therefore there was interest in seeing how these measures related to other outcome variables.

Statistically significant, but weak negative correlations where observed for all absolute difference variables with the respondents rating of whether they would re-use the center for a future testing experience, excluding the two "room size" variables. These correlations were statistically significant and negative, but moderate in strength. With each of these 12 absolute difference variables, the participants' rating of their likelihood to re-use the center decreased as the difference between their preference and perception of the center increased. For example, as the difference between the perception of room temperature and the person's preference for room temperature increased, the participant's likelihood to re-use the center decreased. The same observation was made with the correlation analysis of the 12 absolute difference variables with the rating of the effect the center's environment may have had on NPTE performance. Statistically significant but weak negative correlations were observed between each of the 12 absolute difference variables and the rating on effect on performance. The two room size variables demonstrated a slightly stronger correlation (r = -.33 and -.36) with the rating of effect on performance. For example, as the difference between the perception of room size and the person's preference for room size increased, the participant's rating of the effect that the center on their perception decreased (had more of an effect on performance). A statistically significant strong positive correlation was observed between the likelihood that a person would re-use a center and their rating on the effect that the center may have had on performance. This would indicate that the more the participant believed that the center environment prevented the person from performing at his or her best, the less likely the person would be to choose the same testing center again.

Partitioning the Variability of Perceptions of Environmental Characteristics

Two-level models were built using HLM 6. Analysis began with an unconditional model (no level-1 or level-2 predictors are included) to look at within and between center variability on the perceptions of each of the 12 environmental characteristics. These characteristics were: room size/layout (intimate-spacious, enclosed-open), room climate (cold-hot, dry-humid), room light (dim-bright, soft-intense), room sound (quiet-loud, muffled-clear), workstation chair (soft-hard, not adjustable-highly adjustable), and workstation desktop area (small-large, not adjustable-highly adjustable). These center characteristics were measured through the examinees' perceptions of the test environment in which they took the exam. An intraclass correlation coefficient (ICC) was calculated for each of the environmental characteristics for perceptions (Table 4.8) using a two-level unconditional models in which participants (level-1) were nested within testing centers (level-2). This baseline model allowed for partitioning of the total variability, and identified the amount of variability between centers and the amount within centers.

$$\text{ICC} = \frac{\text{Between center variability}}{\text{Between center variability + within center variability}}$$

Intraclass correlation coefficients of zero indicate that there is no variability between centers. High ICCs indicate that there is heterogeneity between centers and heterogeneity within centers.

Intraclass correlation coefficients for the perceptions of the 12 environmental characteristics ranged from .001 (Light: dim/bright; Chair: not adjustable/Highly adjustable; Station: not adjustable/Highly adjustable) to .111 (Sound: quiet/loud). The mean ICC was .032 and the median ICC was .025. Four environmental characteristics were greater than .050 (Temperature: dry/humid; Temperature: cold/hot; Sound: muffled/clear; Sound: quiet/loud).

Partitioning the Variability of Absolute Differences Between Perceptions and Preferences

As with the analysis of perceptions of environmental characteristics, two-level models were built using HLM 6. Analysis began with an unconditional model (no level-1 or level-2 predictors are included) to look at within and between center variability on the absolute difference between participants' perception and preference for environmental characteristics. Intraclass correlation coefficients for the absolute difference scores between perceptions and preferences for the 12 environmental characteristics ranged from .000 (Light: dim/bright) to .276 (Size: enclosed/open). The mean ICC was .078 and the median ICC was .060. Seven environmental characteristics were greater than .050 (see Table 4.8).

Partitioning the Variability of Perception of Effect On Performance and Likelihood of Choosing the Same Center

As with the analysis of perceptions of environmental characteristics, two-level models were built using HLM 6. Analysis began with an unconditional model (no level-1 or level-2 predictors are included) to look at within and between center variability of the effect on performance of the environmental characteristics. The ICCs for the effect on performance variable and likelihood of choosing the same center again if they needed to retake the NPTE were .004 and .059, respectively.

Table 4.8. Interclass Correlation Coefficients (ICCs) and Reliabilities of Perception, Preference and Absolute Difference Variable for Center Characteristics

	Perceptions	Preferences	Abs Difference
Dependent variable	ICC Reliabilty	ICC Reliability	ICC Reliability
Size (small-spacious)	.101 (.166)	.249 (.345)	.038 (.068)
Size (enclosed-open)	.025 (.046)	.164 (.283)	.280 (.381)
Temp (cold-hot)	.063 (.108)	.004 (.008)	.095 (.156)
Temp (dry-humid)	.055 (.097)	.102 (.162)	.245 (.343)
Light (dim-bright)	.002 (.003)	.138 (.204)	.000 (.001)
Light (soft-intense)	.027 (.050)	.047 (.082)	.001 (.002)
Sound (quiet-loud)	.111 (.173)	.006 (.011)	.056 (.107)
Sound (muff-clear)	.119 (.187)	.090 (.159)	.007 (.013)
Chair (soft-hard)	.063 (.109)	.011 (.021)	.002 (.004)
Chair (not adj-high adj)	.002 (.003)	.019 (.036)	.066 (.109)
Station (sm-large)	.000 (.001)	.022 (.272)	.065 (.112)
Station (not adj-high adj)	.001 (.002)	.191 (.278)	.106 (.165)

Note: Reliability estimates are in parentheses.

Research Question 3:

3a. What is the relationship between examinee characteristics (sex, program GPA, age, online experience, online testing experience, comfort level with online testing, preferred time) and how examinees perceive dimensions of the test environment (room size/layout, climate, lighting, sound, workstation chair and desktop design)?

3b. What is the relationship between examinee characteristics (sex, program GPA, age, online experience, online testing experience, comfort level with online testing, preferred time) and absolute difference scores of examinees' perceptions and preferences of the testing environment (room size/layout, climate, lighting, sound, workstation chair and desktop design)?

3c. What is the relationship between examinee characteristics (sex, program GPA, age, online experience, online testing experience, comfort level with online testing, preferred time) and examinees' perception of the effect on performance and likelihood to use the center again?

The multilevel unconditional models found relatively little variability between centers on the center characteristics that were rated. Much of the variability was within centers. This was consistent for both perception of the centers and the absolute difference scores of examinees' perceptions and preferences of the environmental characteristics. Little between center variability was measured for perception of effect on performance and likelihood to use the center again.

Examinee Characteristics Related to Perceptions of Environmental Characteristics Using Single-Level Analyses

Prior to developing multilevel models, single-level multiple regression analyses were conducted for each of the 12 environmental perception variables. Seven predictor variables (sex, program GPA, age, online experience, online testing experience, comfort level with online testing, and preferred testing time) were entered into each model. These seven predictor variables accounted for < 10% of the variability found in each of the outcome variables (sample sizes for these models ranged from 135 to 166). R-square values ranged from .01 to .10, and none was statistically significant. R-square values, unstandardized regression coefficients and standard errors for each of the 12 models are provided in Table 4.9. None of the predictor variable coefficients were statistically significant for any of the 12 models. These coefficients were very small to small in size, ranging from 0 to .24.

Examinee Characteristics Related to Perceptions of Environmental Characteristics Using Multilevel Analyses

Two-level multilevel modeling with examinees (level-1) nested within centers (level-2) was conducted to further explore the individual characteristics and how they explain the variance within centers. Level-1 models were generated for each of the environmental perception variables using the participant characteristics as predictor variables (Table 4.10). No level 2 (center) predictors were included in the analysis at this point. Results from the multilevel analyses were very similar to those from the single-level multiple regression analyses. Coefficients ranged from 0.00 to 0.89 and only the coefficient for the predictor “online test experience” was statistically significant (p<.001) for the outcome variable room temperature (cold/hot). The regression coefficient of -0.83 indicated that those with more online testing experience tended to perceive the testing environment as colder.

Table 4.9. Regression Model Summary of Participants' Environmental Perceptions

Predictor Variable	Dependent Variable							
	Size (small-spacious)		Size (enclosed-open)		Temp (cold-hot)		Temp (dry-humid)	
	r	*B*	*r*	*B*	*r*	*B*	*r*	*B*
Sex [a]	.10	-0.24 (0.29)	.06	0.19 (0.31)	-.10	-0.19 (.23)	-.06	-0.11 (0.22)
GPA	.03	0.04 (0.44)	-.03	-0.50 (0.47)	.07	0.46 (.35)	.06	0.26 (0.32)
Age	-.06	-0.02 (0.03)	-.21	-0.07 (0.03)	-.05	0.00 (.20)	.08	0.02 (0.02)
Online [b]	.05	0.09 (0.24)	.01	-0.07 (0.26)	-.01	0.01 (.19)	.04	0.10 (0.18)
Online-test [c]	-.15	-0.51 (0.26)	-.11	-0.32 (0.28)	-.24	-0.60 (.20)	-.13	0.38 (0.19)
Comfort [d]	.02	0.12 (0.12)	-.03	0.05 (0.13)	-.15	-0.16 (.09)	.08	0.08 (0.09)
Preferred time [e]	-.00	-0.03 (0.10)	.00	-0.02 (0.11)	.04	0.03 (.08)	.10	0.09 (0.02)
R^2		.05		.07		.10		.06
n		*166*		*164*		*164*		*164*

Note. None of the coefficients was statistically significant (all *p*-levels were > .001). Standard errors are in parentheses.
[a] Sex (1 = Male, 2 = Female); [b] Online – percentage of coursework 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%;[c] Online test – percentage of testing 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%; [d] Comfort – online testing 1 = very comfortable, 5 = very uncomfortable; [e] preferred time for taking tests 1 = early morning (before 10 am), 2 = late morning (10 am – noon); 3 = early afternoon (after noon but before 3 pm); 4 = late afternoon (3 pm or after).

Table 4.9 (Continued) Regression Model Summary of Participants' Environmental Perceptions

Predictor Variable	Dependent Variable: Light (dim-bright)		Light (soft-intense)		Sound (quiet-loud)		Sound (muff-clear)	
	r	*B*	*r*	*B*	*r*	*B*	*r*	*B*
Sex [a]	.06	0.16 (.24)	-.04	0.03 (0.22)	.06	0.25 (0.28)	-.15	-0.65 (0.31)
GPA	-.07	-0.34 (.35)	-.04	0.14 (0.33)	.02	-0.00 (0.42)	-.04	-0.37 (0.46)
Age	.00	-0.00 (.02)	.12	0.04 (0.02)	-.09	-0.01 (0.03)	-.09	-0.05 (0.03)
Online [c]	-.02	-0.03 (.19)	.07	0.20 (0.18)	.24	0.66 (0.24)	-.01	-0.15 (0.25)
Online-test [d]	-.02	-0.05 (.20)	.02	0.02 (0.19)	.05	0.08 (0.25)	.07	0.34 (0.27)
Comfort [e]	.04	0.06 (.09)	-.19	-0.23 (0.09)	-.05	-0.10 (0.12)	.00	0.08 (0.12)
Preferred time [f]	-.05	-0.04 (.08)	.05	0.04 (0.08)	-.14	-0.20 (0.10)	.08	0.10 (0.11)
R^2		.01		.07		.09		.05
n		*162*		*161*		*163*		*164*

Note. None of the coefficients was statistically significant (all *p*-levels were > .001). Standard errors are in parentheses.
[a] Sex (1 = Male, 2 = Female); [b] Online – percentage of coursework 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%;[c] Online test – percentage of testing 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%; [d] Comfort – online testing 1 = very comfortable, 5 = very uncomfortable; [e] preferred time for taking tests 1 = early morning (before 10 am), 2 = late morning (10 am – noon); 3 = early afternoon (after noon but before 3 pm); 4 = late afternoon (3 pm or after).

Table 4.9 (Continued) Regression Model Summary of Participants' Environmental Perceptions

	Dependent Variable							
Predictor Variable	Chair (soft-hard)		Chair (not adj-high adj)		Station (sm-large)		Station (not adj-high adj)	
	r	*B*	*r*	*B*	*r*	*B*	*r*	*B*
Sex [a]	-.06	-0.11 (0.26)	-.06	-0.26 (0.37)	-.08	-0.29 (0.26)	-.15	-0.74 (0.37)
GPA	.07	0.61 (0.39)	-.10	-0.99 (0.59)	.02	-0.05 (0.39)	-.01	-0.09 (0.55)
Age	.08	0.04 (0.02)	-.15	-0.06 (0.03)	-.05	-0.02 (0.02)	-.01	-0.01 (0.03)
Online [b]	.05	0.15 (0.22)	.06	0.16 (0.37)	.06	0.19 (0.21)	.01	0.00 (0.31)
Online-test [c]	-.01	-0.06 (0.23)	-.16	-0.57 (0.35)	-.13	-0.38 (0.23)	.03	0.11 (0.36)
Comfort [d]	-.09	-0.17 (0.11)	-.02	0.11 (0.16)	.08	0.14 (0.11)	.03	0.11 (0.15)
Preferred time [e]	.01	0.00 (0.09)	.00	0.02 (0.03)	.02	0.03 (0.09)	.10	0.15 (0.13)
R^2		.04		.07		.04		.04
n		*162*		*135*		*163*		*149*

Note. None of the coefficients was statistically significant (all *p*-levels were > .001). Standard errors are in parentheses.
[a] Sex (1 = Male, 2 = Female); [b] Online – percentage of coursework 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%; [c] Online test – percentage of testing 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%; [d] Comfort – online testing 1 = very comfortable, 5 = very uncomfortable; [e] preferred time for taking tests 1 = early morning (before 10 am), 2 = late morning (10 am – noon); 3 = early afternoon (after noon but before 3 pm); 4 = late afternoon (3 pm or after).

Table 4.10. Multilevel Models for Participants' Environmental Perceptions

	Fixed Effects			
	Dependent Variable			
Predictor Variable	Size (small-spacious)	Size (enclosed-open)	Temp (cold-hot)	Temp (dry-humid)
	B	*B*	*B*	*B*
Sex	-0.10 (0.27)	0.15 (0.29)	-0.15 (0.22)	-0.13 (0.20)
GPA	0.10 (0.43)	-0.39 (0.46)	0.43 (0.35)	0.32 (0.32)
Age	-0.01(0.03)	-0.07 (0.03)	0.01 (0.02)	0.02 (0.02)
Online	0.18 (0.23)	0.04 (0.24)	0.19 (0.18)	0.25 (0.17)
Online-test	-0.34 (0.27)	-0.36 (0.30)	-0.83 (0.22)**	-0.46 (0.20)
Comfort	0.10 (0.11)	0.07 (0.12)	-0.16 (0.09)	0.03 (0.08)
Preferred time	-0.25 (0.15)	0.06 (0.16)	0.04 (0.12)	0.07 (0.12)
	Random Effects			
Intercept Variance (tau 00) (between Center)	0.21	0.03	0.12	0.07
Level-1 Variance (σ^2) (within Center)	1.90	2.23	1.19	1.04

Note. ** Statistically significant at $p < .001$

Table 4.10 (Continued) Multilevel Models for Participants' Environmental Perceptions

	Fixed Effects Dependent Variable			
Predictor Variable	Light (dim-bright) *B*	Light (soft-intense) *B*	Sound (quiet-loud) *B*	Sound (muff-clear) *B*
Sex	0.26 (0.22)	0.01 (0.20	0.31 (0.28)	0.43 (0.29)
GPA	-0.54 (0.33)	-0.22 (0.31)	0.07 (0.43)	-0.23 (0.45)
Age	0.01 (0.02)	0.05 (0.02)	-0.03 (0.03)	-0.05 (0.03)
Online	-0.03 (0.18)	0.20 (0.17)	0.57 (0.23)	-0.06 (0.24)
Online-test	-0.22 (0.21)	-0.20 (0.20)	0.23 (0.27)	0.35 (0.29)
Comfort	-0.09 (0.08)	-0.12 (0.08)	-0.15 (0.12)	0.10 (0.12)
Preferred time	-0.09 (0.12)	-0.04 (0.11)	-0.29 (0.15)	-0.09 (0.17)
	Random Effects			
Intercept Variance (tau 00) (between Center)	0.00	0.06	0.09	0.12
Level-1 Variance (σ^2) (within Center)	1.17	1.00	1.92	2.08

Note. ** Statistically significant at p < .001

Table 4.10 (Continued) Multilevel Models for Participants' Environmental Perceptions

	Fixed Effects Dependent Variable			
Predictor Variable	Chair (soft-hard) *B*	Chair (not adj-high adj) *B*	Station (sm-large) *B*	Station (not adj-high adj) *B*
Sex	-0.03 (0.25)	0.47 (0.36)	-0.33 (0.24)	-0.57 (0.34)
GPA	0.49 (0.40)	0.21 (0.61)	-.0.00 (0.38)	0.14 (0.53)
Age	0.04 (0.03)	0.09 (0.04)	-0.02 (0.02)	-0.01 (0.03)
Online	0.19 (0.21)	-0.03 (0.36)	0.13 (0.20)	-0.04 (0.28)
Online-test	-0.06 (0.25)	0.89 (0.39)	-0.49 (0.24)	0.20 (0.37)
Comfort	-0.16 (0.10)	-0.36 (0.12)	0.15 (0.10)	0.11 (0.14)
Preferred time	0.11 (0.15)	0.52 (0.21)	-0.16 (0.14)	-0.08 (0.20)
	Random Effects			
Intercept	0.02	0.27	0.00	0.00

Table 4.10 (Continued)

Variance (tau 00) (between Center)				
Level-1 Variance (σ^2) (within Center)	1.68	3.12	1.57	2.84

Note. ** Statistically significant at p < .001

Examinee Characteristics Related to the Absolute Difference Between Participants' Environmental Preferences and Perceptions of Environmental Characteristics using Single-level Analyses

Prior to developing multilevel models, single-level multiple regression analyses were conducted for each of the 12 environmental absolute difference variables. Seven predictor variables (sex, program GPA, age, online experience, online testing experience, comfort level with online testing, and preferred testing time) were entered into each model. These seven predictor variables accounted for < 10% of the variability found in each of the outcome variables (sample sizes for these models ranged from 156 to 193). R-square values ranged from .03 to .10, and only two models were statistically significant ($p < .01$). The two models were room light dim/bright and room light soft/intense. R-square values, unstandardized regression coefficients and standard errors for each of the 12 models are provided in Table 4.11. Two of the predictor variable coefficients were statistically significant for the 12 models. Comfort with online testing was statistically significant ($B = -.39$, $p < .01$) for Room size (small/spacious) and Online test experience was statistically significant (B = .59, $p < .001$) for Room light (dim/bright). Overall coefficients were very small to large in size, ranging from 0 to .85.

Examinee Characteristics Related to Absolute Difference Between Preference and Perception of Environmental Characteristics Using Multilevel Analyses.

Two-level multilevel modeling with examinees (level-1) nested within centers (level-2) was conducted to further explore the individual characteristics and how they explain the variance of the absolute difference variable within centers. A level-1 model was generated for each of these absolute difference variables using the participant

Table 4.11 Regression Model Summary of Absolute Difference Between Preference and Perception

Predictor Variable	Size (small-spacious)		Size (enclosed-open)		Temp (cold-hot)		Temp (dry-humid)	
	r	*B*	*r*	*B*	*r*	*B*	*r*	*B*
Sex [a]	- .04	-0.04 (.31)	-.04	-0.06 (0.32)	.13	0.33 (.21)	.06	0.16 (0.14)
GPA	- .16	-0.64 (.48)	.01	0.53 (0.50)	-.00	-0.14 (.33)	-.18	-0.45 (0.21)
Age	-.06	0.03 (.03)	.12	0.07 (0.03)	-.02	-0.00 (.20)	.05	0.01 (0.01)
Online [b]	.07	0.32 (.26)	.07	0.32 (0.27)	.16	0.33 (.18)	.01	0.06 (0.12)
Online-test [c]	-.01	-0.02 (.29)	.02	0.03 (0.31)	.20	0.46 (.20)	-.04	-0.08 (0.13)
Comfort [d]	-.23	-0.39 (.13)**	-.13	-0.28 (0.13)	.00	-0.28 (.09)	-.11	-0.06 (0.06)
Preferred time [e]	-.06	-0.13 (.11)	.04	0.04 (0.12)	-.11	-0.12 (.08)	.04	0.02 (0.05)
R^2		.08		.05		.08		.05
n		*193*		*191*		*191*		*190*

Note. ****S*tatistically significant p > .001. ** Statistically significant p < .01 Standard errors are in parentheses.
[a] Sex (1 = Male, 2 = Female); [b] Online – percentage of coursework 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%;[c] Online test – percentage of testing 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%; [d] Comfort – online testing 1 = very comfortable, 5 = very uncomfortable; [e] preferred time for taking tests 1 = early morning (before 10 am), 2 = late morning (10 am – noon); 3 = early afternoon (after noon but before 3 pm); 4 = late afternoon (3 pm or after).

Table 4.11 (Continued) Regression Model Summary of Absolute Difference Between Preference and Perception

	Dependent Variable							
Predictor Variable	Light (dim-bright)		Light (soft-intense)		Sound (quiet-loud)		Sound (muff-clear)	
	r	*B*	*r*	*B*	*r*	*B*	*r*	*B*
Sex [a]	.16	0.39 (.19)	-.01	-0.01 (0.21)	.08	0.31 (0.27)	.08	.37 (0.27)
GPA	-.09	-0.41 (.28)	-.09	-0.34 (0.32)	-.01	-0.04 (0.41)	.16	.16 (0.42)
Age	-.02	-0.01 (.02)	.19	0.05 (0.02)	-.03	0.01 (0.02)	.05	.03 (0.03)
Online [c]	-.04	-0.17 (.16)	.13	0.32 (0.17)	.22	0.67 (0.23)	.04	.18 (0.23)
Online-test [d]	.23	0.59 (.17)***	.19	0.39 (0.19)	.09	0.20 (0.25)	-.04	-.21 (0.26)
Comfort [e]	-.06	-0.07 (.08)	.04	0.01 (0.08)	-.10	-0.18 (0.11)	-.08	-.15 (0.11)
Preferred time [f]	.01	-0.13 (.11)	-.08	-0.08 (0.07)	-.05	-0.08 (0.10)	-.01	-.01 (0.10)
R^2		.10**		.09**		.07		.03
n		*188*		*187*		*189*		*191*

Note. ****S*tatistically significant p > .001. ** Statistically significant p < .01 Standard errors are in parentheses.
[a] Sex (1 = Male, 2 = Female); [b] Online – percentage of coursework 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%;[c] Online test – percentage of testing 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%; [d] Comfort – online testing 1 = very comfortable, 5 = very uncomfortable; [e] preferred time for taking tests 1 = early morning (before 10 am), 2 = late morning (10 am – noon); 3 = early afternoon (after noon but before 3 pm); 4 = late afternoon (3 pm or after).

Table 4.11 (Continued) Regression Model Summary of Absolute Difference Between Preference and Perception

	Dependent Variable							
Predictor Variable	Chair (soft-hard)		Chair (not adj-high adj)		Station (sm-large)		Station (not adj-high adj)	
	r	*B*	*r*	*B*	*r*	*B*	*r*	*B*
Sex [a]	-.04	-0.07 (0.25)	.12	0.59 (0.36)	.03	0.17 (0.30)	.15	-0.85 (0.39)
GPA	-.05	0.10 (0.39)	-.08	-0.17 (0.60)	-.05	-0.05 (0.47)	-.09	-0.49 (0.60)
Age	.13	0.04 (0.02)	.19	0.08 (0.03)	.03	0.02 (0.03)	.05	0.03 (0.04)
Online [b]	-.03	-0.04 (0.21)	-.04	-0.02 (0.37)	.01	0.00 (0.26)	.00	0.08 (0.33)
Online-test [c]	.02	0.09 (0.24)	.19	0.76 (0.36)	.07	0.31 (0.29)	.05	0.23 (0.40)
Comfort [d]	-.14	-0.21 (0.10)	-.09	-0.27 (0.16)	-.16	-0.29 (0.12)	-.13	-0.30 (0.16)
Preferred time [e]	-.01	-0.03 (0.09)	.05	0.10 (0.14)	.01	-0.01 (0.11)	-.05	-0.08 (0.14)
R^2		.04		.09		.04		.06
n		*188*		*156*		*190*		*170*

Note. ***Statistically significant p > .001. ** Statistically significant p < .01 Standard errors are in parentheses.
[a] Sex (1 = Male, 2 = Female); [b] Online – percentage of coursework 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%; [c] Online test – percentage of testing 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%; [d] Comfort – online testing 1 = very comfortable, 5 = very uncomfortable; [e] preferred time for taking tests 1 = early morning (before 10 am), 2 = late morning (10 am – noon); 3 = early afternoon (after noon but before 3 pm); 4 = late afternoon (3 pm or after).

characteristics. Coefficients ranged from 0.00 to 0.89. The coefficient for the predictor "online test experience" was statistically significant (*B =0.68, p*<.001) for the outcome variable room lighting (dim/bright). Lastly, the intercept variance for the model for room size (enclosed/open) demonstrated statistical significance (*p*<.001) demonstrating between center variance for only this outcome variable. However, no predictor variable coefficients were statistically significant for this model (Table 4.12).

Examinee Characteristics Related to Perceptions of Effect On Performance and Use of Center Again Using Single-Level Analyses.

Prior to developing multilevel models, single-level multiple regression analyses were conducted for the variable 'effect on performance' and 'center again'. Seven predictor variables (sex, program GPA, age, online experience, online testing experience, comfort level with online testing, and preferred testing time) were entered into each model. These seven predictor variables accounted no more than 10 to 12% of the variability found in each of the outcome variables (sample size for these models were 183 and 186, respectively). R-square values, unstandardized regression coefficients and standard errors for the two models are provided in Table 4.13: Regression Model Summary. Both models were statistically significant at the $p < .01$ level. When effect on performance was used in the regression as an outcome variable, there were two statistically significant coefficients ($p < .01$). Age ($B = -.07$, $p < .01$) and comfort with online testing ($B = .34$, $p < .01$) were significant for explaining some of the variance observed in the outcome variable of 'effect on performance', however the coefficients were very small to small. When 'center again' was used in the regression as an outcome variable, the coefficient for comfort with online testing was statistically significant ($B = .39$, $p < .01$) and most of the coefficients were very small. These findings for both 'effect on performance' and 'center again' are similar to the findings for the

Table 4.12. Multilevel Models for Absolute Difference Between Preference and Perception

	Fixed Effects			
	Dependent Variable			
Predictor Variable	Size (small-spacious)	Size (enclosed-open)	Temp (cold-hot)	Temp (dry-humid)
	B	*B*	*B*	*B*
Sex	0.01 (0.31)	-0.07 (0.31)	0.42 (0.21)	0.12 (0.14)
GPA	-0.75 (0.49)	0.49 (0.52)	-0.24 (.34)	-0.41 (0.22)
Age	0.01 (0.03)	0.06 (0.03)	-0.01 (0.02)	0.01 (0.01)
Online	0.30 (0.26)	0.41 (0.27)	0.31 (0.18)	0.09 (0.11)
Online-test	0.00 (0.31)	-0.19 (0.31)	0.57 (0.22)	-0.09 (0.14)
Comfort	-0.37 (0.13)	-0.19 (0.13)	0.01 (0.09)	-0.04 (.06)
Preferred time	-0.02 (0.17)	-0.09 (0.18)	0.08 (0.12)	0.10 (0.08)
	Random Effects			
Intercept Variance (tau 00) (between Center)	0.09	0.71**	0.14	0.10
Level-1 Variance (σ^2) (within Center)	2.26	2.25	1.15	0.43

Note. ** Statistically significant at p < .001

Table 4.12 (Continued). Multilevel Models for Absolute Difference Between Preference and Perception

	Fixed Effects Dependent Variable			
Predictor Variable	Light (dim-bright) *B*	Light (soft-intense) *B*	Sound (quiet-loud) *B*	Sound (muff-clear) *B*
Sex	0.37 (0.20)	-0.06 (0.21)	0.45 (0.27)	0.51 (0.27)
GPA	-0.40 (0.30)	-0.37 (0.33)	-0.25 (0.43)	-0.09 (0.43)
Age	-0.00 (0.02)	0.05 (0.02)	-0.01 (0.03)	0.02 (0.03)
Online	-0.20 (0.16)	0.34 (0.17)	0.66 (0.23)	0.19 (0.23)
Online-test	0.68 (0.19)**	0.39 (0.21)	0.29 (0.27)	-0.09 (0.27)
Comfort	-0.08 (0.08)	0.01 (0.08)	-0.12 (0.11)	-0.11 (0.11)
Preferred time	0.07 (0.11)	-0.03 (0.12)	-0.22 (0.15)	-0.19 (0.16)
	Random Effects			
Intercept Variance (tau 00) (between Center)	0.00	0.01	0.12	0.02
Level-1 Variance (σ^2) (within Center)	0.96	1.13	1.85	1.99

Note. ** Statistically significant at $p < .001$

Table 4.12 (Continued). Multilevel Models for Absolute Difference Between Preference and Perception.

Predictor Variable	Chair (soft-hard) *B*	Chair (not adj-high adj) *B*	Station (sm-large) *B*	Station (not adj-high adj) *B*
	Fixed Effects (Dependent Variable)			
Sex	0.10 (0.25)	0.47 (0.36)	0.19 (0.30)	0.84 (0.40)
GPA	0.00 (0.39)	0.21 (0.61)	-0.04 (0.47)	-0.61 (0.63)
Age	0.04 (0.03)	0.09 (0.04)	0.01 (0.03)	0.03 (0.04)
Online	-0.04 (0.21)	-0.03 (0.36)	0.01 (0.25)	0.08 (0.33)
Online-test	0.19 (0.25)	0.89 (0.39)	0.45 (0.30)	0.11 (0.44)
Comfort	-0.19 (0.10)	-0.36 (0.16)	-0.30 (0.12)	-0.29 (0.16)
Preferred time	0.04 (0.15)	0.52 (0.21)	0.22 (0.17)	0.11 (0.23)
	Random Effects			
Intercept Variance (tau 00) (between Center)	0.00	0.27	0.07	0.18
Level-1 Variance (σ^2) (within Center)	1.68	3.12	2.38	3.81

Note. ** Statistically significant at p < .001

Table 4.13. Regression Model Summary of Participants' Environmental Perceptions for Effect on Performance and Center Again

	Dependent Variable			
Predictor Variable	Effect on Performance		Center Again	
	r	*B*	*r*	*B*
Sex [a]	-.05	-0.34 (0.26)	-.07	-0.43 (0.31)
GPA	.12	0.28 (0.41)	.11	0.34 (0.48)
Age	-.16	-0.07 (0.03)**	-.13	-0.06 (0.03)
Online [b]	-.12	-0.44 (0.22)	-.11	-0.46 (0.27)
Online-test [c]	-.10	-0.21 (0.26)	-.09	-0.29 (0.29)
Comfort [d]	.20	0.34 (0.12)**	.20	0.39 (0.13)**
Preferred time [e]	-.02	-0.08 (0.15)	.04	0.09 (0.11)
R^2		.12**		.10**
n		*183*		*186*

Note. *** Statistically significant at $p < .001$. ** Statistically significant at $p < .01$. Standard errors are in parentheses.

[a] Sex (1 = Male, 2 = Female); [b] Online – percentage of coursework 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%; [c] Online test – percentage of testing 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%; [d] Comfort – online testing 1 = very comfortable, 5 = very uncomfortable; [e] preferred time for taking tests 1 = early morning (before 10 am), 2 = late morning (10 am – noon); 3 = early afternoon (after noon but before 3 pm); 4 = late afternoon (3 pm or after). Effect Performance (1 = prevented from performing at best, 7 = allowed to perform at best); Center Again (1 = not likely, 7 = highly likely).

single level analysis conducted for the perceptions of the environment.

Examinee Characteristics Related to Perceptions of Effect On Performance and Use of Center Again Using Multilevel Level Analyses

Two-level multilevel modeling with examinees (level-1) nested within centers (level-2) was conducted to further explore the individual characteristics and how they explain the variance of the variables ‘effect on performance’ and ‘center again’. A level-1 model was generated for each of these two variables using the participant characteristics. Coefficients ranged from 0.07 to 0.53. The coefficient for the predictor variable ‘age’ was statistically significant in both models ($B = -.07$ *and* $B = -.08$, *respectively,* $p < .001$). The coefficient for the predictor variable ‘comfort’ was also statistically significant in both models ($B = .33$ *and* $B = .40$, *respectively,* $p < .001$). However, the intercept variance for each of the two models did not demonstrate statistical significance ($p < .001$) indicating little to no between center variance for these two outcome variables.

Table 4.15 provides a summary of significant individual characteristic predictors ($p < .05$) for the outcome variables comparing single level analysis and multilevel analysis where these data are nested within centers.

Table 4.14. Multilevel Models for Effect on Performance and Would Use the Center Again.

	Fixed Effects	
	Dependent Variable	
Predictor Variable	Effect Performance	Center Again
	B	*B*
Sex	-0.31 (0.27)	-0.37 (0.31)
GPA	0.35 (0.42)	0.53 (0.50)
Age	-0.07 (0.03)**	-0.08 (0.03)**
Online	-0.45 (0.22)	-0.50 (0.27)
Online-test	-0.22 (0.27)	-0.35 (0.31)
Comfort	.33 (0.11)**	0.40 (0.13)**
Preferred time	-0.09 (0.15)	-0.21 (0.18)
	Random Effects	
Intercept Variance (tau 00) (between Center)	0.06	0.36
Level-1 Variance (σ^2) (within Center)	1.85	2.35

Note. ** Statistically significant at p < .01 *** Statistically significant at p <.001
[a] Sex (1 = Male, 2 = Female); [b] Online – percentage of coursework 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%;[c] Online test – percentage of testing 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%; [d] Comfort – online testing 1 = very comfortable, 5 = very uncomfortable; [e] preferred time for taking tests 1 = early morning (before 10 am), 2 = late morning (10 am – noon); 3 = early afternoon (after noon but before 3 pm); 4 = late afternoon (3 pm or after). Effect Performance (1 = prevented from performing at best, 7 = allowed to perform at best); Center Again (1 = not likely, 7 = highly likely).

Table 4.15. Summary of Significant Predictors for Research Question Three Outcome Variables Using Single-level and Multilevel Analysis

Outcome Variable		Predictor Variables						
Room size		**Sex**	**GPA**	**Age**	**Online Exp**	**Online Test**	**Online Comfort**	**Preferred Time**
Intimate[a]/Spacious[b]	Perception						**ML(-)**	**SL(-)**
	|Difference|						**ML**	
Enclosed[a]/Open[b]	Perception			**SL (-) ML (-)**				
	|Difference|			**ML**			**ML**	
Room temperature								
Cold[a]/Hot[b]	Perception					**SL(-) ML(-)**		
	|Difference|					**SL ML**		
Dry[a]/Humid[b]	Perception					**SL(-) ML(-)**		
	|Difference|							
Room light								
Dim[a]/Bright[b]	Perception							
	|Difference|	**ML**				**SL ML**		
Soft[a]/Intense[b]	Perception			**SL ML**				
	|Difference|			**ML**				
Room sound								
Quiet[a]/Loud[b]	Perception				**SL ML**			
	|Difference|				**SL ML**			
Muffled[a]/Clear[b]	Perception							
	|Difference|							
Workstation chair								
Soft[a]/Hard[b]	Perception						**ML**	
	|Difference|					**ML**	**ML**	

No Adjust[a]/High Adjust[b]	Perception		**SL(-) ML**		**ML**		**ML**
	|Difference|	**SL ML**	**SL ML**		**ML**	**ML(-)**	**SL**
Workstation desktop							
Small[a]/Large[b]	Perception				**ML(-)**		
	|Difference|					**ML**	
No Adjust[a]/High Adjust[b]	Perception	**ML**					
	|Difference|						
Effect Performance			**SL(-) ML(-)**	**ML(-)**		**SL ML**	
Center Again			**SL(-) ML(-)**			**SL ML**	

Note. SL = single level analysis, ML = multilevel analysis, $p < .05$

Research Question 4

4a.What is the relationship between center characteristics (room size, room density, center 'newness', presence of ambient light, break space, and access to food/drink) and examinees' perceptions of the testing environment (room size/layout, climate, lighting, sound, workstation chair and desktop design)?

4b. What is the relationship between center characteristics (room size, room density, center 'newness', presence of ambient light, break space, and access to food/drink) and the absolute difference between examinees' preferences and perceptions of the testing environment (room size/layout, climate, lighting, sound, workstation chair and desktop design)?

4c What is the relationship between center characteristics (room size, room density, center 'newness', presence of ambient light, break space, and access to food/drink) and examinees' perceptions of the effect on performance?

Center level data analyses were conducted to address these research questions. Because these were individuals' perception (ratings) of the center, it was important to identify the reliability of those ratings. ICCs of the center-level variables by center were run to determine the reliability of the ratings on these center characteristics. Table 4.16 shows the ICCs for the seven center characteristics.

Table 4.17 presents the statistically significant relationships between the center-level predictor variables and outcome variables using multilevel modeling (examinees nested in centers). This includes the outcome variables of perception of the testing environment (actual), absolute difference between participants' preferences and perceptions of the testing environment (absolute) and the examinee's perceptions of the effect the environment had on performance. Only level-2 predictors (nested data) were entered in the equations for this analysis.

Table 4.16. Intraclass Correlation Coefficients of Center-level Variables by Center

Variable	ICC	Reliability
Room size	.353	(.457)
Room density	.032	(.057)
Center newness	.384	(.069)
Ambient light	.134	(.211)
Visual distractions	.004	(.007)
Break room	.238	(.336)
Access food/drink	.013	(.182)

Note. Reliability estimates are in parentheses.

Table 4.17. Summary of Significant Predictors for Outcome Variables using Single Level and Multilevel Analysis

Outcome Variable						
Room size		**Room Size**	**Room Density**	**Window**	**Center New**	**Break Food**
Intimate[a]/Spacious[b]	Perception	**SL**	**SL (-)**		**SL**	
	|Difference|		**SL**		**SL (-)** **ML (-)**	
Enclosed[a]/Open[b]	Perception		**SL (-)** **ML (-)**	**SL**		
	|Difference|		**SL** **ML**		**SL (-)**	
Room temperature						
Cold[a]/Hot[b]	Perception					
	|Difference|					
Dry[a]/Humid[b]	Perception					
	|Difference|					
Room light						
Dim[a]/Bright[b]	Perception					**SL** **ML**
	|Difference|				**SL (-)**	
Soft[a]/Intense[b]	Perception	**SL (-)** **ML (-)**				
	|Difference|			**ML**	**SL (-)**	
Room sound						
Quiet[a]/Loud[b]	Perception				**SL (-)** **ML (-)**	
	|Difference|				**ML (-)**	
Muffled[a]/Clear[b]	Perception					
	|Difference|					
Workstation chair						
Soft[a]/Hard[b]	Perception				**SL (-)**	
	|Difference|					**ML**
No Adjust[a]/High Adjust[b]	Perception		**ML (-)**		**SL**	
	|Difference|					**SL**
Workstation desktop						
Small[a]/Large[b]	Perception				**SL**	
	|Difference|				**ML (-)**	
No Adjust[a]/High Adjust[b]	Perception				**SL**	
	|Difference|					
Effect Performance					**ML**	
Center Again					**ML**	

Note: SL = single level analysis, ML = multilevel analysis, $p < .05$

CHAPTER FIVE:

DISCUSSION

The purpose of this study was to assess the testing environment as perceived by individuals taking the National Physical Therapy Examination. This study explored examinees' environmental preferences and perceptions of the environment in which they were administered the NPTE. Data related to examinees' demographic information, academic ability, previous test taking experiences, program characteristics, test administration information, as well as test center characteristics were also collected. These data were collected through the use of an on-line survey instrument that included open and closed-item formats.

Previous research has not explored the relationship between the testing environment and the unique characteristics of the individuals who take the examinations administered in these environments. This exploratory study attempted to establish baseline data on center characteristics, the characteristics of test takers, their preferences for the testing environment, and their experience with the testing environment. The potential relationships between these variables were explored. The research questions that were considered were:

Research Question 1:

1a. What are the environmental preferences for the NPTE testing environment (e.g., room size/layout, climate, lighting, sound, workstation chair and desktop design) of examinees who have taken the NPTE?

1b. What is the relationship between participants' background characteristics (e.g., sex, , program GPA, age, online experience, online testing experience, comfort level with online testing, and preferred testing time) and environmental preferences (e.g., room size/layout, climate, lighting, sound, workstation chair and desktop design)?

Research Question 2:

2a. On the dimensions of the test environment identified (e.g., room size/layout, climate, lighting, sound, workstation chair and desktop design), how much variability of examinees' perceptions of the testing environment exists between testing centers and how much is within testing centers administering the NPTE?

2b. On the dimensions of the test environment identified (e.g., room size/layout, climate, lighting, sound, workstation chair and desktop design), how much variability in the absolute difference scores (i.e., difference in examinees' perceptions and preferences of the testing environment) exists between testing centers and how much is within testing centers administering the NPTE?

2c. How much variability in the "effect on performance" and "use center again" variables exists between testing centers and how much is within testing centers administering the NPTE?

Research Question 3:

3a. What is the relationship between examinee characteristics (sex, degree, program GPA, age, online experience, online testing experience, comfort level with online testing) and how examinees perceive dimensions of the test environment (room size/layout, climate, lighting, sound, workstation chair and desktop design)?

3b. What is the relationship between examinee characteristics (sex, degree, program GPA, age, online experience, online testing experience, comfort level with online testing) and absolute difference scores of examinees' perceptions and preferences of the testing environment (room size/layout, climate, lighting, sound, workstation chair and desktop design)?

3c. What is the relationship between examinee characteristics (sex, degree, program GPA, age, online experience, online testing experience, comfort level with online testing) and perception of effect on performance and likelihood to use the center again?

Research Question 4:

4a.What is the relationship between center characteristics (room size, room density, center 'newness', presence of ambient light, break space, and access to food/drink) and examinees' perceptions of the testing environment (room size/layout, climate, lighting, sound, workstation chair and desktop design)?

4b. What is the relationship between center characteristics (room size, room density, center 'newness', presence of ambient light, break space, and access to food/drink) and the absolute difference between examinees' preferences and perceptions of the testing environment (room size/layout, climate, lighting, sound, workstation chair and desktop design)?

4c What is the relationship between center characteristics (room size, room density, center 'newness', presence of ambient light, break space, and access to food/drink) and examinees' perceptions of the effect on performance and likelihood to use the center again?

Two hundred and sixteen participants completed an online survey that gathered data on participant characteristics (sex, degree, curriculum type, experience with online courses/tests, comfort in taking online tests, and preference for testing time), center characteristics, and participant's preferences and experience with environmental factors in the testing room. Most participants had taken the NPTE within the previous 12-months, with seven of the participants taking the NPTE within 15-months of completing the survey. Participant demographics were representative of the demographic characteristics of graduates of PT programs nationally who would be sitting for the NPTE.

Background to the Study

Industry standards exist for the delivery of computer-based and internet-delivered exams and address the test user. The International Test Commission (ITC) Computer-based and Internet delivered testing guidelines (2005) address accessibility needs of test users from different cultural groups and minimizing barriers (p. 26) and the physical needs of the test taker. Section 3a, number 34-36 (p. 17) makes note of the need for comfort and proper ergonomics of the test user (see Appendix A for details).

Although these guidelines/standards exist, it is unclear what is currently being implemented in testing centers, as well as what oversight there is on ensuring test user comfort and that the environments in which the tests are administered address the ergonomic, lighting, heating, and ventilation standards. There is no apparent monitoring of these factors. Prior to this study, there was also little known of the specific preferences of test users and how they perceive the testing environment. This study provided an initial look into these issues.

This study considered individual test user preferences and experiences for the test environment, sources of distraction, and their perception on whether their performance was affected. Individuals' perceptions of whether they would use the center again, if necessary, were also considered. Through gathering these baseline data, test developers and administrators can begin to analyze the variability that may exist in the conditions under which tests are administered. It may also provide guidance as to the design of these centers and the scheduling of examinations within these centers.

Examinee Characteristics

Gaining a better understanding of the participants' prior experience using online learning and computer-based exams was essential in order to avoid making assumptions about recent graduates' computer comfort levels. The survey results found participants had a surprisingly low amount of experience taking online courses and testing, and a high level of discomfort with taking fully online tests. A ten year analysis of 2800 higher education institutions demonstrated that greater than 6.7 million students, which represents 32 percent of the total higher education enrollment took at least one higher education course online during the Fall 2011 semester. This represents a .6 million student growth from the previous year (http://www.onlinelearningsurvey.com/reports/changingcourse.pdf). Online education participation has grown from 1.6 million students enrolling in at least one course to 7.1 million in 2012 (http://www.onlinelearningsurvey.com/reports/gradechange.pdf). Despite the continually growing amount of online education being offered in higher education, there may be certain fields in higher education that online educational experiences (courses, and in particular, exams) are less utilized. It could be questioned that within the testing industry and licensure agencies, there may be an assumption that current test takers are far more seasoned in their online/computer-based testing experience. Since the computer-based testing experience was already relatively foreign to many test takers in this study, it may be necessary for more attention to be given to the orientation to test takers and to the computer-user interaction. The objective of licensure examinations, particularly in the health professions, must always be on public/consumer protection, and not on the person's ability to overcome the testing process and environment.

Another unknown aspect of test administration is what examinees need with regard to physical and mental breaks from the test taking situation. The NPTE is unique in its length of administration. Examinees have up to five hours to complete the exam, and this amount of time is fully utilized by most examinees. By looking at test takers' behaviors during this long administration period, more information was gathered about these physical and mental needs for 'breaks' during the examination process. In this study, 96% used the allotted break time (exam clock pauses) and left the exam room. Nearly half (42%) found it necessary to take an additional break during which time the exam clock continues to run. Most of those individuals (88%) left the room during this additional break. It should be questioned whether there is a physical need for additional scheduled break time during this long exam where the timer is paused. The addition of a scheduled break may take away testing effects and allow the examinee to perform at a more typical level. Open-ended comments regarding breaks included the distraction of people taking breaks at different times, coming in and out of the room. Several commented on the lack of a dedicated break space where one could relax during a break. Of the 41 comments related to the break space, 30 were neutral in nature and eight were negative. The three positive comments related to break space were also connected to the Prometric center being located in a relatively new facility.

Test taker preferences include the time of day in which they prefer to take exams. Although examinees may have preferences for testing times, test administration may have restrictions on when exams can be offered. The length (five hours) of the NPTE administration creates a need for most exams to be started in earlier times of the day. During this study, the vast majority (> 90%) preferred a morning testing time (start time before noon). Ninety-three percent of the participants reported a morning testing start time. However of the sixty-five percent of participants who were scheduled for an early morning testing time (start time before 10 am), only 40% of those participants reported this time to be their preferred time. These data were collected at a time where the NPTE was being offered using continuous testing scheduling. With exams that are offered at fixed times during the year, there may be even more restrictions as to when exams are scheduled during the day. Although this may benefit some examinees, the fixed date/time may negatively impact others'.

Examinees do not have a great deal of control over their testing environment. One option that may be offered during testing is the use of headphones provided by the testing center. The NPTE test administration protocol allows for the use of headphones. The results of this study demonstrated that there was mixed use of headphones. Approximately 50% of participants reported using headphones at some time during the examination process. Responses on the open ended question provided some insight as to the perceived benefits of headphone use, or the reason participants chose to stop using them. Thirty-nine participants commented on headphones. Twelve made neutral comments on the use of headphones or the availability of them. Positive comments related to the effectiveness of the sound dampening provided by the headphones. However, the majority of the comments were negative related to the use of headphones. Comments included their ineffectiveness, the creation of a “tunnel” effect, the discomfort produced by using the headphones, and even the concern of their cleanliness. A sampling of the comments made regarding use of the headphones is included in Appendix F. It may be considered by testing organizations to explore the type of headphones made available to examinees since room quietness was shown as strong preference by participants in this study.

Center Characteristics

Having direct measures of center characteristics would be the strongest evidence as to the physical environment in which examinees take tests. These measures may include measures of sound levels, light levels, and dimensions of the space and furniture. Being able to record observations of test takers’ movement while working at a workstation, as well as when entering and departing the testing room, would provide other direct measures of the testing environment. However, access to testing centers is restricted to minimize violations to test security. Being able to secure evaluative data reported by the testing centers themselves would also provide another level of evidence regarding testing centers. However, Prometric was not willing to provide or did not have these data on the individual testing centers. It is unclear if customer satisfaction data are collected regularly by Prometric, nor was it available on the company website or by request. Therefore, in this study, the physical environment in testing centers needed to be measured through self-report of examinees. Self-report of room characteristics provided a good baseline for future studies where more direct measures could be sought.

In the initial design of the study, multiple examinees were sought from each center to provide a more reliable assessment of the environmental characteristics of the center. Despite multiple attempts to increase the sample size, 63 of the 103 centers described in this exploratory study were reported on by only one participant. The remainder of the centers represented had two or more participants' data, with a range of 2-16 participants' data for any of these centers. Therefore, these data only allow for an initial, baseline description of a sample of testing centers across the United States.

Individuals' perception of room size and density of the same size varied. Using an indirect measure of how individuals perceive a room allowed for differing ratings in multi-participant centers. However ratings were typically within one category difference. Nearly half the center size mean ratings were "2" meaning that centers were 11-20 workstations in size. The 'density' of the center was examined as a measure of how crowded the center was at the time the participant was present. The majority of the mean ratings of the density of the center demonstrated 0-10 workstations in use (34 centers) or 11-20 workstations in use (37 centers). Comments made for the open ended question were sometimes very specific such as estimating the room dimensions, stating the number of stations and even how the workstations/cubicles were organized within the room. Comments specific to the number of stations ranged from as few as four to as many as 35. Another center provided an individual room for the examinee because "PT testers are known to get up and move around a lot." This is not consistent with the policies for administration of the NPTE, and should require additional inquiry.

At the time of this study, examinees would register for an exam time that ranged from morning to afternoon and various days of the week. Self-reporting of the center characteristics "room size" and "newness" should have less variability than "room density" and "visual distractions" due to the time of day or day of the week during which the participant was using the center. Room size and newness are relatively stable concepts about the size of a room and whether it is new/newly renovated or worn and outdated. The concept of "newness" may also have a personal value to it, with some participants being much more critical of the degree to which things are worn or outdated. However, room density and visual distractions could be affected by time of day, day of week, etc. Some days are busier than others within testing centers, with some days having many different exams being administered than other days. Once again, this may stabilize due to the fact that fixed date/time testing will decrease the variability of volume. However, it is anticipated that the volume of examinees on these fixed dates will be maximized.

Other comments about the testing room related to the ceiling height and color of the walls ("drab" from one participant, needs "lighter colored paint" by another). One participant stated the presence of a wall and the specific color of the wall (blue) and another spoke of "white walls and standard issue grey carpet". A third commented on the "greyish color" carpet and the walls being light blue/grey. The level of detail which examinees recalled the details of the space in which they took the exam may indicate the level of importance that the physical environment has on individuals in this high stakes testing situation.

Research Questions

In addition to exploring the examinee and center characteristics, this study sought to answer four specific questions. Each of those will be now discussed.

Research Questions 1a-1b

These questions provide a description of the environmental preferences for participants taking the NPTE as well as explore the relationship between participants' background characteristics and their environmental preferences. Since little is known about the characteristics of these examinees and the environment in which they prefer to take a high-stakes examinations, the results of this research attempted to develop baseline data.

Research question 1a. What are the environmental preferences for the NPTE testing environment of examinees who have taken the NPTE?

The results of this study demonstrate that there was considerable variability in participants' preferences, with only two of the 12 environmental factors having a large percentage of the respondents in the modal class. No strong preference for either end of the scale was observed for responses to items related to room size (intimate/spacious, enclosed/open), room temperature (cold/hot, dry/humid), room light (soft/intense), and room sound (muffled/clear) which were clustered toward the middle of the scale. Responses to the open-ended question confirmed this variability in preferences for the environmental characteristics of room size, room climate (dry/humid), room light intensity (soft/intense) and the clarity of the room sound by either not being commented upon or being mentioned in a neutral way. Specific comments on lighting demonstrated the most variability in that some thought the lighting was "a little too bright, but I personally like lower lighting in general" and others thought it was consistent for their preference of "not too bright" and not using the florescent lights "which I don't like too much."

The item with the greatest departure from normality was room sound/quiet loud, with the strong preference toward a quiet room. Sound and the presence of noise during the exam was the greatest area of comment on the open ended question (Appendix F). Fifty-six of the respondents who commented on sound reported that the constant typing of test takers taking other writing based examinations and the noise from examinees and staff entering and exiting the room were a negative factor and led to distraction and decreased concentration. This led some to say that it took them additional time to take the exam as they had to reread items. Examinees commented about the staff footwear contributing to the noise (high heels), other examinees reading the items of their exam aloud, and even the presence of a big storm on the day of the exam (which was not attributed to the testing center). There were comments about their location in the room and proximity to the proctor's desk or entrance to the room, and the individuals' proximity to an outside door that allowed sound to pass through. Comments also included the distraction of a ticking clock, people coughing, foot tapping, and the "whine/buzz" of the computer. These comments would indicate the high level of awareness that examinees have during a high stakes examination; awareness that is retained even a period of time after the testing experience.

As stated in Chapter 2, Errett, Bowden, Choiniere, and Wang (2006) considered whether individuals become increasingly more aggravated by background noise the longer it persists or whether individuals habituate to the sound in their environment. They discussed that the exposure to background sound may not have an impact on performance but that the perception of effect was more evident. Bowden and Wang (2005), studying architectural acoustics, observed in individual's data that some subjects were more able to "tune out" the noise in the background than others. Landstrom (2004) concluded that background noise in a silent environment is more detectable and may have a larger impact on a cognitive task. It could be hypothesized that since the testing environment is a relatively silent environment, examinees may be more sensitive to sounds in the environment and therefore more easily distracted by these noises. It would appear that there continues to be much to learn about human factors related to sound in the environment in which individuals function.

Slight preferences were noted for the softness and adjustability of the desk chair, which were supported by the responses to the open ended question. Physical comfort and the need for adjustability because of the long duration of the exam were noted by many respondents. Fifty-six percent of the individuals who commented about the chair commented negatively and focused on the lack of adjustability which led to postural fatigue and even back/neck pain. Several respondents mentioned that the workstation did not allow for the accommodation of their short stature. One participant noted the irony of the fact that the individual was taking test questions regarding ergonomics yet was experiencing back pain due to the chair/workstation arrangement.

Respondents much more strongly preferred a desktop area that had a great deal of adjustability. Based on the open ended responses, this may include the adjustability of the chair height in relation to the keyboard/mouse, the orientation of the screen to the seating position, and accommodating left and right handedness. Forty-seven percent of those who commented on the workstation made negative comments about the workstation at the testing center, while forty-two percent made neutral comments. Ergonomics was mentioned by several of the participants. One respondent commented that the position of the monitor and keyboard required them to sit diagonally in the workstation the entire time. The stated "after five hours of testing with my cervical and thoracic spine in left rotation, I was fairly sore and distracted." Participants in this study may have an elevated awareness of certain human factors related to workstation design due to the fact that the respondents were graduates of physical therapy programs and have been educated about the importance of proper ergonomics. The testing industry may consider the ergonomic literature to identify potential simple modifications that can be made to testing environments to decrease examinee fatigue and discomfort, and allow greater focus on the process of taking the test.

Research question 1b. What is the relationship between participants' background characteristics (e.g., sex, , program GPA, age, online experience, online testing experience, comfort level with online testing, and preferred testing time) and environmental preferences (e.g., room size/layout, climate, lighting, sound, workstation chair and desktop design)?

There were no statistically significant relationships between the 12 environmental preference variables and the seven participant background characteristic variables. No statistically significant predictors were identified using both single-level multiple regression analysis and multilevel analysis for nested data. It would appear that preferences are unique to an individual and not predictable based on the background of the person. The generalizability of this finding was limited since the sample represented a volunteer sample. It does not appear that similar research has been conducted and therefore a comparison of this research and previous research findings cannot be made. These baseline data may demonstrate the variability of preferences for environmental factors and indicate a challenge for test administrators to meet the needs of their population served. While it is impossible to meet the needs of all individuals using a physical space, more attention to this aspect of testing may be warranted.

Research Questions 2a-2c

These questions explored how different examinees perceptions were of the Prometric testing centers when looking at 12 environmental dimensions as well as the center's effect on their testing performance and whether they would test in the center again. The variability of perception scores between and within testing centers provided a baseline view of the stability of these self-report ratings of the environment.

Research question 2a. On the dimensions of the test environment identified (e.g., room size/layout, climate, lighting, sound, workstation chair and desktop design), how much variability of examinees' perceptions of the testing environment exists between testing centers and how much is within testing centers administering the NPTE?

This study explored participants' perceptions of the testing environment in which they took the NPTE. Descriptive statistics demonstrated that ratings were toward the middle of the scale on each of the 12 perception variables, with mean scores ranging from 2.81 to 3.89 on the 1 to 7 scale. Further discussion of these data for center characteristics is found in the results section for research question 4.

An intraclass correlation coefficient (ICC) was calculated for each of the environmental characteristic variables for participants' preferences, perceptions, and the absolute difference between these two ratings (Chapter 4,Table 4.8) The ICCs were calculated from two-level unconditional models (i.e., no level-1 or level-2 predictors) in which participants (level-1) were nested within testing centers (level-2). The unconditional model provided a baseline model to analyze the total variability in preferences, perceptions, and absolute difference for each of these variables. ICCs helped to identify the amount of total variability between centers and the amount of variability within centers. This exploratory study was one of the first studies that have looked at ICCs to determine whether there may be a large or small amount of within center variability on the variables measured. The decision was made that if ICCs were greater than .05, there would appear to be between group variability that would warrant further investigation.

ICCs for the 12 perception variables were small (*M=.032, median=.025, SD=.063)* which indicates that there was little between center variability on the center characteristics that were rated. The range of ICCs for these perception variables were as small as .001 (Light – dim/bright, Chair – non-adjust/adjust, and workstation – non-adjust/adjust) to as large as .111 (Sound – quite/loud). It should be noted that these perceptions are only an indirect measure of the characteristics of the center. Low ICCs for the 12 perception variables may have occurred for a number of reasons. Artificially low ICCs may have occurred because of using the variables that are not predictive of the construct. It may also have occurred due to the constructs being measured by a single item. Additionally, ICCs may have been low because they are based on limited data. Organizational research using ICCs typically is based on sample sizes of at least 25-30 subjects per group. Since this study had very small numbers of participants (level-1 units) at each center, the confidence intervals may have been very large. Since this study is exploratory, it is difficult to determine if the low variability between centers is due to design error or whether it demonstrates fairly stable environments from center to center. Responses to the open-ended comment suggest that there actually may be between center variability. These comments include one from a repeat test taker who commented “The center where I took my first exam was in the city of Chicago and I had none of these complaints about that center.” Another stated “I took a Florida state exam [jurisprudence] at a testing center and work stations faced each other, staff was a little more noisy when bringing test takers into the room, [and] it was a little harder to concentrate. I was thankful that I did not take the five hour test there.”

Following an examination of the ICCs for participants' perceptions of the center characteristics, similar analyses were conducted to explore between- and within-center variability of participants' preferences and the absolute difference between perception and preference ratings. ICCs for the 12 preference variables were slightly larger than the ICCs of the perception variables (*M=.087, median=.069, SD=.083)* but relatively small. The range for ICCs for these preference variables were as small as .004 (Temp – cold/hot) and .006 (Sound – quiet/loud) to as large as .249 (Size – small/spacious). Small ICC values for the preference variables may indicate that the individuals across the centers were fairly similar in their preferences (little between center variability) for these environmental characteristics.

Research question 2b. On the dimensions of the test environment identified (e.g., room size/layout, climate, lighting, sound, workstation chair and desktop design), how much variability in the absolute difference scores (i.e., difference in examinees' perceptions and preferences of the testing environment) exists between testing centers and how much is within testing centers administering the NPTE?

The absolute difference score was created to look at the difference between examinees' perceptions of the testing environment and their environmental preferences. Analysis of this difference score provides insight into the degree to which the testing environment matched the preferences of the examinees. The absolute difference ICC values were also relatively small (*M=.078, median=.060, SD=.090*). The range of ICCs for the absolute difference were as small as .000 (Light – dim/bright) to as large as .276 (Size – enclosed/open). This measure may indicate that the difference between what examinees prefer in a testing environment and what the centers can offer is relatively similar from center to center. Variability may be more affected by the variability of individual preferences versus that of the centers on these 12 environmental dimensions. Again, since the sample size is relatively low (< 200 centers), findings cannot be generalized to all testing environments. Further analysis of these differences between what examinees want for a testing environment and what testing environments offer is needed to better understand this issue.

Research Question 2c. How much variability in the "effect on performance" variable and likelihood of choosing the same center again if they needed to retake the NPTE exists between testing centers and how much is within testing centers administering the NPTE?

The ICCs for the effect on performance variable and likelihood of choosing the same center again if they needed to retake the NPTE were .004 and .059, respectively. As with the variability discussed for research question 2a and 2b, these small ICCs demonstrate little between center variability. Further research using a more robust sample size could help explore what factors affect examinees' perception of the how aspects of the testing environment affect test performance or whether they would choose to take an exam in a particular environment again. Responses to the open- ended question did identify the perception that the environment offered in the testing centers, represented by this study, did have an effect on their performance, but it is unclear if there is variability between testing centers or if this is just an individual's perception of an effect.

Research Questions 3a-3c

Further analysis explored the relationship between examinee characteristics and their perceptions of the testing environment using single level analysis as well as multilevel analysis. This exploration was the first of its kind to explore if examinees' experiences and achievements in their education, as well as their personal characteristics, predict the perception of a testing environment as well as the perceived environmental effect on test performance.

Research question 3a. What is the relationship between examinee characteristics (sex, program GPA, age, online experience, online testing experience, comfort level with online testing, preferred time) and how examinees perceive dimensions of the test environment (room size/layout, climate, lighting, sound, workstation chair and desktop design)?

Analysis of individual differences was indicated since unconditional models found relatively small variability between centers on the environmental perception variables that were measured. Additionally, small levels of variability were also noted between centers for individuals' preferences for these environmental characteristics. The variability of absolute difference between individual's preferences and perceptions on these characteristics was also low.

When data were not considered in a nested structure, multiple regression analysis demonstrated no statistically significant coefficients for the predictor variables (examinee characteristics) on environmental perception variables. These seven predictor variables accounted for < 10% of the variability found in each of the outcome variables (sample sizes for these models ranged from 135 to 166). R-square values ranged from .01 to .10, and none was statistically significant. Examinee characteristics do not appear to explain the variability in each of the environmental perception variables measured when center membership is not considered.

Two-level multilevel modeling, which takes into account the nesting of examinees within centers, demonstrated a similar result with only one examinee characteristic variable explaining the variability within centers on the 12 perception variables (p<.001). Unstandardized regression coefficients ranged from 0.00 to 0.89 and only the coefficient for the predictor "online test experience" was statistically significant (p<.001) for the outcome variable room temperature (cold/hot). The regression coefficient of -0.83 indicated that those with more online testing experience tended to perceive the testing environment as colder. However, in general, it would appear that the majority of individual variables included in this study do not explain the variability observed in environmental perception variables.

Research Question 3b. What is the relationship between examinee characteristics (sex, , program GPA, age, online experience, online testing experience, comfort level with online testing, preferred time) and absolute difference scores of examinees' perceptions and preferences of the testing environment (room size/layout, climate, lighting, sound, workstation chair and desktop design)?

Single level analysis identified 2 significant predictor coefficients, but each for different outcome variables. The coefficient for the predictor “online test experience” was statistically significant (*B =0.68, p*<.001) for the outcome [absolute difference] variable room lighting (soft/intense). Therefore those participants who had more online test experience had a larger gap between their preference for lighting (soft/intense) and what they perceived in the environment in a given center. This may suggest that those with more online testing experience had less satisfaction with the lighting in the testing center. The coefficient for the predictor “comfort” was statistically significant (*B = -0.37, p*<.001) for the outcome variable room size (small/spacious). Therefore, those with more comfort in taking tests in a fully online format appeared to have less difference between what they prefer in testing room size (small/spacious) and what they experienced in the testing center. This may suggest that there was increased satisfaction by these participants.

The lack of statistical significance for many of the predictor variables may be explained by the number of examinees that responded from each of the testing centers. As with research question 3a, it would appear that the individual variables included in this study do not explain the variability observed in environmental perception variables.

Research question 3c. What is the relationship between examinee characteristics (sex, program GPA, age, online experience, online testing experience, comfort level with online testing, preferred time) and perception of effect on performance and choosing the center again?

Participants were asked to rate the environment on its effect of their performance on the NPTE. A low rating indicated that it was perceived that the environment prevented the individual from performing at their best and a high rating indicated that it allowed them to perform at their best. Participants were asked this question to measure their perception of the impact the environment had on testing performance/outcome. Participants were also asked, based solely on the testing environment, if they would likely choose the same testing center if they had to take the exam again. A low rating indicated that the individual would not likely choose the center again and a high rating indicated that they were highly likely to choose the testing center attain if necessary. Participants were asked this question to identify to what degree the environmental factors were concerning enough that the individual would not use the center again. Since examinees are kept from moving forward to employment until they successfully pass this exam, it is believed that individuals would not take the risk of using a center again if they believed that it has a significant impact on their performance. Statistically significant negative relationships existed between the participant's age and the degree to which the perceived that the environmental characteristics affected performance ($B = -.07$, $p < .001$) and their likelihood to use the center again ($B = -.08$, $p < .001$). These results indicate that participants who were older may be more sensitive to the environment in which they take exams and have a greater perception of effect on performance. Further investigation of the influence that age has on these testing issues would allow for an improved understanding of older test takers.

Lastly, to complete the consideration of research questions 3a-3c, a comparison between single-level and multilevel analysis was made. Further analysis of the relationship between the significant predictor variables (individual characteristics) and outcome variables (perception and absolute difference of environmental characteristics) was conducted to compare the type of significant findings identified with single-level analysis versus multilevel analysis (Table 4.xx - pg 70). It was observed that the majority of significant findings identified through single level analysis was also found through multilevel analysis. The similarity in findings may be due in part to the relatively low ICCs for the various outcome variables. Nesting the data within centers did not provide additional insight into the variability observed in individual characteristics and prediction of perception and absolute difference ratings. Once again, this may be due to the small sample size for each of the centers. If a larger overall sample size was obtained, and more participants within each center were analyzed, the outcome of these statistical analyses may demonstrate a greater effect of nesting the data.

Research Question 4 a- 4c

Research Question 4a. What is the relationship between center characteristics (room size, room density, center 'newness', presence of ambient light, break space, and access to food/drink) and examinees' perceptions of the testing environment (room size/layout, climate, lighting, sound, workstation chair and desktop design)?

Research Question 4b. What is the relationship between center characteristics (room size, room density, center 'newness', presence of ambient light, break space, and access to food/drink) and the absolute difference between examinees' preferences and perceptions of the testing environment (room size/layout, climate, lighting, sound, workstation chair and desktop design)?

Research Question 4c. What is the relationship between center characteristics (room size, room density, center 'newness', presence of ambient light, break space, and access to food/drink) and examinees' perceptions of the effect on performance and choosing the center again?

Research questions 4a-4c explored the relationship between the characteristics of the centers and how examinees perceived the environment, the difference between what they prefer and what they experienced, and how the examinees perceived the environments effect on their test performance.

Multilevel analysis identified several significant relationships between center-level predictor variables and outcome variables which may begin to identify important factors for testing centers to consider. The predictor "room size" (number of workstations in the room) had a statistically significant negative relationship with the perception of room lighting (soft/intense). Larger rooms were associated with softer room lighting. This effect was observed using either single level or multilevel analyses. The predictor “room density” (number of workstations in use) showed a statistically significant negative relationship with room size (enclosed/open) using either analyses. Rooms with more workstations in use were perceived as being more enclosed. Additionally the predictor “room density” showed a statistically significant relationship with the absolute difference between preference and perception of the “room size” (small/spacious) with both single level and multilevel analyses. Rooms with more stations in use were associated with larger discrepancies between preference and perception of room size (small/spacious).

The predictor “center newness” demonstrated a statistically significant negative relationship with the outcome variable noise (quiet/loud). Newer centers were perceived as quieter. This may be a real feature of the centers as materials in newer spaces may have more sound dampening characteristics. The predictor center newness also demonstrated a statistically significant negative relationship with absolute difference between preference and perception of the outcome variable room size (small/spacious), room noise (quiet/loud), as well as desk size (small/large). This indicates that the newer the facility the less discrepancy existed between what participants preferred for room size, sound, and desk size and what they actually experienced in the testing center. Center newness was also positively associated with a positive effect on performance and likelihood to use the center again. The newness of the center was the predictor of the greatest number of outcome variables for both actual and absolute difference scores on the environmental characteristics. It could be inferred that overall satisfaction of participants was greater and therefore there was less perceived difference between what they wanted and what they received in the testing environment. Lastly, the availability of food, drink or both during testing breaks was a significant predictor for light (dim/bright) and the absolute difference between preferred and actual chair softness/hardness in both analyses. It is unclear as to the reason for this relationship given that these characteristics differ greatly.

Limitations and Future Research

This study served as a baseline, exploratory study to begin to identify the characteristics of the individuals who sit for the NPTE, the preferences that individuals have for testing environment, and to begin to explore the variability that exists in high stakes testing environments. Research focused on the human factors is limited and there is no previous baseline research looking at the environmental characteristics, preferred and perceived, of testing environments in which adults take high-stakes, computer-based exams.

The sample size was limited and caution is warranted in interpreting the findings. Additional research with larger samples of examinees is needed to determine if the results of the present study are replicated. Difficulties in recruiting large samples of examinees were noted throughout this document. Participants could not be recruited directly from registration lists for the NPTE exam. Therefore, recruitment occurred through physical therapy programs forwarding the recruitment emails to their recent graduates. Multiple attempts to gain participation from recent examinees were made. The number of participants, as well as the number of centers represented, did increase from the first to the second email request, however, overall participation was limited to just over 200 individuals with just over 100 centers being represented. Some centers had low usage due to their location, while others were large and had high usage patterns. Although the number of participants per center did not lead to many statistically significant findings, the open-ended comments demonstrated the variability between centers on size, conditions, distractions and even the degree to which the security rules were followed. The collection of these baseline data will assist in the design of future research looking at test administration and testing environments for adults.

A direct measure of the environment was not possible at the time that this study was conducted. Indirect measures of participants' recall and perception of the environment were collected by self-report. The use of self-report may have had the greatest impact on the center characteristics data. The variability of data for the center characteristics may demonstrate some individuals' limited ability to estimate certain aspects of their environment, even those as concrete as the size of the room. This effect was minimized to some degree by the inclusion criteria being limited to individuals who took their exam within the 12 months preceding data collection. The addition of participant interviews may have added to the researcher's ability to clarify the perceptions of the environment.

Variability may also have affected certain findings as some characteristics may be different within a center depending on the day of the week, time of day, and time of year (e.g., room temperature, noise, etc.). Although an attempt was made to analyze these fluctuations, it is unclear if the degree of these differences was detected. Future studies should re-attempt to gain direct access to testing centers for observation of centers themselves while examinees are present, while still maintaining test administration security. These direct measures may include lighting levels, sound levels, measured dimensions and room layout, temperature readings, ergonomic analysis of chair and workspace, and how many examinees were in the testing space at any given time.

Lastly, there may have even been errors in how the participant reported the particular center that they went to take the exam. Participants were encouraged to not only provide the city and state in which the exam was taken, but to inform the researcher if there was more than one testing center in the city. If so, the participant was asked to provide the street on which the center was located. Through having a master list of all Prometric centers, some errors could be corrected through cross referencing these data with the master list of locations. Any remaining errors could not be detected in the raw data set without having other evidence of the participants' location. Future research may want to include pull-down menus of testing locations as part of the online survey to decrease the potential errors on the identification of the testing center.

Further exploration into the testing environment may provide follow-up to several comments made in the open-ended questions. Eight individuals mentioned going outside the testing center during the exam. Two of these eight mentioned this as a desired break but were not allowed to go outside by the testing center staff, however four were permitted to walk around outdoors. Two others were permitted to walk in the building, but outside the Prometric suite. These inconsistencies in application of policies, as well as preferences for fresh air and physical activity during testing, require more exploration. The open-ended question identified a variety of perspectives on the test environment and identified categories of areas for which participants commented. Investigating these categories of comments would benefit further research in this area of test administration and lead to a deeper understanding of what is most important to examinees.

The use of effective sound dampening devices, such as headphones, should be explored further to determine if this would assist examinees during test taking. Future studies may also want to explore the effects of physical discomfort and poor ergonomic design on examinee performance. Several participants mentioned physical fatigue, as well as pain, as being a source of distraction. Currency of technology appears also to be a source of variability in the examinees' testing experience from center to center based on comments within the open ended question responses. Lastly, concerns were raised about the time taken with fingerprinting and its effect on the amount of break time that was allowed. Other security practices, such as the adjustment of cameras, were also mentioned by several participants. With the ever increasing amount of security measures that must be put in place, the human factors of the implementation of these measures should be considered. Continued investigation of security measures that do not increase distractions and stress for examinees is needed.

Since the NPTE is now offered on fixed dates, where all examinees in a center are taking the same exam during the same period of time, some of the strongest comments related to visual and sound distractions may be diminished. Scheduled breaks will occur at the same time for everyone, yet unscheduled breaks will still be taken by some. Examinees will begin at the same time, yet exiting behaviors will vary. The need for a dedicated break space may have more importance as breaks will be taken simultaneously by more people. Fixed date testing for the NPTE may allow for additional study through the use of focus groups and individual interviews. Focus groups and individual qualitative interviews with examinees shortly after taking the NPTE at different centers may provide rich insights into the testing experience.

Summary and Conclusions

Test administrators have many obstacles to offering valid and reliable high-stakes exams as well as providing a secure testing environment. These challenges may decrease the current focus on maximizing and regulating the test environment itself. In-depth exploration of the variability between centers may expose tests to challenges by examinees that may or may not be valid. However, this study illustrates that testing agencies may want to reconsider the time and emphasis that need to be placed on monitoring these testing sites. Human factors, such as individuals' response to their environment, may need more attention in both the learning and testing environment.

REFERENCES

Al-Hindawe, J. (2006). *Considerations when constructing a semantic differential scale*. LaTrobe Papers in Linguistics, 9. Retrieved from https://www.latrobe.edu.au/linguistics/LaTrobePapersinLinguistics/Vol%2009/03 lHindawe.pdf.

Altenberg, E., & Ferrand, C. (2006). Perception of individuals with voice disorders by monolingual English, bilingual Cantonese-English, and bilingual Russian-English women. *Journal of Speech, Language, and Hearing Research*, *49*, 879-887.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

American National Standards Institute. (2002). Acoustical performance criteria, design requirements, and guidelines for schools. ANSI S12.60-2002. Retrieved from http://asa.aip.org/classroom.html.

American Physical Therapy Association. (2008). *2007-08 Fact Sheet of Physical Therapist Educational Programs*. Alexandria, VA: APTA.

American Physical Therapy Association. (2007). *2007 Annual Accreditation Report.* Alexandria, VA: APTA.

Association of Test Publishers. (2002). *Testing guidelines: Guidelines for computer-based testing*. Washington, DC: ATP.

Banbury, S., & Berry, D. (1998). Disruption of office-related tasks by speech and office noise. *British Journal of Psychology*, *89*, 499-517.

Bartram, D. (2009), The International Test Commission Guidelines on Computer-Based and Internet-Delivered Testing. *Industrial and Organizational Psychology*, *2*, 11–13. doi: 10.1111/j.1754-9434.2008.01098.

Bekkering, G., Hendriks, H., van Tulder, M., Knol, D., Hoeijenbos, M. Oostendorp, R., & Bouter, L. (2005). Effect on the process of care of an active strategy to implement clinical guidelines on physiotherapy for low back pain: A cluster randomised controlled trial. *Quality & Safety in Health Care*, *14*, 107-112.

Bekkering, G., van Tulder, M., Hendriks, E., Koopmanschap, M., Knol, D., Bouter, L., & Oostendorp, R. (2005). Implementation of clinical guidelines on physical therapy for patients with low back pain: Randomized trial comparing patient outcomes after a standard and active implementation strategy. *Physical Therapy*, *85*, 544-555.

Bell, P. A., Fisher, J. D., & Loomis, R. J. (1978). *Environmental psychology*. Philadelphia, PA: W.B. Saunders.

Bechtel, R., & Churchman, A. (2002). *Handbook of environmental psychology*. New York, NY: Wiley.

Bernecker, C., Brainard, G., Fernsler, F., Rollag, M., Long, R., Tierney, S., & Gaddy, J. (1994). Biological effects of architectural lighting and their associated energy utilization. *Journal of the Illuminating Engineering Society*, *25*(2), 41-49.

Brill, M., Weidemann, S, Alard, L., Olson, J., & Keable, E. (2001). *Disproving widespread myths about workplace design.* Jasper, IN: Kimball International.

Butler, D. L., & Biner, P. M. (1987). Preferred lighting levels: Variability among settings, behaviors, and individuals. *Environment and Behavior*, *19*, 695-721.

Canadian Centre for Occupational Health and Safety. (October, 2007). Thermal comfort for office work. Retrieved from www.ccohs.ca/oshanswers/physical_agents/thermal_comfort.html.

Charles, K. E., & Veitch, J. A. (2002). *Environmental satisfaction in open-plan environments: 2. Effects of workstation size, partition height and windows*. National Research Council Canada, Internal Report (IRC-IR-845). Ottawa, Ontario, Canada.

Cohen, S. (1978). Environmental load and the allocation of attention. In A. Baum & S. Valins (Eds.), *Advances in environmental research* (pp 1-29). Hillsdale, NJ: Lawrence Erlbaum.

Cohen, S., Evans, G., Stokols, D., & Krantz, D. (1991). *Behavior, health, and environmental stress*. New York, NY: Plenum Press.

Christensen, R. & Knezek, G. (2009). Construct validity for the teachers' attitudes toward computers questionnaire. *Journal of Computing in Teacher Education*, *25*, 143-155.

Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory* (3rd ed.) Belmont, CA: Wadsworth.

De Young, R. (1999). Environmental psychology. In D.E. Alexander & R.W. Fairbridge (Eds.), *Encyclopedia of environmental science* (pp 223-224). Hingham, MA: Kluwer.

Dockter, M. (2001). An analysis of physical therapy preadmission factors on academic success and success on the national licensing examination. *Journal of Physical Therapy Education*, *15*, 60-64.

Dunn, R., & Dunn, K. (1978). *Teaching students through their individual learning styles.* Reston, VA: Reston.

Engers, A., Wensing, M., van Tulder, M., Timmermans, A., Oostendorp, R., Koes, B., & Grol, R. (2005). Implementation of the Dutch low back pain guideline for general practitioners: A cluster randomized controlled trial. *Spine*, *30*, 559-600.

Fang, L., Clausen, G., & Fanger, P. O. (1998). Impact of temperature and humidity on the perception of indoor air quality. *Indoor Air*, *8*, 80-90.

FSBPT. (2009). *2009 NPTE Candidate Handbook for the National Physical Therapy Examinations: PT, PTA*. Retrieved from fsbpt.org/download/Candidatehandbook20090406.pdf.

Fulton, R. (1991). A conceptual model for understanding the physical attributes of learning environments. In R. Hiemstra (Ed.), *New Directions for Adult and Continuing Education: Creating Environments for Effective Adult Learning*, *50*, 13-22.

Gordon, H. (1996). *Analysis of productivity and learning style preferences of participants in distance education*. Paper presented at the Annual Conference of the American Educational Research Association, New York, NY. (ERIC Document Reproduction Service No. ED397132)

Guffey, J., Farris, J., Aldridge, R., & Thomas, T. (2002). An evaluation of the usefulness of noncognitive variables as predictors of scores on the National Physical Therapy Examination. *Journal of Allied Health, 31,* 78-86.

Hamilton, M. B. (2003). Online survey response rates and times: Background and guidance for industry. Retrieved from www.supersurvey.com/papers/supersurvey_white_paper_response_rates.pdf.

Heise, D. (2010). *Surveying cultures: Discovering shared conceptions and sentiments*. Hoboken, NJ: Wiley.

Hendriks, E., Kerssens, J., Dekker, J., Nelson, R., Oostendorp, R., & van der Zee, J. (2003). One-time physical therapist consultation in primary health care. *Physical Therapy*, *83*, 918-931.

Himmelfarb, S. (1993). The measurement of attitudes. In A. Eagly. & S. Chaiken. (Eds.), *Psychology of attitudes* (pp. 23-88). Florence, KY: Thomson/Wadsworth.

Hollman, J., Rindflesch, A., Youdas, J., Krause, D., Hellyer, N., & Kinlaw, D. (2008). Retrospective analysis of the behavioral interview and other preadmission variables to predict licensure examination outcomes in physical therapy. *Journal of Allied Health*, 37, 97-104.

International Test Commission. (2006). International Guidelines on Computer-Based and Internet Delivered Testing. *International Journal of Testing*, *6* (2), 143-171.

Jones, J. (2000). Promoting stakeholder acceptance of CBT. *Journal of Applied Testing Technology*, retrieved January 26, 2007, from www.testpublishers.org/journal02.htm.

Kerssens, J., Sluijs, E., Verhaak, P., Knibbe, H., & Hermans, I. (1999). Back care instructions in physical therapy: A trend analysis of individualized back care programs. *Physical Therapy*, *79*, 286-295.

Lallh, A., & A. Rochet. (2000). The effect of information on listeners' attitudes toward speakers with voice or resonance disorders. *Journal of Speech, Language, and Hearing Research*, *43*, 782-783.

Lee, S. Y., & Brand, J. L. (2005). Effects of control over office workspace on perceptions of the work environment and work outcomes. *Journal of Environmental Psychology*, 25, 323-333.

Lusk, C., Delclos, G. L., Burau, K., Drawhorn, D. D., & Aday, L. A. (2007). Mail versus internet surveys: Determinants of method of response preferences among health professionals. *Evaluation & the Health Professions*, *30*, 186-201.

Mehrabian, A., & Russell, J. (1974). *An Approach to Environmental Psychology.* Cambridge, MA: MIT Press.

Mohr, T., Ingram, D., Hayes, S., & Du, Z. (2005). Educational program characteristics and pass rates on the National Physical Therapy Examination. *Journal of Physical Therapy Education*, *19*, 60-66.

Occupational Safety and Health Administration. (n.d.). OSHA ergonomic solutions: Computer workstation eTool - Evaluation checklist. Retrieved October 12, 2007, from http://www.osha.gov/SLTC/etools/computerworkstations/.

Osborne, J., & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical Assessment, Research & Evaluation*, *8*(2). Retrieved from http://PAREonline.net/getvn.asp?v=8&n2.

Osgood, C., Suci, G., & Tannenbaum, P. (1957). *The measurement of meaning*. Urbana, IL: University of Illinois Press.

Osgood, C., May, W., & Miron, M. (1975). *Cross-cultural universals of affective meaning*. Urbana, IL: University of Illinois Press.

Parshall, C., Spray, J., Kalohn, J., & Davey, T. (2002). *Practical considerations in computer-based testing*. New York, NY: Springer.

Paterson, L., & Goldstein, H. (1991). New statistical methods for analyzing social structures: An introduction to multilevel models. *British Educational Research Journal*, *17*, 387-393.

Price, G., Dunn, R., & Dunn, R. (1991). *Productivity environmental preference survey. PEPS Manual*. Lawrence, KS: Price Systems.

Rea, M., Oulette, J., & Kennedy, J. (1985). Lighting and task parameters affecting posture, performance, and subjective ratings. *Journal of the Illuminating Engineering Society*, *14*, 231-238.

Reising, D. L. (2003). The relationship between computer testing during a nursing program and NCLEX performance. *Computers Informatics Nursing*, *21*, 326-329.

Ribich, F., Barone, W., & Agostino, R. (1998). Semantically different: Preservice teachers' reactions to the gifted student concept. *The Journal of Educational Research*, *91*, 308-312.

Riddle, D., Utzman, R., Jewell, D., Pearson, S., & Kong, X. (2009). Academic difficulty and program-level variables predict performance on the National Physical Therapy Examination for licensure: A population-based cohort study. *Physical Therapy, 89*, 1182-1190.

Roehrig, S. M. (1988). Prediction of licensing examination scores in physical therapy graduates. *Physical Therapy*, *68,* 694-698.

Sailer, U., & Hassenzahl, M. (2000). Assessing noise annoyance: An improvement-oriented approach. *Ergonomics*, *43*, 1920-1938.

Sakuta, Y., & Gyoba, J. (2006). Affective impressions and memorability of color-form combinations. *The Journal of General Psychology*, *133*, 191-208.

Short, R., & Magana, L. (2002). Political rhetoric, immigration attitudes, and contemporary prejudice: A Mexican American dilemma. *The Journal of Social Psychology*, *142*, 701-712.

Smellie, S. (2003). The limitations of a standard workstation for its user population. *Clinical Chiropractic*, *6*, 101-108.

Stillman, A., Braitman, L., & Grant, R. (1998). Are critically ill older patients treated differently than similarly ill younger patients? *The Western Journal of Medicine*, *169*, 162-165.

Sundstrom, E., Bell, P. A., Busby, P. L., & Asmus, C. (1996). Environmental psychology 1989-1994. *Annual Review of Psychology*, *47*, 485-512.

Swartz, E., Gabel, R., & Irani, F. (2009). Speech-language pathologists' attitudes towards people who stutter. *Canadian Journal of Speech-Language Pathology & Audiology*, *33*, 99-105.

Sweet, G. (1989). Training 101: Arranging the training room and the trainer. *Training & Development Journal*, *43*, 19-23.

Swinkles, I., Wimmers, R., Groenewegen, P., van den Bosch, W., Dekker, J., & van den Ende, C. (2005). What factors explain the number of physical therapy treatment sessions in patients referred with low back pain: A multilevel analysis. *BMC Health Services Research*, *74* retrieved from http://www.biomedcentral.com/1472-6963/5/74.

Toplyn, G. (1988). The differential effect of noise on creative task performance. *Dissertation Abstracts International, 48*, 3718. Dissertation, Published.

Tracey, J., Arroll, B., Richmond, D., & Barham, P. (1997). The validity of general practitioners' self assessment of knowledge: Cross sectional study. *British Medical Journal*, *315*, 1426-1428.

University of California – Irvine. (n.d.) *Computer Ergonomic Review Tool*. Retrieved from http://www.ehs.uci.edu/programs/safety/COMPUTER%20ERGONOMIC%20REVIEW%20TOOL.pdf.

U.S. Environmental Protection Agency. (August, 2003). Indoor air quality and student performance. Retrieved March 4, 2008, from www.epa.gov/iaq/schools/pdfs/publications/iaq_and_student_performance.pdf.

Utzman, R., Riddle, D., & Jewell, D. (2007). Use of demographic and quantitative admissions data to predict performance on the National Physical Therapy Examination. *Physical Therapy*, *87*, 1-13.

Van Auken, S., Barry, T., & Anderson, R. (1993). Observations: Toward the internal validation of cognitive age measures in advertising research. *Journal of Advertising Research*, *33*, 82-84.

Veitch, J., & Newsham, G. (1998). Determinants of light quality I: State of the science. *Journal of Illuminating Engineering Society*, *27*(1), 92-106.

Vendrely, A. M. (2007). An investigation of the relationships among academic performance, clinical performance, critical thinking, and success on the physical therapy licensure examination. *Journal of Allied Health*, *36*, 108-123.

Wallace, P., & Clariana, R. (2005). Test mode familiarity and performance: Gender and race comparisons of test scores among computer-literate students in advanced information systems courses. *Journal of Information Systems Education*, *16*, 177-182.

Workers Compensation Fund. (2016). *Ergonomics in Workstation Design.* Retrieved from https://www.wcf.com/ergonomics-workstation-design.

Yu, J., Albaum, G., & Swenson, M. (2003). Is there a central tendency error inherent in the use of semantic differential scales in different cultures? *International Journal of Market Research*, *45*, 213-230.

Zandvliet, D., & Straker, L. (2001). Physical and psychosocial aspects of the learning environment in information technology rich classrooms. *Ergonomics*, *44*, 838-857.

Zevin, J., & Corbin, S. (1998). Measuring secondary social studies students' perceptions of nations. *The Social Studies*, *89*, 35-38.

Zugazaga, C., Surette, R., Mendez, M., & Otto, C. (2006). Social worker perceptions of the portrayal of the profession in the news and entertainment media: An exploratory study. *Journal of Social Work Education*, *42*, 621-636.

APPENDICES

Appendix A

International Test Commission (ITC) Computer-based and Internet Delivered Testing Guidelines (2005)

3. Provide appropriate levels of control over CBT and Internet testing

a. Detail the level of control over the test conditions

35. Test Publishers:

3) Inform test users of the need to consider health and safety rules during CBT/Internet testing. For example, identify whether an Internet test has the facility for breaks if the testing process is lengthy (pg 17).

36. Test Users

2) When testing at a specific test centre, ensure that the test-taker is comfortable with the workstation and work surface (e.g., the ergonomics are suitable). For example, test-takers should:
- be encouraged to maintain proper seating posture,
- be able to easily reach and manipulate all keys and controls,
- have sufficient leg room, and
- not be required to sit in one position for too long. (pg 27)

4) Ensure that the facilities, conditions, and requirements of the testing conform to national health and safety, and union rules. For example, there may be rules governing the length of time a person should work at a monitor before having a break, or rules as to adequate lighting, heating, and ventilation. When testing over the Internet, inform test-takers of such rules and regulations. (pg 27-28)

Appendix B

Descriptive Statistics for Examinee Online Experience, Experience with Online Testing, and Comfort W/Online Testing

		Percentage				
Variable	N	0-25%	26-50%	51-75%	76-100%	
% of online coursework	211[a]	85.6	9.8	1.4	.9	
% of exams online	211[a]	87.9	7.0	2.8	0.0	
		Percentage				
		1[b]	2	3	4	5
Comfort level w/online	210[a]	2.4	8.1	25.2	39.0	25.2

Note. [a] One case was eliminated; no response was received from four participants. [b] 1 equals very comfortable, 5 equals very uncomfortable

Appendix C

Descriptive Statistics for Examinee Use of Breaks During Exam

		Percentage			
Variable	N	0	1	2	3
# of Scheduled	206 [a]	3.3	96.6	0	0
# of Schedule – out	205	n/a	96.1	0	0
# of Unscheduled	205	57.6	22.4	9.8	10.2
# of Unscheduled –out	203	n/a	20.2	8.4	8.4

Note. [a] Eleven cases were eliminated because participants exited survey

Appendix D

Responses to Open-Ended Question Related to "Breaks"

Behaviors:
"I walked around the building for fresh air during my scheduled break. I brought my own snack."(#29955)
"The breaks were taken in a separate room with a water dispenser; however, I was not allowed to go outside for fresh air."(#29991)
"I was allowed to take as many breaks as needed and left the building for the long break that was mandatory. "(#30331)
" The break space was the waiting room but I went outside to get some fresh air."(#30128)
"I did not give much attention to break space, all I need is space to walk so I can walk around during my break to get some exercise. The walking space outside the testing room was adequate."(#30338)
Space:
The breakspace was the waiting area.(#29962)
"small with a small waiting/break area"(#29955)
"There was no break space, when checking out one could go to the general waiting room and there was restroom, but not real space to relax and take a break." (#30038)
There were no break rooms, only an area for small lockers and restrooms.(#30053)
"There was not a specified "break room" in the facility so I just wandered out into the hall with the snack and drink that I brought on my own. '(#30056)
"small lobby area to take breaks "(#30013)
"no break room other than the waiting area"(#29972)
"Facility was new and break areas were nice. "(#30138)
'We were not shown a break area, but there was a room by our lockers where you could sit & a bathroom."(#30200)
"Had small break room outside main room."(#30207)
"The break room was spacious with plenty of magazines to help me relax during my scheduled break."(#30246)
"I had access to a locker where I stowed food and drink that I could have during the breaks (there were no vending machines). There was a water dispenser. "(#30330)
"There was no specific "break space" but there was rather a waiting area where there were bathrooms/lockers." (#30367)
"We did not have a separate room for a break room. You just walked out into the main lobby where all the lockers were. If you brought food you were allowed to eat it. They had a water fountain available."(#30371)

Appendix E

HLM Models for Nested Data – Preferences

	Dependent Variable			
Predictor	Size (small-spac) *B*	Size (encl-open) *B*	Temp (cold-hot) *B*	Temp (dry-humid) *B*
Sex	-.096 (.36)	.102(.38)	.241 (.20)	.238 (.37)
Degree	.532 (.34)	.315 (.35)	.112 (.16)	.205 (.30)
GPA	-.195 (.58)	.690 (.61)	.211 (.30)	-.132 (.56)
Age	.015 (.04)	-.003 (.04)	.002 (.02)	-.031 (.03)
Online	.511 (.30)	.770 (.32)	-.121 (.16)	-.015 (.31)
Online-test	-.427 (.35)	-.079 (.37)	.252 (.17)	-.047 (.33)
Comfort	-.071 (.15)	-.029 (.16)	.069 (.08)	-.125 (.15)
Preferred time	-.113 (.14)	.092 (.14)	-.030 (.07)	-.219 (.13)

Note. None of the coefficients was statistically significant (all *p*-levels were > .001). Standard errors are in parentheses.
[a] Sex (1 = Male, 2 = Female); [b] Degree (1= Masters, 2 = DPT); [c]Online – percentage of coursework 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%; [d] Online test – percentage of testing 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%; [e] Comfort – online testing 1 = very comfortable, 5 = very uncomfortable; [f] preferred time – for taking tests 1 = early morning (before 10 am), 2 = late morning (10 am – noon); 3 = early afternoon (after noon but before 3 pm); 4 = late afternoon (3 pm or after), 5 = no preference.

Appendix E (Continued)

HLM Models for nested data -- preferences

	Dependent Variable			
Predictor	Light (dim-bright)	Light (soft-intense)	Sound (quiet-loud)	Sound (muff-clear)
	B	*B*	*B*	*B*
Sex	.402 (.24)	.148 (.25)	-.007 (.02)	.-.613 (.45)
Degree	.173 (.19)	.013 (.21)	-.004 (.13)	.123 (.38)
GPA	-.596 (.35)	-.268 (.39)	.520 (.24)	-.942 (.69)
Age	.007 (.02)	.023 (.02)	-.007 (.02)	-.049 (.04)
Online	-.135 (.19)	-.290 (.21)	.089 (.13)	-.173 (.38)
Online-test	.266 (.21)	-.168 (.23)	-.066 (.14)	-.161 (.42)
Comfort	-.019 (.09)	-.154 (.10)	-.067 (.07)	-.128 (.18)
Preferred time	.062 (.08)	.044 (.09)	-.030 (.06)	.065 (.16)

Note. None of the coefficients was statistically significant (all *p*-levels were > .001). Standard errors are in parentheses.
[a] Sex (1 = Male, 2 = Female); [b] Degree (1= Masters, 2 = DPT); [c]Online – percentage of coursework 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%; [d] Online test – percentage of testing 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%; [e] Comfort – online testing 1 = very comfortable, 5 = very uncomfortable; [f] preferred time – for taking tests 1 = early morning (before 10 am), 2 = late morning (10 am – noon); 3 = early afternoon (after noon but before 3 pm); 4 = late afternoon (3 pm or after), 5 = no preference.

Appendix E (Continued)

HLM Models for nested data -- preferences

	Dependent Variable			
Predictor	Chair (soft-hard)	Chair (not adj-high adj)	Station (sm-large)	Station (not adj-high adj)
	B	*B*	*B*	*B*
Sex	-.035 (.25)	.441 (.27)	-.021 (.03)	.007 (.35)
Degree	-.327 (.21)	.068 (.22)	.007 (.23)	.278 (.32)
GPA	.064 (.38)	.175 (.41)	.359 (.41)	-.168 (.56)
Age	-.003 (.02)	.030 (.03)	-.021 (.03)	-.031 (.04)
Online	.243 (.21)	.146 (.22)	.181 (.22)	.021 (.30)
Online-test	-.172 (.23)	.204 (.24)	-.035 (.25)	-.075 (.34)
Comfort	-.062 (.10)	-.145 (.11)	-.165 (.11)	-.109 (.15)
Preferred time	.064 (.09)	.070 (.10)	-.089 (.10)	-.209 (.13)

Note. None of the coefficients was statistically significant (all *p*-levels were > .001). Standard errors are in parentheses.
[a] Sex (1 = Male, 2 = Female); [b] Degree (1= Masters, 2 = DPT); [c]Online – percentage of coursework 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%; [d] Online test – percentage of testing 1 = 0-25%, 2 = 26-50%, 3 = 51-75%, 4 = 76-100%; [e] Comfort – online testing 1 = very comfortable, 5 = very uncomfortable; [f] preferred time – for taking tests 1 = early morning (before 10 am), 2 = late morning (10 am – noon); 3 = early afternoon (after noon but before 3 pm); 4 = late afternoon (3 pm or after), 5 = no preference.

Appendix F

Responses to Open-Ended Question Related to Use of Headphones

---

"I chose to use headphones because I am easily distracted by noise. However, I questioned how clean the headphones were." (#29944)
"Yes, headphones were provided. I attempted to wear them, but the headphones created a echo tunnel sound of the typing. " (#29962)
"Cubicles were in the testing area with optional headphones that sounded like a tunnel when placed on. They were large headphones. I wore them at times. They muffled the coughing and moving of other test takers." (#29955)
" I wore headphones the entire time because of this and because other people were coming and going because their tests were a lot shorter than mine.(#29968)"
" I usually don't have to use the headphones, but there was a guy next to me with a nervous tick and kept tapping on his chair with his foot, therefore I decided to use the headphones. After wearing them awhile, they dug into the area around my ears and I found that to be very uncomfortable." (#29980)
"Sound was not a factor for me and I did not use the headphones during the exam. "(#30056)
"The main problem i had was someone taking some sort of exam with a lot of typing next to me. it was very distracting, even with headphones on. " (#30268)
" I wore the headphones but by the end of the test, they were uncomfortable and hurting my ears." (#30394)
"I did wear the headphones provided to me to help block out the sound of the air conditioner turning on and off and the other people coming in and out of the testing room. Not all the people in the testing room were taking the NPTE so there was extra noise coming from them typing on the keyboards." (#30579)
"It was easy to hear other people moving chairs, coming and going. Even with the headphones it was sometimes hard to focus. " (#30588)
"The headphones provided for the tests were very uncomfortable and didn't block any sounds." (#30793)
"I wore headphones to drown out the constant clicking of other test takers. Also, there was an elderly gentleman behind me that was having great difficulty w/ his exam and got up to get an employee several times throughout my time there. That made it hard to concentrate on my exam b/c I could still hear through the headphones. " (#30849)

"Since there were so many people in the room and many were having to type essays I heard a lot of clicking on the keyboard, even with the headphones on. I caught myself not focusing on questions frequently and I was having to re-read things/focus more."(#30862)
" I had a difficult time concentrating, even with the headphones, because of noise from someone constantly coughing/sniffling. It would be nice to put someone who is sick in a seperate area." (#30864)
"The people coming in and out of the room were some what distracting. I automatically put my headphones on when I sat down at my station. " (#30925)
"I used the noise eliminating headphones in the beginning and towards the middle felt they were a nuisance and I removed them. ... I had the noise eliminating headphones on and I could still hear that clip-clip noise. That was extremely irritating. "(#31628)
"Even with the headphones, the noise was very audible and distracting. "(#31741)
"The testing room was adequate but the clicking of the keyboards around me was very distracting. I could hear them through the headphones and I ended up having to take the headphones off because they bothered my ears after an hour or so."(#34287)

“Headphones were provided, however, they were oversized, worn, and not adjustable. They did not fit me so I was unable to use them. However, the room was quiet enough that the background noise did not bother me….. I felt even with the headphones there were many distractions.”(#35127)

“Earphones were uncomfortable to wear but the noise would have been more distracting if I did not wear them.”(#30598)

“Earphones were too heavy to use comfortably.” (#35006)

“Several of the other test-takers @ the facility were taking standardized tests that required a lot of typing, which was loud & was not muffled out by the headphones the testing facility provided (they did not provide earplugs, unfortunately).”(#36442)

Appendix G

IRB Exemption Letter

USF
UNIVERSITY OF SOUTH FLORIDA

DIVISION OF RESEARCH INTEGRITY AND COMPLIANCE
Institutional Review Boards, FWA No. 00001669
12901 Bruce B. Downs Blvd., MDC035 • Tampa, FL 33612-4799
(813) 974-5638 • FAX (813) 974-5618

August 18, 2008

Ellen Williamson, MS PT
Department of Educational Measurement and Research
21040 North River Rd.
Alva, FL 33920

RE: **Exempt Certification** for IRB#: 107172 G
Title: *Adults' Perceptions of the Testing Environment During the National Physical Therapy Examination*

Dear Ms. Williamson:

On August 15, 2008, the Institutional Review Board (IRB) determined that your research **meets USF requirements and Federal Exemption criteria two (2).** It is your responsibility to ensure that this research is conducted in a manner reported in your application and consistent with the ethical principles outlined in the Belmont Report and with USF IRB policies and procedures.

Please note that changes to this protocol may disqualify it from exempt status. It is your responsibility to notify the IRB prior to implementing any changes.

The Division of Research Integrity and Compliance will hold your exemption application for a period of five years from the date of this letter or for three years after a Final Progress Report is received. If you wish to continue this protocol beyond those periods, you will need to submit an Exemption Certification Request form at least 30 days before this exempt certification ends. If a Final Progress Report has not been received, the IRB will send you a reminder notice prior to end of the five year period; therefore, it is important that you keep your contact information current with the IRB Office. Should you complete this study prior to the end of the five-year period, you must submit a Final IRB Progress Report for review.

Please reference the above IRB protocol number in all correspondence to the IRB c/o the Division of Research Integrity and Compliance. In addition, we have enclosed an Institutional Review Board (IRB) Quick Reference Guide providing guidelines and resources to assist you in meeting your responsibilities when conducting human subjects research. **Please read this guide carefully.**

We appreciate your dedication to the ethical conduct of human subject research at the University of South Florida and your continued commitment to human research protections. If you have any questions regarding this matter, please call 813-974-9343.

Sincerely,

Paul G. Stiles, J.D., Ph.D., Chairperson
USF Institutional Review Board

Enclosures: IRB Quick Reference Guide

Cc: Anna Davis/cd, USF IRB Professional Staff
Robert Dedrick, PhD

SB-EXEMPT-0602

Appendix H

Testing Environment Survey

Page 1 of 12

Thank you for your participation in this study. The survey will take you through 9 short pages of questions and should take you approximately 10 minutes to complete. The first page is an Informed Consent Form which requires your agreement to proceed to the survey. Please continue to the next page to provide your informed consent.

Save Progress and Exit   CONTINUE

Informed Consent to Participate in Research

Information to Consider Before Taking Part in this Research Study

Researchers at the University of South Florida (USF) and Florida Gulf Coast University (FGCU) study many topics. To do this, we need the help of people who agree to take part in a research study. This form tells you about this research study.

We are asking you to take part in a research study that is called: Adults' Perceptions of the Testing Environment During the National Physical Therapy Examination

The person who is in charge of this research study is Ellen K. Williamson, MS PT, Doctoral Candidate under the supervision of Robert Dedrick, PhD, Faculty Advisor.

The research will be conducted on-line through your completion of a web-based survey.

Purpose of the study

The purpose of the study is to begin to understand the physical environment in which test takers take licensure examinations. This will be accomplished through collecting information from about 900 recent test takers of the National Physical Therapy Examination (NPTE) and how they perceived the testing center in which they sat for the exam.

Study Procedures

If you take part in this study, you will be asked to complete a 9-page on-line survey that is estimated to take about 10 minutes to complete. The survey asks about your preferences when taking a test such as the NPTE (e.g., room temperature) and your actual perceptions of the environmental conditions of the testing environment the last time you took the NPTE. You will only be asked to participate one time and your involvement in the study ends when you submit your survey electronically.

Alternatives

Your participation in the study is voluntary and you have the alternative to choose not to participate in this research study.

Benefits

The potential benefits to you are that you will reflect on your preferences for aspects of the physical environment around you when you take a test. This may allow you to be more aware of your preferences and will aid your decision making the next time you have to take a computer-based examination.

Risks or Discomfort

There are no known risks to those who take part in this study.

Compensation

We will not pay you for the time you volunteer while being in this study.

Confidentiality

We must keep your study records confidential. Upon your submission of your responses to the survey, your responses will be entered into an electronic database. Your internet provider address will be removed from your record and you will be assigned a participant code. Data using only participant codes will be used for the study and retained for five years.

However, certain people may need to see your study records. By law, anyone who looks at your records must keep them completely confidential. The only people who will be allowed to see these records are:

- The Principal Investigator,
- Faculty Advisor and a professional staff member at Florida Gulf Coast University, who supports the survey software, will be the

only individuals who will have access to the data.

- Certain government and university people who need to know more about the study. For example, individuals who provide oversight on this study may need to look at your data. This is done to make sure that we are doing the study in the right way. They also need to make sure that we are protecting your rights and your safety.

These include:

- the University of South Florida Institutional Review Board (IRB) and the staff that work for the IRB.
- Other individuals who work for USF that provide other kinds of oversight may also need to look at your records.

We may publish what we learn from this study. If we do, we will not let anyone know your name. We will not publish anything else that would let people know who you are.

Voluntary Participation / Withdrawal

You should only take part in this study if you want to volunteer. You should not feel that there is any pressure to take part in the study, to please the investigator. You are free to participate in this research or withdraw at any time. There will be no penalty or loss of benefits you are entitled to receive if you stop taking part in this study.

Questions, concerns, or complaints

If you have any questions, concerns or complaints about this study, please call Ellen Williamson at 239-590-7531 or e-mail to ekwill@fgcu.edu.

If you have questions about your rights, general questions, complaints, or issues as a person taking part in this study, call the Division of Research Integrity and Compliance of the University of South Florida at (813) 974-9343.

If you experience an adverse event or unanticipated problem call Ellen Williamson at 239-590-7531 or e-mail ekwill@fgcu.edu.

If you have questions about your rights as a person taking part in this research study you may contact the Florida Department of Health Institutional Review Board (DOH IRB) at (866) 433-2775 (toll free in Florida) or 850-245-4585.

Consent to Take Part in this Research Study

It is up to you to decide whether you want to take part in this study. If you want to take part, please, please click on the "I agree" button below. By clicking this button you are freely giving your consent to participate in the research study.

*1. Do you agree to participate?

A response is required for this item - Thank you!

○Yes ○No

BACK | Save Progress and Exit | CONTINUE

*2. Have you taken the NPTE?

- Yes
- No

BACK | Save Progress and Exit | CONTINUE

The following survey will ask you questions about your background information, the PT program from which you graduated, your preferences for testing environments, and experiences while taking the National Physical Therapy Examination. Please provide your exact information or your best estimate when answering each question. The survey should take you approximately 10 minutes to complete. You will click on your response to a variety of questions using radio buttons, pull down menus, and an open response box. Click on "continue" to move forward and "back" to return to a previous page. You will also be able to save and return to the survey if you must leave the survey. You will see your progress on the progress bar at the top of each page. Once you have finished the survey, you will submit your final page and be thanked for your participation.

Please provide the following information about yourself:

3. 1. What was your program GPA at the conclusion of your PT program? (e.g. 3.52)

4. What was the highest verbal GRE score you earned (if taken)? (e.g. 450)

5. What was the highest quantitative GRE score you earned (if taken)? (e.g. 500)

6. What age were you when you last took the NPTE? (e.g. 27)

7. What sex are you?

Select:

8. What was the month/year that you graduated with your entry-level PT degree?

(e.g. May 2008)

BACK Save Progress and Exit CONTINUE

9. Which entry-level PT degree did you earn?

Select:

10. Which description BEST describes the type of curriculum you completed for your entry-level PT education?

- CASE-BASED: utilizes patient cases as unifying themes throughout the curriculum
- LIFESPAN-BASED: built around the physical therapy needs of individuals throughout the lifespan
- PROBLEM-BASED: entire curriculum (including basic and clinical science content) is built around patient problems, using a facilitation and independent learning model
- MODIFIED PROBLEM-BASED: uses the problem-based model in the later stages, but the early courses (primarily basic sciences) are presented in the more traditional format of lecture and laboratory
- SYSTEMS-BASED: built around physiological systems (musculoskeletal, neuromuscular, cardiopulmonary)
- GUIDE-BASED: built around the disability model, the patient management model, and the prferred practice patterns in the Guide to Physical Therapist Practice
- TRADITIONAL: begins with basic science, followed by clinical science and then by physical therapy science
- HYBRID: designed as a combination of two or more of the above models

BACK | Save Progress and Exit | CONTINUE

11. Estimate the percentage of your PT program coursework that was delivered on-line/computer-based?

Select:

12. Estimate the percentage of quizzes/tests/exams that you took on-line/computer-based during your PT program.

Select:

13. How would you rate your comfort level for taking exams in a fully on-line format?	Very comfortable				Very uncomfortable
Rate your level from 1-5	○	○	○	○	○

14. What time of day do you prefer to take tests/exams?

- ○ early morning (before 10 am)
- ○ late morning (10am - noon)
- ○ early afternoon (after noon but before 3 pm)
- ○ late afternoon (3 pm or after)
- ○ no preference

BACK | Save Progress and Exit | CONTINUE

The following questions relate to your experience taking the National Physical Therapy Examination (NPTE).

15. What was the total number of times you took NPTE?
(e.g. 1)

16. Have you received the results of your most recent attempt at the NPTE?

◯Yes ◯No

17. What was your total score for your most recent attempt at taking the NPTE (do not answer if you have not received most recent scores)?
(e.g. 610)

18. What was the month/year that you last took the NPTE?
(e.g. 09/10/2007)

19. What was the city/state in which you last took the NPTE (testing center location)?
(Please use format -- Anytown, FL)

20. Was there more than one testing center that offered the NPTE in that specific city?

◯Yes

◯No

21. If yes, would you provide the street on which the center was located?

22. Was there more than one testing room at the testing center (not including break space, etc)?

◯Yes

◯No

BACK | Save Progress and Exit | CONTINUE

Please answer the remaining questions in terms of the LAST time you took the NPTE.

23. When you last took the NPTE, what time of day were you scheduled to begin?

- early morning (before 10 am)
- late morning (10 am – noon)
- early afternoon (after noon but before 3 pm)
- late afternoon (3 pm or after)

24. How many of the SCHEDULED breaks did you take?

Select:

25. During how many of the SCHEDULED breaks did you leave the testing room?

Select:

26. How many of the UNSCHEDULED/OPTIONAL breaks did you take?

Select:

27. During how many of the UNSCHEDULED/OPTIONAL breaks did you leave the testing room?

Select:

BACK | Save Progress and Exit | CONTINUE

28. Estimate the number of workstations that were in the testing room.

Select:

29. Estimate the number of workstations that were in use for the majority of the time that you were taking the exam.

Select:

30. How would you rate the center in which you took the NPTE?
Rate your level from 1-5

New/newly renovated				worn/outdated
5	4	3	2	1

31. Were there exterior windows in the testing room which let in natural light?

Select:

32. What was the level of visual distractions in the room while you were taking the exam?

No distractions			Constant distractions	
1	2	3	4	5

33. Did you use headphones during the exam?

Yes

No

34. Was there a separate room provided for taking breaks?

Select:

35. Was there access to food/drink in the break room?

Select:

BACK Save Progress and Exit CONTINUE

The following section contains various aspects of the physical environment in which you last took the NPTE. Read each line and click the radio button that best describes your preference (left column) and your actual experience when you last took the exam (right column) on the scale provided. Please provide your best estimate of your preference even if you do not feel strongly about the preference.

	What is your preference?		What was your actual experience?	
	Small	**Spacious**	**Small**	**Spacious**
Testing room				

	What is your preference?		What was your actual experience?	
	Enclosed	**Open**	**Enclosed**	**Open**
Testing room				

	What is your preference?		What was your actual experience?	
	Cold	**Hot**	**Cold**	**Hot**
Room climate				

	What is your preference?		What was your actual experience?	
	Dry	**Humid**	**Dry**	**Humid**
Room climate				

	What is your preference?		What was your actual experience?	
	Dim	**Bright**	**Dim**	**Bright**
Room lighting				

	What is your preference?		What was your actual experience?	
	Soft	**Intense**	**Soft**	**Intense**
Room lighting				

	What is your preference?		What was your actual experience?	
	Quiet	**Loud**	**Quiet**	**Loud**
Room				

sound

	What is your preference?							What was your actual experience?						
	muffled						clear	muffled						clear
Room sound														

	What is your preference?							What was your actual experience?						
	Soft						Hard	Soft						Hard
Workstation chair														

	What is your preference?							What was your actual experience?							
	Not adjustable						Highly adjustable	Not adjustable						Highly adjustable	Don't know
Workstation chair															

	What is your preference?							What was your actual experience?						
	Small						Large	Small						Large
Workstation desktop area														

	What is your preference?							What was your actual experience?							
	Not adjustable						Highly adjustable	Not adjustable						Highly adjustable	Don't know
Workstation desktop area															

BACK | Save Progress and Exit | CONTINUE

48. How would you rate the environment in which you took the NPTE?

	It prevented me from performing at my best						**It allowed me to perform at my best**
Effect on performance	○	○	○	○	○	○	○

49. Based solely on the testing environment, how likely would you be to choose the same testing center if you had to take the exam again?

	Not likely						**Highly likely**
Likelihood	○	○	○	○	○	○	○

50. Using your own words, please describe the testing center at which you took the exam. Think about the testing room itself (size, temperature, lighting, sound, etc), the workstation, the chair, the computer set up, the break space. If you took the exam more than once, and used a different center, please describe the differences in the centers.

BACK | Save Progress and Exit | CONTINUE

Thank you for taking the survey.