# INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

**The quality of this reproduction is dependent upon the quality of the copy submitted.** Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

# Validity Evidence of a Computer Simulation Examination to Measure Decision-making Skills in Athletic Training

Jason P. Bennett

A dissertation presented to the Graduate Faculty of Middle Tennessee State University in partial fulfillment of the requirements for the Doctor of Arts degree in Physical Education in the Department of Health, Physical Education, Recreation, and Safety.

August, 2002

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

UMI Number: 3057597

UMI®

UMI Microform 3057597
Copyright 2002 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Validity Evidence of a Computer Simulation Examination to Measure Decision-making Skills in Athletic Training

APPROVED:

Graduate Committee:

Dr. Bill Whitehill

Dr. Doug Winborn

Dr. Kim Shibinski

Head of the Department of Health, Physical Education, Recreation, and Safety

Dean of the Graduate College

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ABSTRACT

Validity Evidence of a Computer Simulation Examination to Measure Decision-making Skills in Athletic Training

The purpose of the study was to collect validity evidence on a computer simulation program to measure decision-making skills in athletic training. Content-related, construct-related, and criterion-related validity data were collected.

Twelve program directors of CAAHEP-accredited athletic training education programs served as subject-matter experts (SMEs). SMEs completed a content validity form and score forms for three simulations. Results demonstrated that all simulations had a minimum mean validity score of 6.33 out of a maximum of eight. Based on these results, and the methods of accumulating content validity data, the Computerized Simulation Test in Athletic Training (CSTAT) contained a high level of content validity in measuring decision-making skills in athletic training.

Proficiency, efficiency, and omission scores were collected from the CSTAT examination. The proficiency score is the sum of the points assigned to all selected options divided by the maximum score possible. The efficiency score is the proportion of an examinee's

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

selections that are indicated options. The omission score is the proportion of contraindicated options that were selected.

Three groups in this study were ATCs (n=23), senior-level undergraduate students (n=31), and junior- or sophomore-level undergraduate students (n=16). ANOVA results established significant group differences ($p<.05$) for overall proficiency, efficiency, and omission scores. Post-hoc testing revealed that for overall proficiency, there were only significant differences ($p<.05$) between the ATC group and the junior-level group, and the senior-level group and the junior-level group. For overall proficiency, there was no significant difference between ATCs and senior-level students. For overall efficiency and omission, only the junior-level and ATC group were significantly different. ANOVA results for each simulation revealed significant group differences for all simulations except simulation #1 and #5 for all three CSTAT scores.

Ten subjects submitted their NATABOC certification exam results. No significant correlations were found between CSTAT and NATABOC written simulation score, but when one outlier scores was removed, two significant correlations ($p<.01$) were found. Correlations between

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

efficiency and written simulation score, and omission and written simulation score were significant ($p<.01$).

It was concluded that the CSTAT examination contained sufficient levels of content-related, construct-related, and criterion-related validity evidence to measure decision-making skills in athletic training.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

# DEDICATION

For Jennifer - thank you for being my partner in everything we do. You are simply the best!

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

TABLE OF CONTENTS

LIST OF TABLES .......... vii

LIST OF FIGURES .......... ix

LIST OF APPENDICES .......... x

CHAPTER

I. INTRODUCTION .......... 1

Statement of the Problem .......... 6

Purpose of the Study .......... 6

Research Questions .......... 8

Limitations .......... 9

Delimitations .......... 9

Significance of the Study .......... 10

Definition of Terms .......... 12

II. REVIEW OF LITERATURE .......... 14

Obtaining Validity Evidence .......... 14

Measuring Decision-making Skills .......... 18

Psychometric Issues of Simulations .......... 22

Scoring of Simulations .......... 24

Development of a Computerized Simulation .......... 27

Computerized Simulations in the Health Professions .......... 32

III. METHODS AND PROCEDURES .......... 38

Content-related Evidence .......... 39

CSTAT Instrument Development .......... 43



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Pilot Study .................................... 45

Construct-related Evidence ...................... 47

Criterion-related Evidence ...................... 49

Data Analysis ................................... 51

IV. RESULTS AND ANALYSIS OF DATA ..................... 52

Description of Participants ..................... 52

Content-related Validity Evidence ............... 54

Construct-related Validity Evidence ............. 59

Criterion-related Validity Evidence ............. 73

V. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS .......... 79

Summary ......................................... 79

Conclusions ..................................... 83

Discussion ...................................... 84

Recommendations ................................. 91

REFERENCES .......................................... 102



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

## LIST OF TABLES

Table

1. Existing and revised domains of athletic training.................................. 31

2. List of software programs and functions........... 43

3. Descriptive data on subjects who attempted the CSTAT examination.................. 53

4. Mean content validity scores by simulation........ 55

5. NATA domain relevancy mean scores and ranking..... 58

6. Mean and standard deviation of overall proficiency, efficiency, and omission scores by group....................................... 59

7. ANOVA results for overall proficiency, efficiency, and omission scores.................. 60

8. Results of Tukey post-hoc test for overall proficiency, efficiency, and omission scores...... 62

9. ANOVA results for proficiency, efficiency, and omission scores for simulation #1................. 64

10. ANOVA results for proficiency, efficiency, and omission scores for simulation #2................. 65

11. ANOVA results of proficiency, efficiency, and omission scores for simulation #3................. 66

12. ANOVA results of proficiency, efficiency, and omission scores for simulation #4................. 67



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

13. ANOVA results of proficiency, efficiency, and omission scores for simulation #5................. 68
14. ANOVA results of proficiency, efficiency, and omission scores for simulation #6................. 69
15. ANOVA results of proficiency, efficiency, and omission scores for simulation #7................. 70
16. ANOVA results of proficiency, efficiency, and omission scores for simulation #8................. 71
17. Significantly different groups by simulation...... 72
18. Raw scores for seniors reporting NATABOC certification results.............................. 74
19. Correlation coefficients between proficiency, efficiency, omission, and written simulation score............................................... 75
20. Correlation coefficients between proficiency, efficiency, omission, and written simulation score without outlier data........................ 77



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

LIST OF FIGURES

Figure

1. Scatter plot of proficiency, efficiency, omission, and NATABOC written-simulation scores ........................................ 76

2. Scatter plot of proficiency, efficiency, omission, and NATABOC written-simulation scores without outlier data ...................... 78



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

LIST OF APPENDICES

Appendix

A. Content Validity Form.............................92

B. Scenario Relevance to NATA Performance Domains.........................................94

C. Web Page for Submission of NATABOC Certification Examination Scores..................96

D. IRB Approval.......................................98

E. Letter of Consent for SEATA Conference Attendees...........................................100



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

# CHAPTER 1

## INTRODUCTION

The National Center for Catastrophic Sports Injury Research (NCCSIR) identified 809 direct catastrophic injuries in high school and college sports from 1982 through the spring of 2000 (NCCSIR, 2002). Between June 1983 and June 1993, 160 athletes died as a result of non-traumatic causes (Van Camp, Bloor, Mueller, Cantu, Olson, 1995). Of these non-traumatic causes, 78% of deaths were due to cardiac causes in high school and college sports (Van Camp et al). In addition to these catastrophic injuries, it was estimated that over 6.5 million youths participated annually in high school athletics. Out of the 6.5 million youths that participated in high school athletics, there are approximately 2 million injuries, 500,000 doctor visits, and 30,000 hospitalizations annually (National Federation of State High School Associations, 1997).

Due to the increase in injuries to athletes, in 1998 the American Medical Association made the recommendation that each high school in the United States employ a certified athletic trainer to care for high school athletes (Lyznicki, Riggs, & Champion, 1999). A certified athletic trainer (ATC) is a skilled professional specializing in

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

athletic health care (National Athletic Trainers' Association, 2001). An ATC predominantly deals with the prevention, recognition and treatment, and rehabilitation of athletic injuries. There are currently over 21,000 certified athletic trainers (ATCs) in the United States (B. Donahoo, personal communication, August 27, 2001). Due to the fact that ATCs may make critical decisions, one component of the ATC's certification process is designed to measure their decision-making skills (NATABOC, 2002b).

The term clinical decision-making has been defined in a variety of ways. Croke (1999) defined decision-making as a course of action utilized to solve client care problems. Leighton & Sheldon (1997) defined clinical-decision making as a combination of creative thinking, clarification, and understanding which allows a person to be an effective problem solver. The National Athletic Trainers' Association Board of Certification (NATABOC) made no distinct clarification between decision-making and problem solving (NATABOC, 2002b). While decision-making has somewhat different definitions in the literature, one method of assessment is accomplished through the use of simulations (Norcini et al., 1986; Tharp, 1990; Engberg & White, 1991; Henry & Holzemer, 1993; Bersky & Yocom, 1994; Mansen & Haak, 1996; Johnson et al., 1998).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Currently, the only way to become an ATC is to pass the NATABOC certification examination. The mission of the NATABOC is to "certify athletic trainers and to identify for the public, quality healthcare professionals through a system of certification, adjudication, standards of practice and continuing competency programs" (NATABOC, 2002a). The first certification exam was conducted in August of 1969 (Grace, 1999). This first examination was a two-part test, the first part contained 150 multiple-choice questions, and the second part was an oral-practical examination. Through 2002, the NATABOC certification examination is a three-part test consisting of a 150 question multiple-choice component, a practical-skills component, and a written simulation examination. The written simulation portion of the examination was added in 1987 (Grace, 1999).

The written simulation examination was intended to measure the decision-making skills or problem solving skills of the examinees (NATABOC, 2002b). The written simulation portion of the examination was developed as a cued-format examination. This cued-format examination contained a list of options for examinees to choose from and the examinee then "selected" those options on the test booklet. "The candidates indicate their selections of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

actions/decisions on the answer booklet that is printed on latent image paper. The candidates utilize a special latent image pen to indicate their selections on the answer booklet. When he/she [sic] makes [sic] a selection, a response/consequence to their action is revealed" (NATABOC, 2002b).

This method of testing is also known as patient-management problems, or PMPs (Swanson, Norman, & Linn, 1995). While this method of simulation testing for the NATABOC certification examination has been shown to be effective with a Kuder-Richardson reliability coefficient of .95, the fidelity of this type of testing is low (NATABOC, 2001). Fidelity was defined as how closely a simulation experience imitates reality (Alessi, 1988). When high stakes examinations are administered, such as licensure and certification examinations, it has been suggested that high levels of fidelity are needed (Gagne, 1954). Since ATCs may face life-threatening situations and critical decisions, it is more appropriate to give a test with higher fidelity than the current written simulation examination. A computer-based simulation exam with video and audio of patients' actions would be an exam with higher fidelity than the NATABOC written simulation exam. However, unlike other health care fields, there is no evidence of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

investigation of the use of computerized simulations to measure decision-making or problem solving skills in athletic training.

In 1988, the National Council of State Boards of Nursing, concerned about an effective way to evaluate nursing clinical competence and problem solving, began a three-year study to develop a computerized clinical simulation test (Bersky, 1996). From this initial investigation, the Computerized Clinical Simulation Test (CST) exam was created and, up until 1999, was to be incorporated into the National Council of State Boards of Nursing Licensure Exam for Registered Nurses (Bersky, Yocom, 1994). Nursing is not the only profession to research the benefits of computerized clinical simulations.

The National Board of Medical Examiners began development of a computer-based examination (CBX) in the early 1970s (Orr, 1988). Also, the dental licensure boards created a not-for-profit corporation, the Dental Interactive Simulations Corporation, to create and validate simulation tests for licensure exams (Alessi & Johnson, 1992). While other health care professions have investigated the implementation of computerized clinical simulations, there is currently only one study in athletic training that has

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

investigated the use of computerized clinical simulations (Castle, 2000).

## Statement of the Problem

In 1999, the passing rate for first-time examinees attempting the NATABOC certification examination was 31.26% (NATABOC, 2000). Of these first-time examinees, 54.81% of students passed the written simulation portion of the examination. Of the three sections of the NATABOC certification examination, the written simulation portion yielded the lowest passing percentage for the 1997 and 1999 examinations (NATABOC, 2000).

While first-time examinees had the lowest passing percentage on the written simulation portion, examinees re-taking all three sections of the certification examination had the highest passing percentage on the written simulation portion (NATABOC, 2000). The preparation of examinees for the written simulation examination could contribute to the initial low passing percentage and the subsequent high passing percentage for the written simulation section.

## Purpose of the Study

The purpose of the study was to collect evidence to validate a computer simulation program to measure decision-making skills in athletic training. The strongest case for

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

test validity is made when a variety of types of evidence agree in supporting validity of the test (Alessi & Johnson, 1992). While there are many types of validity evidence, the various means of collecting validity evidence have been grouped into three main categories: content-related, criterion-related, and construct-related evidence of validity (APA, AERA, & NCME 1985; Smith & Hambleton, 1990; McKenzie, Wood, Kotecki, Clark, & Brey, 1999).

According to Sireci (1998), "content validity refers to the degree to which a test measures the content domain it purports to measure". Sireci also wrote that using content experts to ascertain the domain representation and domain relevance of the test typically produces content validity. Criterion-related validity evidence "demonstrates that test scores are systematically related to one or more outcome criteria" (APA, AERA, & NCME, 1985). According to the Standards for Educational and Psychological Testing (1985), criterion-related evidence is sub-categorized into predictive validity and concurrent validity. Predictive validity is used to predict future performance, while concurrent validity is used when two tests supposedly measuring the same construct are correlated (Dick & Hagerty 1971).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Construct-related evidence "focuses primarily on the test score as a measure of the psychological characteristic of interest" (APA, AERA, & NCME 1985). One method of collecting construct-related evidence is to collect and compare group differences on a test (Payne, 1997). The sampled individuals would be assumed to differ in the construct of decision-making skills.

Therefore, three types of validity evidence were needed to determine if a computerized simulation test had a high level of validity. Several research questions were investigated in the areas of content-related, criterion-related, and contruct-related validity evidence.

**Research Questions**

The following research questions were investigated:

1. Does the CSTAT examination have a high level of content-related validity?
2. Do scores on the computer simulation program differ between known groups of users? (Construct-related validity evidence)
3. What is the correlation of senior-level subjects' scores on the computer simulation program and those same subjects' scores on the written simulation portion of the NATABOC certification examination? (Criterion-related validity evidence)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

## Limitations

This study had the following limitations:

1. Computers at the subjects' institution had to be powerful enough to run the computer simulation program, and some may not have had sufficient resources to obtain such computers. The computerized simulation program required sufficient hard drive space, 32mb of random-access memory (RAM), a sound card, and QuickTime 4.0 or later be installed on the computer. All computers ran the Windows 95, 98, or 2000 operating systems.

## Delimitations

The following were the delimitations of the study:

1. Content experts were ATCs who were employed as an instructor of athletic training in a CAAHEP-accredited athletic training education program.
2. Subjects were all either: ATCs, senior-level, junior-level, or sophomore-level athletic training students.
3. Subjects were athletic training students or certified athletic trainers from District 9 of the National Athletic Trainers' Association (NATA). NATA District Nine included the states of Kentucky, Tennessee, Georgia, Alabama, Florida, Mississippi, and Louisiana.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

4. The content of the computer simulation program was limited to 650 megabytes of space, which was the maximum possible amount of data that can be stored on a CD-ROM.

**Significance of the Study**

The results of this study will help determine if there is an alternate method of assessing clinical decision-making skills in athletic training. In addition, the computer simulation program could be used as a resource tool to detect deficiencies in students preparing to take the NATABOC certification exam. Currently, there is no research available in athletic training that has focused on the investigation of the specific preparation of students for the written simulation portion of the NATABOC certification examination.

The written simulation portion of the certification examination differs from a traditional multiple-choice examination. An instrument that familiarizes the students with this type of test may benefit students preparing for the certification exam. Since a computerized simulation program would address this familiarity issue as well as increase the fidelity of the experience by the student, a computerized simulation program could be a valuable tool to prepare students for the NATABOC certification examination.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A computerized simulation program could also be created for the purpose of continuing education for ATCs. Simulations in the dental profession are already being used for continuing education (Johnson, Wohlgemuth, Cameron, Caughman, Koertge, Barna, Schulz, 1998). A life-threatening skill, such as spine boarding an athlete, may not be frequently used by the ATC, and a computerized simulation could be used to educate and test ATCs on this skill. This type of continuing education may be beneficial because it would allow certified athletic trainers to remain knowledgeable about subjects that they may not be exposed to frequently.

The computer simulation program could also help educators recognize weaknesses, either in their students or in their own educational program. These educators could then make modifications to specific courses to address these weaknesses. Finally, if the computer simulation program can increase the performance of students and certified athletic trainers in real-world situations, then the general public that comes into contact with these professionals will benefit as well.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

## Definition of Terms

- Decision-making skills - course of action utilized to solve client care problems (Croke, 1999).
- Validity - extent to which a test measures what it is supposed to measure (Baumgartner & Jackson, 1999).
- Fidelity - how closely a simulation imitates reality (Alessi, 1988).
- Athletic Training - the allied health profession dedicated to the prevention, evaluation, treatment, and rehabilitation of injuries to the physically active population (National Athletic Trainers' Association, 2000).
- Athletic Trainer - The certified athletic trainer (ATC) is a highly educated and skilled professional specializing in the prevention, treatment and rehabilitation of injuries (National Athletic Trainers' Association, 2000).
- Content Validity - the degree to which a test measures the content domain it purports to measure (Sireci, 1998).
- Criterion Validity - test scores are systematically related to one or more outcome criteria (APA, AERA, & NCME 1985).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

- Construct Validity - focuses primarily on the test score as a measure of the psychological characteristic of interest (APA, AERA, & NCME 1985).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

# CHAPTER II

## REVIEW OF LITERATURE

Currently there is only one other study in athletic training that has investigated the use of a computerized simulation program (Castle, 2000). In order to collect evidence to validate a computer simulation program to measure decision-making skills in athletic training, several areas of the literature needed to be reviewed. This chapter addresses the following areas: 1) obtaining validity evidence, 2) measuring decision-making skills, 3) psychometric issues of simulations, 4) scoring of simulations, 5) development of a computerized simulation, and 6) computerized simulations in the health professions.

### Obtaining Validity Evidence

According to the Standards for Educational and Psychological Tests by the American Psychological Association (1985), there are three types of validity evidence: content, criterion, and construct. While the methods used to collect content evidence vary, collecting criterion and construct-related evidence has become more standardized.

Criterion-related validity evidence "demonstrates that test scores are systematically related to one or more outcome criteria" (APA, AERA, & NCME 1985). According to

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the Standards for Educational and Psychological Testing (1985), criterion-related evidence is sub-categorized into predictive validity and concurrent validity. Predictive validity is used to predict future performance, while concurrent validity is used when two tests supposedly measuring the same construct are correlated (Dick & Hagerty 1971).

Construct-related evidence "focuses primarily on the test score as a measure of the psychological characteristic of interest" (APA, AERA, & NCME 1985). One method of collecting construct-related evidence is to compare group differences on a test (Payne, 1997; Melnick, 1990). In this study, if there are significant group differences on scores earned from the CSTAT examination, then the sampled individuals could differ in the construct of decision-making skills.

"Content validity refers to the degree to which a test measures the content domain it purports to measure" (Sireci, 1998). Collecting content validity evidence is typically judgmental and can be obtained in different ways (Melnick & Henk, 1997). Crocker (1997) noted that content validity studies have been traditionally about collecting evidence on "a) the relevance of the test item content to the knowledge domain of interest and b) the balance of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

coverage of the items in relation to the breadth of the domain." In addition, Sireci (1998) wrote that using content experts to ascertain the domain representation and domain relevance of the test typically establishes content validity.

Crocker (1997) also stated that evidence of content validity is obtained by having qualified experts review test items and matching those items to a domain. Sireci (1998) stated, "content validity studies typically involve a relatively small number of participants who are required to make a variety of important judgments." Sireci gave several recommendations regarding the use of subject-matter experts (SMEs) in content validity studies. His recommendations were:

1. Use competent and representative SMEs
2. Use 15 or fewer SMEs
3. Minimize the burden on SMEs
4. Use rating scales of sufficient length
5. Provide monetary compensation for SMEs

For recommendation 1, Sireci stated that SMEs must be familiar with the content tested and knowledge and skill levels of the tested population. In addition, geographic, racial, and ethnic diversity in the population should be represented in the sample. For recommendation 3, Sireci advised researchers to prepare individual content validity forms for each SME with a rating scale on each form. For

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

recommendation 4, Sireci stated that an 8-point scale would be best so that SMEs would avoid the neutral response.

However, the methods for obtaining content validity evidence were not consistent within the literature. Yalow and Popham (1983) stated that there is no uniformity for content reviewers or in the way the resulting data have been analyzed. Examples of this non-uniformity were apparent in several studies regarding content validity.

Gill and Deeter (1988), in their development of a sport orientation questionnaire, used five graduate students in an advanced sport psychology seminar class as the content experts. This study is in stark contrast to the study by Kulinna and Silverman (1999) who used 28 experts in sport pedagogy in development of an instrument to measure teachers' attitudes toward teaching physical activity and fitness. These experts were defined as persons who were employed in tenure-track professorial positions in higher education. In addition, Barrett (1992) gave the recommendation of using six to eight SMEs in collecting content validity evidence.

Non-uniformity in the methodology for measuring content validity in computer-simulation studies has also been apparent. In Henry and Holzemer's (1993) study, using the commercially available Tach-Man simulation program,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

content validity was established by "… comparison with clinical cases and a review of the literature according to the author". No quantifiable evidence was given for the content validity of this simulation. In contrast, the Dental Interactive Simulations Corporation used two teams of oral health professionals to ensure content validity (Johnson, et al., 1998). One team, the design/content team consisting of 7 experts, focused on the design of the simulations, and the scoring team, consisting of 10 experts, focused on the scoring component of each simulation. Two members of the design/content team were also members of the scoring team. In a similar study, content validity for the Computerized Clinical Simulation Test for nursing used two different 12-member panels of master's prepared RNs from across the country (Bersky, 1996).

**Measuring Decision-making Skills**

Traditionally, multiple-choice questions have been used to measure decision-making skills. However, in the 1950s, in response to dissatisfaction with traditional multiple choice tests, patient-management problems (PMPs) were developed (Swanson, Norcini, & Grosso, 1987). PMPs begin with an opening scenario and then the examinee proceeds through a series of "scenes" in which information

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

is gathered relating to the problem in order to determine an appropriate intervention for the patient (Swanson, Norman, & Linn, 1995). While this format of testing is more difficult to administer, write, and score, this format does provide feedback to examinees throughout the testing process (Norcini et al., 1986).

Written simulations, also known as PMPs, were seen as an objective, standardized technique which could pose important clinical problems in realistic ways (Swanson et al., 1987). Norcini, Meskauskas, Langdon, & Webster (1986) stated that the reason PMPs were developed was to more closely simulate the clinical encounter as compared to multiple-choice questions. Simulations are attempts to imitate reality, placing students in realistic settings and having them perform in realistic ways (Alessi & Johnson, 1992).

Written simulations are the current standard of testing decision-making skills in athletic training (NATABOC, 1999). According to the NATABOC (1999), the simulation portion of the exam is appropriate because ATCs use professional judgement to make important decisions in situations where the element of time is critical.

Several advantages of clinical simulations, both written and computerized, have been documented. One benefit

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of using simulations is that it allows the subject to practice in a non-threatening environment (Engberg & White, 1991; Henry & Holzemer, 1993). Engberg and White stated that simulations allow the subject to receive feedback and suggestions for additional self-study, and it allows subjects to gain exposure to clinical problems that may not be available in a clinical rotation. Forker and McDonald (1996) stated that computer simulation testing allows the presentation of a fuller scope of client situations that are truer to reality than multiple-choice exams. Alessi and Johnson (1992) stated that simulations reduce costs and provide replicable education experiences as compared to live patient performance exams.

Scheuneman, Van Fan, & Clyman (1998) stated that computer-based case simulations as a performance assessment instrument could also be used to measure patient management skills in a realistic environment with simulated time and unfolding clinical situations. Other authors suggest that the use of clinical simulations eliminates the issue of inter-rater reliability that is associated with traditional performance evaluations (Henry & Holzemer, 1993; Henry & Waltmire, 1993). According to Henry and Holzemer (1993), clinical simulations also allow the evaluation of decisions without a threat to patient safety, which gives the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

opportunity to evaluate decision-making during infrequent, but life-threatening situations.

Gates, Greene, & Kris-Etherton (1990) noted that written simulations are most frequently used because they are less expensive, easier to administer, and easier to score than live simulations. McGuire and Babbott (1967) also suggested that simulation techniques are of value in assessing problem-solving skills in clinical medicine. In addition, Iserson (1990) commented that computer simulations will decrease the inconsistency of clinical experience among students who only see procedures that are done during their rotation. Forker and McDonald (1996) also stated that a computer simulation "...expands the scope of the evaluation program beyond knowledge acquisition, bringing an assessment approach that offers greater fidelity to the measurement of clinical competency in actual performance".

In summary, utilizing simulations is an efficient method of assessing decision-making skills. Furthermore, there are a myriad of benefits to using simulations. The most common advantages are 1) it allows users to gain knowledge and skill in an area that may not be common-place in their clinical rotation, 2) it eliminates the inter-rater reliability factor of live simulations, and 3)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

students can be assessed in a safe environment, without risk to patients.

**Psychometric Issues of Simulations**

PMPs were commonly used on medical licensing examinations until the late 1980s, but recently questions regarding the psychometric soundness of this type of examination have risen (Swanson et al., 1987). Swanson suggested that cued versus uncued PMPs can lead to psychometric problems. In cued formats, all options are presented to the examinee, while in uncued formats, options must be self-generated by the examinee.

It was concluded that more data were collected in cued simulations than uncued (Swanson et al., 1987; Newble, Hoare, & Baxter, 1982; McCarthy, 1966). McCarthy (1966) stated that since history-taking and physical examination are non-cued activities of physicians, a cued examination may not be appropriate. However, despite these issues, Swanson et al. (1987) stated that cueing problems were more severe in the written simulations rather than the computer-based simulations because examinees may read ahead or go back and reveal answers based on what was found in later sections. Simulations can be made less cued through the use of long, standardized lists of options in both written and computer-based simulations (Swanson et al., 1987).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In their review of written and computer-based simulations, Swanson et al. (1987) noted that while there were a large number of studies that contained two or more simulations, very few studies had reported intercase correlations. Intercase reliability was defined as the reliability of profiency scores between a set of simulations. Of the studies reviewed, intercase reliabilities ranged from 0.12 to 0.49, a low reliability. Swanson recommended that at least eight to ten cases be used in a simulation exam.

Interestingly, Swanson et al. (1987) found little evidence that simulations provided unique measurement information that was not available through traditional multiple-choice formats. In Swanson's review, there were several studies that had small convenient groups and inadequate data analysis. Based on these studies Swanson concluded that it was unclear if simulations have any real advantages beyond greater face validity.

While Swanson concluded that the advantages of simulations are unclear, other authors disagree. Peitzman, Nieman, and Gracely (1990) investigated the correlation of medical students' ability to answer "higher order" multiple-choice questions and their perceived level of performance in the clinical setting. In their conclusions,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Peitzman et al. noted that higher order questions fail to measure effective clinical performance. One suggestion for measuring clinical performance was the use of simulations. "Ultimately, the use of multiple high-fidelity clinical simulation may emerge as the only satisfactory controlled assay of clinical competence" (Peitzman, Nieman, and Gracely, 1990).

Although there are disagreements regarding the psychometric issues with cued PMPs, the written simulation portion of the NATABOC certification exam is a series of cued PMPs. One method of decreasing the psychometric issues that were noted by Swanson et al. (1987) was to create a computerized version of the written simulation portion of the NATABOC certification exam. By computerizing this portion of the test, examinees will not be able to "look ahead" or "go back" to select answers as suggested by Swanson.

**Scoring of Simulations**

Several different types of scoring procedures have been implemented in both written and computer simulations. McGuire and Babbott (1967) explained that the most promising system of scores was the inclusion of efficiency, proficiency, errors of omission, errors of commission, and a composite index of overall competence.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Swanson et al. (1987) in their review of simulation experiments noted that the three most common scores for simulation tests were profiency, efficiency, and omission scores. Swanson et al. defined these scores as:

> The proficiency score is the sum of the points assigned to all selected options divided by the maximum score possible. The efficiency score is the proportion of an examinee's selections that are indicated options. The omission score is the proportion of contraindicated options that were selected.

According to Swanson et al., obtaining three scores instead of one can give a further indication of the subjects' decision-making skills. These three scores can help determine if a subject is choosing too many incorrect options or choosing too many or too few options altogether.

Different scoring systems have also been used to measure each action within a simulation. In Norcini et al. (1986), subjects were given either +9, +4 or +1 points depending on the importance of the action taken. Another study utilized scoring of +1 or 0 for each action, with no weight given for importance of the action (Clauser, Subhiyah, Piemme, Greenberg, Clyman, Ripkey, & Nungester, 1993). In Henry and Waltmire's (1992) study, essential

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

items were scored as +1, items that were neither harmful nor helpful were scored a 0, and items that were harmful were scored -1. The commercial simulation program Medi-Sims used a scoring range of +3 for actions that are very important to patient care to a -3 for actions that are very detrimental to patient care (Lowdermilk & Fishel, 1991).

The NATABOC's written simulation examination has five categories for scoring (NATABOC, 1999). The highest category is "Clearly Indicated" and is defined as being essential in resolving the problem. The next category is "Indicated" and is defined as being helpful in resolving the problem. The third category is "Neutral" and is defined as being neither helpful nor harmful in resolving the problem. The fourth category is "Contraindicated" and is defined as being detrimental in resolving the problem. The last category is "Clearly Contraindicated" and is defined as detrimental in resolving the problem and resulting in immediate harm (NATABOC, 2000).

In the review of computer and written simulations by Swanson et al. (1987), the authors noted that "changes in the number of categories into which options are classified and alterations in the weights assigned to those categories make little difference". Based on Swanson's conclusions,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the computer simulation program developed for this research project followed the system utilized by the NATABOC.

**Development of a Computerized Simulation**

McGuire and Babbott (1967) stated that a simulation must have five characteristics. The first characteristic was that the simulation must be initiated by information of the type a patient gives a physician. Secondly, the simulation must require a series of sequential, interdependent decisions representing the various stages of the management of the patient. Thirdly, the examinee must be able to obtain information about the results of each decision. Fourthly, it must be impossible for the examinee to retract a decision. And finally, the problem must be constructed to allow different medical approaches.

In their suggestions for creating simulation exercises, the authors stated that each section must give the appearance of random listing but in fact be a carefully structured group of procedures. Another suggestion was that minimal, if any, data are gratuitously provided so that all decisions become the responsibility of the examinee.

Friedman (1995) stated that simulation designers have two options when designing simulations, either create a cued or uncued simulation. Each of these approaches has a

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

distinct disadvantage. The disadvantage of the cued simulation is related to the problems mentioned by Swanson et al. (1987) where subjects collect more data as compared to the uncued simulation. The disadvantage of the uncued simulation is the need for the computer to interpret "natural language" requests from the user (Friedman, 1995).

Although there are disadvantages, a cued computerized simulation was created since the NATABOC's written simulation portion of the certification exam is a cued simulation. Computerizing a cued simulation can limit the disadvantages according to Swanson et al. (1987), by not allowing a user to "look ahead" or "go back" to select answers.

Currently, there is only one commercially available computerized simulation program for the profession of athletic training, SIMS Master, developed by Shearer and Breitbach (1997). While this program is similar to the written simulation portion of the NATABOC certification exam, it does not contain multimedia capabilities such as video or audio. Engberg and White (1991) suggested that incorporating video into simulations could increase the effectiveness.

Therefore, a new computerized simulation program was developed that incorporated video and audio to assess

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

athletic trainers' clinical decision-making skills. A non-commercial computerized simulation program was previously created by Barker in 1997 called Computerized Athletic Training Simulation (CATS). With permission from the author of CATS, this program was used as a basis for creating a new simulation program.

The computerized simulation that was developed needed to be representative of the current performance domains set forth by the NATABOC. According to the 1999 Role Delineation study of the National Athletic Trainers' Board of Certification (NATABOC, 1999), there are six performance domains for certified athletic trainers. These include: prevention; recognition, evaluation and assessment; immediate care; treatment, rehabilitation and reconditioning; organization and administration; and professional development and responsibility.

In the NATABOC's study of the importance of each domain, 716 certified athletic trainers responded, and the importance of the performance domains was ranked. These rankings of importance of domains were: immediate care; recognition, evaluation and assessment; treatment, rehabilitation and reconditioning; prevention; professional development and responsibility; and organization and administration.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Additionally, in the NATABOC study the criticality of each of these domains was assessed. The criticality was defined as the lack of knowledge which could have significant adverse safety consequences. The same order of domains was given for the criticality as was the importance.

However, the NATA is in the process of updating these performance domains to more accurately reflect the performance duties of certified athletic trainers. The Education Council, a committee of the NATA, has made revisions and has further delineated these domains into 12 new educational domains (Educational Council Web Site, 2000). Table 1 contains the 12 new educational domains presented to the NATA Board of Directors by the Educational Council.

Therefore, the eight simulations that were developed took into consideration the importance and criticality of the domains of athletic training, with more emphasis placed on the three most important and critical performance domains: immediate care; recognition, evaluation, and assessment; and treatment, rehabilitation and reconditioning.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 1

*Existing and revised domains of athletic training*

| Existing Domain | Revised Domain |
|---|---|
| 1. Prevention | • Risk Management |
| 2. Recognition, Evaluation And Assessment | • Assessment and Evaluation |
| 3. Immediate Care | • Acute Care;<br>• General Medical Conditions and Disabilities;<br>• Pathology of Injury and Illness;<br>• Pharmacological Aspects of Injury and Illness;<br>• Nutritional Aspects of Injury and Illness |
| 4. Treatment, Rehabilitation and Reconditioning | • Therapeutic Exercise;<br>• Therapeutic Modalities |
| 5. Organization and Administration | • Health Care Administration;<br>• Professional Development and Responsibilities |
| 6. Professional Development and Responsibility | • Psychosocial Intervention and Referral |

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Computerized Simulations in the Health Professions**

Only a small number of studies have been conducted in the health professions to investigate different aspects of computerized simulations. These aspects include simulations versus multiple-choice questions, attitudes towards computerized simulations, and the effectiveness of measuring decision-making skills with simulations.

Two studies have investigated the comparison of computerized clinical simulation and multiple-choice questions. In the study done by Norcini et al. (1986), the purpose of the study was to compare the computer-aided simulation of the clinical encounter (CASE) to multiple choice questions and PMPs. In development of the CASE program, 10 cases were created that reflected general problems within the internal medicine profession. The researchers of the study concluded that multiple-choice questions and patient management problems were more reliable and efficient than the CASE program. However, in a subjective rating of the CASE program by subjects, it was judged to be a more relevant measure of history and physical examination skills than multiple-choice questions and patient-management problems.

In a similar study comparing simulations to multiple-choice questions, Henry and Holzemer (1993) found that

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

there was a moderate relationship between proficiency score on a written simulation, self-evaluation of expertise, and knowledge on a multiple-choice cognitive examination for the management of myocardial infarction patients. Since their results were only moderately correlated, Henry and Holzemer recommended that clinical simulations would be useful as an adjunct to other methods of evaluation, and may not be appropriate as a stand-alone measurement of decision-making skills.

Another aspect of research on computerized simulations in the health professions has been the attitudes of subjects. Wong, Wong, and Richard (1992) described the implementation of a computerized clinical simulation as an evaluation strategy for nursing clinical decision-making. Specifically, the authors were interested in the subjects' attitudes towards the computerized clinical simulation. The researchers concluded that subjects' attitudes as a whole were positive. A majority of subjects (64%) perceived the use of computer simulations to be a valid method of assessing clinical decision-making, and 75% thought that it was less time consuming than traditional written evaluations. Conclusions based on this study are difficult to interpret since the authors never stated the total number of subjects in the study. Even with the low

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

passing rates, the authors concluded that "implementing CAI clinical simulation as a strategy for evaluating student's decision-making skills has convinced us that it is an objective, efficacious, and pragmatic method for measuring an aspect of clinical performance that seems somewhat elusive".

Interestingly though, the study by Wong, Wong, and Richard (1992) reported that about half of the subjects perceived the computer simulation to be inferior to the written method of evaluation. Some reasons for this perception was that the simulation gave a limited number of choices for intervention, did not allow for personal style of nursing practice, and that answers were slightly different than what was learned in class.

The main research focus has been the effectiveness of a computerized clinical simulation to measure decision-making skills. The National Board of Medical Examiners has investigated the use of the Computer Based Examination (CBX) since the 1970s (Melnick, 1990). This CBX test was an uncued-type of clinical simulation examination. Melnick noted that psychiatric, behavioral, and social problems were difficult to simulate because of the central role of interpersonal communication in managing these problems. In addition, Melnick noted that neither subjects' prior use of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

computers nor computer anxiety had a significant effect on performance on the CBX. A similar type simulation has been developed for licensure in nursing.

The National Council of State Boards of Nursing began research into the development of a computerized simulation to measure clinical decision-making skills (Bersky & Yocom, 1994). From this research, the Computerized Clinical Simulation Testing (CST) was developed. Bersky and Yocom (1994) stated that "the degree of realism available through a computer-administered interactive simulation surpasses that available via a paper-and-pencil simulation but falls short of that available through direct observation of live patient encounters." Bersky and Yocom also noted that CST could be used as a teaching tool for assessment of students within an educational program. Since CST was designed to measure the application of the nurse's clinical decision-making for the management of patient care, CST could also be used as an assessment tool to identify re-education content for previously licensed nurses who have been out of clinical practice for a period of time.

In a smaller study, Lowdermilk and Fishel (1991) found that computer-assisted instruction (CAI) using simulations in nursing increased subjects' decision-making skills, although not significantly. In this experimental study, an

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

instructional simulation was used for the treatment group and scores were compared to the control group. The 26 subjects in the experimental group did not have a significantly higher decision-making score than the 20 subjects in the control group, although mean decision-making scores were higher for the experimental group. Lowdermilk and Fishel used a commercially available simulation program (Medi-Sims) as their evaluation tool.

While there have been several studies using computerized simulations in the health professions, there has been only one study in athletic training that has investigated the use of computerized simulations. Castle (2000) investigated the effectiveness of the Computerized Traditional Athletic Training Simulation Instrument (CTATSI). In the development and content validity portion of this research, Castle utilized 3 subject-matter experts that had at least 5 years experience as a certified athletic trainer and were involved with a CAAHEP-accredited athletic training education program. Castle's conclusion was that the CTATSI was not an effective predictor of subjects' score on the written simulation portion of the NATABOC certification examination.

Castle included several recommendations for further research. One recommendation was for further investigation

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

into the predictive success of a computerized simulation exam. Another recommendation was to include video and audio components in a computerized simulation to create a "virtual" simulation. This recommendation was similar to a recommendation from a different study.

Engberg and White (1991), in their summary of using interactive video simulations, noted that clinical decision-making would be most effective when learned within the context of actual clinical problems. Since athletic trainers, like other health professionals, gather information from patients, clinical decision-making could be most effective with the addition of video of patients answering questions from the examinee.

Based on the review of literature, there was a variety of psychometric issues with PMPs and computerized simulations. These psychometric issues can be limited by following the recommendations of Swanson et al (1987). In addition, other medical professions such as nursing, dentistry, and medicine have investigated the effectiveness of computerized simulation examinations. In the profession of athletic training, there has been only one study investigating the effectiveness of a computerized simulation examination.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

# CHAPTER III

## METHODS

The purpose of the study was to collect evidence to validate a computer simulation program to measure decision-making skills in athletic training. The following research questions were investigated:

1. Does the CSTAT examination have a high level of content-related validity?
2. Do scores on the computer simulation program differ between known groups of users? (Construct-related validity evidence)
3. What is the correlation of senior-level subjects' scores on the computer simulation program and those same subjects' scores on the written simulation portion of the NATABOC certification examination? (Criterion-related validity evidence)

The methods used to collect content-related evidence were a) subject recruitment, b) scoring of scenarios, c) and ranking of scenarios. The CSTAT examination was created and the pilot study conducted after the content-validity evidence was analyzed. After the pilot study was completed, construct and criterion-related evidence was collected.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

## Content-related evidence

Subjects

Subject-matter experts (SMEs) were program directors of CAAHEP-accredited athletic training education programs. Program directors of 131 CAAHEP-accredited athletic training education programs were sent an e-mail asking for participation in the study. Thirty-nine program directors replied indicating their initial agreement to participate in the study. A random sample of 32 program directors from the 39 who responded were selected in order to achieve a geographically diverse sample. These subject-matter experts (SMEs) were all certified athletic trainers (n=12) with a mean of 14.4 years of experience as a certified athletic trainer.

Sireci (1998) noted, "content validity studies typically involve a relatively small number of participants who are required to make a variety of important judgments." The recommendations given by Sireci were followed as best as possible. Sireci's recommendations were:

1. Use competent and representative SMEs
2. Use 15 or fewer SMEs
3. Minimize the burden on SMEs
4. Use rating scales of sufficient length
5. Provide monetary compensation for SMEs

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

## Method

A packet of materials was sent to each of the 32 SMEs. This packet included a descriptive questionnaire, human-subjects form, a score form of 3 random selected scenarios out of 8, a content validation form (Appendix A), and a NATA domain relevancy form (Appendix B). SMEs were asked to return the packet within 1 month. Twelve SMEs completed the packet of information and returned it to the researcher.

The questions used in the content validation form were similar to those used by the NATABOC for the written simulation portion of the NATABOC certification examination (NATABOC, 2001). The questions used by the NATABOC to help determine content validity are:

1) Does the problem flow in a logical manner?
2) Is the opening scene plausible?
3) Is there sufficient information available to the candidate to solve the problem?
4) Are the sections in proper order?
5) Are there appropriate tests/actions/options to cover all of the incorrect actions?
6) How essential is mastery of the knowledge, skills, and abilities in this scenario?

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

7) How important are the knowledge, skills, and abilities assessed by this scenario to the assessment of performance as an entry-level athletic trainer?

8) How frequently would an entry-level athletic trainer be expected to encounter a case similar to that found in the scenario?

The educational domains of athletic training were used as the table of specifications for development of the CSTAT (NATABOC, 1999). The three most important and critical domains were immediate care; recognition, evaluation, and assessment; and treatment, rehabilitation and reconditioning (NATABOC, 1999). Therefore, the topics of the eight simulations were designed to fit into these three areas. The topics of the 8 simulations were: spine boarding of football player; evaluation of shoulder injury; procedures for administering medications; rehabilitation of a patient after anterior cruciate ligament reconstruction; utilization of modalities for treatment of an injury; evaluation of knee injury; cardio-pulmonary resuscitation of an athlete; and prevention of low-back injury. All eight of these topics are a part of the NATA's Athletic Training Clinical Proficiencies (NATA, 1999).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

SMEs gave a score for each of the options listed: a score of +2 for clearly indicated option, +1 for an indicated option, 0 for a neutral option, a -1 score for a contraindicated option, and a -2 for a clearly contraindicated option. This scoring of simulation options is similar to the scoring on the written simulation portion of the NATABOC certification exam (NATABOC, 2000). Based on the results obtained from the SMEs, options on each screen were given a score depending on the majority opinion of the SMEs. If there was no majority of opinion on a particular option, it was either deleted or re-worded and scored subjectively. The SMEs were asked to fill out all forms and return the content validity data to the researcher approximately 1 month after receiving the scenarios.

Collecting content-validity evidence was in contrast to other similar studies. Other studies either did not state a number of SMEs that reviewed the content validity (Lowdermilk & Fishel, 1991; Wong, Wong, & Richard, 1992) or was deemed content valid by the author of the simulation program (Henry & Holzemer, 1993). After collecting content-related validity evidence, the development of the CSTAT instrument began.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

## CSTAT Instrument Development

To create the simulation program, the multimedia software Macromedia Director 8.0 was used. Macromedia Director is widely known as a leader in the multimedia software field. Kerman (2000) wrote on his review of Macromedia Director 8 Shockwave Studio, "It's the best-and only-tool for creating highly interactive multimedia projects and Shockwave movies".

One advantage of Director 8 is its ability to create multimedia programs for both Macintosh and Windows-based machines. Although this version of the CSTAT instrument was a made for the windows platform only, this cross-platform feature was deemed important for future research. A multitude of different media and software programs were necessary to deliver a quality multimedia experience. Table 2 contains a list of software used for creation of the computer simulation program.

Table 2

*List of software programs and functions*

| Software Program | Function of Programs |
|---|---|
| Adobe Photoshop® 5.5 | Graphics |
| Adobe Premiere® 6.0 | Video Editing |
| Macromedia Director® 8.5 | Multimedia Synthesis |
| Adaptec CD Burner® | Creation of CD-ROMs |

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A Dell 4100/800 with 384 megabytes of random-access memory (RAM) was used to create the computerized simulation program. The Dell computer also had a Firewire-video PC card to digitize video, a CD burner, and a 100-megabyte Zip drive. In addition, a Sony DCR-VX2000 digital camcorder was used to videotape simulated patients.

The simulation program was designed for use on a computer with only 32 megabytes of RAM. With this in mind, the color level of all graphics was limited to thousands of colors, all movies were limited to 240 X 180 pixel size and compressed using Sorenson video compression. Throughout the creation of the program, the program was tested on different Windows-based machines to ensure proper functioning.

Upon completing the 8 simulations, the overall scores of proficiency, efficiency, and omission were shown to the subject. These scores were congruent with the recommendations made by Swanson et al. (1987). The user's total score divided by the total score possible calculates proficiency scores. Efficiency scores are calculated by the total number of indicated or correct options selected divided by total number of selections made. Omission scores are the opposite of efficiency scores, it is

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

calculated by total number of contraindicated or incorrect options selected divided by total number of selections made. In addition to the overall score, each scenario had proficiency, efficiency, and omission scores. Upon completion of the CSTAT examination and verification that the CD-ROM would work properly, the pilot study began.

**Pilot Study**

The initial CSTAT examination was administered to 21 subjects. Most subjects were students affiliated with Middle Tennessee State University. Two senior-level subjects were recruited from the University of Alabama for the purposes of the pilot study. Of these 21 subjects, 6 were senior-level, 6 were junior-level, and 9 were sophomore-level students.

Subjects were given the initial CSTAT CD-ROM and instructed to use a computer in the departmental computer lab. These computers had audio available, had QuickTime 4.0 or later installed, and were functioning properly. Upon completion of the CSTAT examination, subjects filled out the informed consent form, a CSTAT score sheet, and a ten-question questionnaire. The purpose of the questionnaire was to receive feedback from the subjects about the CSTAT examination and to find any errors in the program. Feedback from subjects revealed at least 4 errors

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

in the CSTAT examination, which included a misspelled word, and video or images not appearing when they should.

A one-way ANOVA (.05 alpha-level) was performed to find any group differences in proficiency scores among the subjects. A Tukey post-hoc test was also used to determine which groups differed. Results of the one-way ANOVA revealed a significant difference with an F-value of 6.53, significant at the .003 alpha-level. Results of the Tukey post-hoc test indicated that only the senior-level subjects differed from the sophomore-level subjects (.008 significance), and the junior-level subjects scored significantly better than the sophomore-level subjects (.012 significance). However, there were no group differences between senior-level and junior-level subjects.

Based on the pilot study, several modifications of the initial CSTAT examination were made. First, all errors in the program were located and corrected. This included video not appearing at the appropriate time and correction of misspelled words. Also, digital video was recreated due to the low-level of audio that was discovered during the pilot study. The on-screen instructions were improved and other minor imperfections were corrected.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Construct-related Evidence**

Construct-related evidence "focuses primarily on the test score as a measure of the psychological characteristic of interest" (APA, AERA, & NCME 1985). One method of collecting construct-related evidence is to collect group differences on a test (Payne, 1997). If there were group differences on the CSTAT, one possible assumption is that the sampled individuals differed in the construct of decision-making skills. The groups used to collect construct-related validity evidence were certified athletic trainers (ATCs), senior-level undergraduate students, and junior-level or sophomore-level undergraduate students.

Subjects

All undergraduate subjects were recruited at the 2002 Southeast Athletic Trainers' Association (SEATA) Annual Student Symposium. This symposium is divided into two tracks. One is for students who have senior standing at their school and are preparing to take the NATABOC certification exam. The other track is for other students who have junior or lower standing. This symposium is held in District nine of the NATA, but is not limited to students from this district.

In addition to the subjects at the SEATA conference, ATCs were also recruited for this study. Each of the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

original 32 SMEs that were used for content-related validity evidence were asked to recruit 3 ATCs from their respective university.

Methods

Permission was obtained to recruit subjects at the SEATA conference in February, 2002. Subjects were recruited from both tracks at this conference by informing them of the research, obtaining signed informed consent forms, and completing a descriptive questionnaire. Subjects were then given a packet that contained the CSTAT CD-ROM, score sheet, and a stamped envelope addressed to the researcher. The descriptive questionnaire was used to identify whether the subject was an ATC, senior-level subject, junior-level subject, or sophomore-level subject.

Subjects were asked to attempt the CSTAT examination within 1 month of the conference. Immediately following completion of the CSTAT examination, subjects recorded all scores earned and mailed the scores to the researcher. The mean proficiency, efficiency, and omission scores of each group were compared to determine if known groups of athletic training students and ATCs differed when taking a computerized simulation program.

A different packet of materials were mailed to each SME which included the CSTAT CD-ROM, 3 informed consent

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

forms, 3 score sheets, 3 descriptive questionnaires, and 3 stamped envelopes addressed to the researcher. The descriptive questionnaire was used to identify the subjects as ATCs. SMEs were asked to have ATCs attempt the CSTAT examination, complete the enclosed forms, and return the forms within one and a half months.

**Criterion-related Evidence**

The American Psychological Association, American Educational Research Association, & National Council on Measurement in Education stated that criterion-related validity evidence "demonstrates that test scores are systematically related to one or more outcome criteria" (APA, AERA, & NCME 1985). According to the Standards for Educational and Psychological Testing (1985), criterion-related evidence is sub-categorized into predictive validity and concurrent validity. Predictive validity is used to predict future performance, while concurrent validity is used when two tests supposedly measuring the same construct are correlated (Dick & Hagerty, 1971).

Concurrent validity was the type of criterion-related validity evidence obtained by correlating scores earned on the CSTAT examination and scores obtained on the written simulation section of the NATABOC certification examination.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Subjects

Senior-level subjects who took the written simulation portion of the NATABOC certification examination for the first time in April or June 2002 were used to collect criterion-related validity evidence. The same senior-level subjects that were recruited at the SEATA conference were used for criterion-related validity evidence.

Methods

Proficiency, efficiency, and omission scores on the CSTAT examination were correlated for each senior-level student to self-reported scores of the written simulation portion of the NATABOC certification examination. Subjects were instructed to self-report their scores on the written simulation portion of the NATABOC certification exam to the researcher. Subjects were reminded, via e-mail, 1 week before attempting the NATABOC certification examination to self-report scores. This e-mail also reminded subjects on how to submit scores electronically.

A web page was developed to allow subjects to submit scores electronically (Appendix C). This web page contained a form where subjects would fill-in their name and NATABOC scores. Upon submission, the researcher would receive these scores as an e-mail. Instructions on how to submit NATABOC written simulation scores to the researcher

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

via the Internet were given via the reminder e-mail and with the packet that contained the CSTAT CD-ROM. These scores were then correlated to the respective student's proficiency, efficiency, and omission scores on the CSTAT examination.

**Data Analysis**

Data were organized in groups based on level of school, i.e., certified, senior, and junior. A one-way analysis-of-variance (ANOVA) was used to determine mean differences between the three groups for construct-related validity evidence. A one-way ANOVA was used for each of the computer score variables, proficiency, efficiency and omission. A Tukey post-hoc test was used to determine which groups were significantly different. All statistical analyses were run at the .05 alpha-level.

For criterion-related evidence, all three computer simulation scores were correlated with the senior-level subjects' score on the written simulation portion of the NATABOC certification exam. A Pearson product moment correlation was used to determine the correlation between each pair of exam scores.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

# CHAPTER IV

# RESULTS

Data analyses were conducted to establish validity evidence for the CSTAT examination. The results described in this chapter are: 1) description of participants, 2) content-validity results, 3) construct-validity results, and 4) criterion-validity results.

## Description of Participants

The study consisted of two sets of participants. The first set of participants (n=12) were the subject-matter experts (SMEs) who provided expertise for the content validity portion of this research. The second set of participants (n=70) were those that earned scores on the CSTAT examination. A total of 82 subjects were recruited to establish validity evidence.

The SMEs were program directors from CAAHEP-accredited athletic training education programs. Of the twelve SMEs, 5 had obtained tenure at their institution, 4 were in tenure-track positions, and 3 were in non tenure-track positions. Five of the twelve SMEs had an earned doctoral degree, while 3 were seeking a doctoral degree, and the remaining 4 SMEs had obtained a master's degree. The mean years of teaching athletic training was 13.91, with a maximum of 21 and a minimum of 3.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The second set of participants were those that earned scores on the CSTAT examination. The data collected from these subjects were used for construct and criterion-related validity evidence. Data were collected on 70 subjects ages twenty through fourty-three. Of these 70 subjects, 23 were certified athletic trainers (ATCs), 31 were senior-level students, and 16 were junior- or sophomore-level students. A description of the subjects that earned scores on the CSTAT examination are listed in Table 3.

Table 3

*Descriptive data on subjects who attempted the CSTAT examination*

| Variable | Juniors (n=16) | Seniors (n=31) | ATCs (n=23) |
|---|---|---|---|
| Mean Age | 22.44 | 23.37 | 30.0 |
| Type of AT Program | | | |
| CAAHEP-Accredited | 7 | 18 | - |
| Candidacy | 6 | 10 | - |
| Internship | 3 | 3 | - |
| Year in AT Program | 2.06 | 3.33 | - |
| Years Experience | - | - | 7.47 |

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Each of the senior-level and junior-level subjects were attending one of three types of athletic training education programs. A candidacy-stage athletic training education program is a program that has initiated the accreditation process, but is not yet accredited. An internship program is an athletic training education program that is not currently seeking accreditation.

**Content-related Validity Evidence**

Sireci (1998) stated that using content experts to ascertain the domain representation and domain relevance of the test typically collects content validity. In addition to collecting this evidence, each SME completed a content validation form (Appendix A) for each of the three simulations randomnly given to the SME.

The content validity form had an 8-point Likert scale and contained the same questions utilized by the NATABOC for the content validation of the written simulation portion of the NATABOC certification exam (NATABOC, 2001). A score of 8 was the highest value and a score of 1 was the lowest value. Table 4 contains the mean scores for each simulation from the content validity form. The range of mean scores from the content validity form were from 6.33 to 7.10.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 4

*Mean content validity scores by simulation*

| Question | Sim1 (n=6) | Sim2 (n=5) | Sim3 (n=3) | Sim4 (n=4) | Sim5 (n=4) | Sim6 (n=5) | Sim7 (n=5) | Sim8 (n=3) |
|---|---|---|---|---|---|---|---|---|
| 1. Does the problem flow in a logical manner? | 6.50 | 6.00 | 7.33 | 6.25 | 5.75 | 6.40 | 7.20 | 7.33 |
| 2. Is the opening scene plausible? | 8.00 | 7.20 | 6.33 | 7.75 | 7.75 | 7.20 | 7.60 | 8.00 |
| 3. Is there sufficient information available to the candidate to solve the problem? | 5.33 | 2.80 | 5.33 | 5.25 | 5.75 | 5.20 | 5.40 | 2.00 |
| 4. Are the sections in proper order? | 6.33 | 7.00 | 8.00 | 7.25 | 6.75 | 7.75 | 7.40 | 7.33 |
| 5. Are there appropriate tests/actions/options to cover all of the incorrect actions? | 5.33 | 5.00 | 7.67 | 6.50 | 7.00 | 5.80 | 6.80 | 7.00 |
| 6. How essential is mastery of the knowledge, skills, and abilities in this scenario? | 8.00 | 7.40 | 7.67 | 7.25 | 7.75 | 7.00 | 7.40 | 8.00 |

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 7. How important are the knowledge, skills, and abilities assess by this scenario to the assessment of performance as an entry-level athletic trainer? | 8.00 | 7.60 | 7.33 | 6.75 | 7.75 | 7.20 | 7.60 | 8.00 |
| 8. How frequently would an entry-level athletic trainer be expected to encounter a case similar to that found in the scenario? | 5.83 | 7.60 | 5.33 | 6.25 | 4.00 | 4.60 | 7.40 | 7.00 |
| Total | 53.33 | 50.60 | 55.00 | 53.25 | 52.50 | 51.15 | 56.80 | 54.67 |
| Mean Validity Score | 6.67 | 6.33 | 6.88 | 6.66 | 6.56 | 6.39 | 7.10 | 6.83 |

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In addition to the content validation form, SMEs used the relevancy form (Appendix C) to determine the domain of best fit for each simulation. The relevancy form was a ten-point Likert scale, with 10 being the most relevant and 1 being the least relevant. Table 5 contains the mean relevancy scores and ranking for each simulation.

Table 5 describes how the simulations ranked compared to the most important and critical domains of athletic training. According to the NATABOC (NATABOC, 1999), the three most important and critical domains of athletic training are 1) immediate care, 2) recognition, evaluation, and assessement, and 3) treatment, rehabilitation and reconditioning. Six of the eight simulations had the highest rankings within these three most important and critical domains.

All eight simulation except #3 and #6 ranked highest in the three most important and critical domains of athletic training. While simulation #3 and #6 did not rank highest in the three most important and critical domains of athletic training, the second and third highest ranking for each domain was within these three important and critical domains.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 5

*NATA domain relevancy mean scores and ranking*

| NATA Domain | Sim1 | Sim2 | Sim3 | Sim4 | Sim5 | Sim6 | Sim7 | Sim8 |
|---|---|---|---|---|---|---|---|---|
| #1 | 3.60 (5.5) | 3.20 (4) | 4.33 (6) | 2.50 (4) | 2.50 (4) | 9.40 (1) | 3.80 (4) | 1.50 (4) |
| #2 | 8.80 (2) | 9.60 (1) | 8.33 (2.5) | 5.50 (3) | 9.75 (2) | 5.80 (3) | 5.80 (3) | 9.25 (1) |
| #3 | 9.40 (1) | 4.60 (2) | 8.33 (2.5) | 6.50 (2) | 10.00 (1) | 2.20 (6) | 9.20 (2) | 5.00 (2) |
| #4 | 4.20 (4) | 4.20 (3) | 5.67 (4.5) | 9.75 (1) | 2.00 (5) | 9.20 (2) | 9.60 (1) | 2.00 (3) |
| #5 | 5.80 (3) | 1.40 (6) | 9.00 (1) | 1.25 (5) | 4.50 (3) | 2.40 (5) | 1.00 (5.5) | 1.00 (5.5) |
| #6 | 3.60 (5.5) | 1.80 (5) | 5.67 (4.5) | 1.00 (6) | 1.00 (6) | 3.20 (4) | 1.00 (5.5) | 1.00 (5.5) |

Note: Domain #1 = Prevention; Domain #2 = Recognition, Evaluation, and Assessment; Domain #3 = Immediate Care; Domain #4 = Treatment, Rehabilitation, and Reconditioning; Domain #5 = Organization and Administration; Domain #6 = Professional Development and Responsibility.
()= Rank, a decimal value indicates a tie for two ranked values.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

## Construct-related Validity Evidence

An analysis of variance (ANOVA) determined significant group differences among the variables of overall proficiency, efficiency, and omission scores. Means scores and standard deviations are shown in Table 6. ATCs scored highest in proficiency and efficiency and lowest in omission, followed by seniors and then juniors for all three variables.

Table 6

*Mean and standard deviation of overall proficiency, efficiency, and omission scores by group*

| Variable | Junior | Senior | ATC |
|---|---|---|---|
| Proficiency | | | |
| Mean | 64.63 | 70.32 | 71.48 |
| Standard Deviation | 10.07 | 6.73 | 6.76 |
| Efficiency | | | |
| Mean | 78.69 | 81.74 | 83.61 |
| Standard Deviation | 4.60 | 4.99 | 4.34 |
| Omission | | | |
| Mean | 13.81 | 11.26 | 9.09 |
| Standard Deviation | 4.13 | 5.54 | 2.92 |

Results revealed a statistically significant $F(2,67) = 4.23$, $p=.019$ for proficiency. For efficiency, a

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

statistically significant F(2,67) = 5.19, p=.008 was found. For omission, a statistically significant F(2,67) = 5.19, p=.008 was also found. Table 7 contains the data from the ANOVA analysis.

Table 7

*ANOVA results for overall proficiency, efficiency, and omission scores*

| Variable | | Sum of Squares | df | F | Sig. |
|---|---|---|---|---|---|
| Proficiency | | | | | |
| | Between Groups | 490.54 | 2 | 4.23 | .019 |
| | Within Groups | 3886.26 | 67 | | |
| | Total | 4376.80 | 69 | | |
| Efficiency | | | | | |
| | Between Groups | 228.92 | 2 | 5.19 | .008 |
| | Within Groups | 1476.85 | 67 | | |
| | Total | 1705.77 | 69 | | |
| Omission | | | | | |
| | Between Groups | 211.64 | 2 | 5.19 | .008 |
| | Within Groups | 1366.20 | 67 | | |
| | Total | 1577.84 | 69 | | |

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

For overall proficiency, Tukey post-hoc testing revealed only significant differences between the ATC group and the junior-level group (p=.020) and the senior-level group and the junior-level group (p=.046). For efficiency and omission, only the junior-level and ATC group (p=.006) were significantly different. Table 8 contains the data for the Tukey post-hoc test. All significant differences in the Tukey post-hoc test were significant at the .05 alpha-level, except for efficiency, which was significant at the .01 alpha-level. Of interest in Table 8 is that ATCs and senior-level subjects were not significantly different among proficiency, efficiency, and omission. Conversely, ATCs and junior-level subjects had the highest mean difference in scores for proficiency, efficiency, and omission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 8

*Results of Tukey post-hoc test for overall proficiency, efficiency, and omission scores*

| Dependent Variable | (I) STATUS | (J) STATUS | Mean Difference (I-J) | Std. Error | Sig. |
|---|---|---|---|---|---|
| Proficiency | | | | | |
| | Junior | Senior | -5.70* | 2.34 | .046 |
| | | ATC | -6.85* | 2.48 | .020 |
| | Senior | Junior | 5.70* | 2.34 | .046 |
| | | ATC | -1.16 | 2.10 | .846 |
| | ATC | Junior | 6.85* | 2.48 | .020 |
| | | Senior | 1.16 | 2.10 | .846 |
| Efficiency | | | | | |
| | Junior | Senior | -3.05 | 1.45 | .095 |
| | | ATC | -4.92** | 1.53 | .006 |
| | Senior | Junior | 3.05 | 1.45 | .095 |
| | | ATC | -1.87 | 1.29 | .324 |
| | ATC | Junior | 4.92* | 1.53 | .006 |
| | | Senior | 1.87 | 1.29 | .324 |
| Omission | | | | | |
| | Junior | Senior | 2.55 | 1.39 | .165 |
| | | ATC | 4.73* | 1.47 | .006 |
| | Senior | Junior | -2.55 | 1.39 | .165 |
| | | ATC | 2.17 | 1.24 | .196 |
| | ATC | Junior | -4.73* | 1.47 | .006 |
| | | Senior | -2.17 | 1.24 | .196 |

* $p < .05$, ** $p < .01$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

In addition to the overall group differences, significant group differences were ascertained for each simulation for proficiency, efficiency, and omission scores. Significant differences between groups ($p<.05$) were found for simulation #2, #3, #4, #6, #7, and #8. No statistically significant differences ($p>.05$) were found between groups for simulation #1 and #5.

For simulation #2, ANOVA results revealed significant group differences for efficiency ($p=.001$) and omission ($p=.001$) only. For simulation #3, significant group differences were found for proficiency ($p=.018$), efficiency ($p=.005$), and omission ($p=.001$). In simulation #4, significant group differences were found for proficiency ($p=.029$) and omission ($p=.018$) only. For simulation #6, significant group differences were found for proficiency ($p=.001$) and omission ($p=.007$) only. In simulation #7, significant group differences were found for proficiency ($p=.027$), efficiency ($p=.025$), and omission ($p=.050$). For simulation #8, significant group differences were found for proficiency ($p=.036$) only.

Tables 9-16 contain ANOVA results for all eight simulations. In addition, Table 17 denotes the groups that were significantly different according to Tukey's HSD post-hoc test.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 9

*ANOVA results for proficiency, efficiency, and omission scores for simulation #1*

| Variable | | Sum of Squares | df | F | Sig. |
|---|---|---|---|---|---|
| Sim1-Prof | | | | | |
| | Between Groups | 452.87 | 2 | .874 | .422 |
| | Within Groups | 17362.63 | 67 | | |
| | Total | 17815.50 | 69 | | |
| Sim1-Eff | | | | | |
| | Between Groups | 72.56 | 2 | .530 | .591 |
| | Within Groups | 4586.93 | 67 | | |
| | Total | 4659.49 | 69 | | |
| Sim1-Omiss | | | | | |
| | Between Groups | 44.26 | 2 | .649 | .526 |
| | Within Groups | 2282.89 | 67 | | |
| | Total | 2327.15 | 69 | | |

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 10

*ANOVA results for proficiency, efficiency, and omission scores for simulation #2*

| Variable | | Sum of Squares | df | F | Sig. |
|---|---|---|---|---|---|
| Sim2-Prof | | | | | |
| | Between Groups | 503.15 | 2 | 2.49 | .091 |
| | Within Groups | 6783.84 | 67 | | |
| | Total | 7286.99 | 69 | | |
| Sim2-Eff | | | | | |
| | Between Groups | 1012.61 | 2 | 8.76 | .000 |
| | Within Groups | 3874.59 | 67 | | |
| | Total | 4887.20 | 69 | | |
| Sim2-Omiss | | | | | |
| | Between Groups | 821.39 | 2 | 8.81 | .000 |
| | Within Groups | 3124.45 | 67 | | |
| | Total | 3945.84 | 69 | | |

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 11

*ANOVA results of proficiency, efficiency, and omission scores for simulation #3*

| Variable | | Sum of Squares | df | F | Sig. |
|---|---|---|---|---|---|
| Sim3-Prof | | | | | |
| | Between Groups | 821.58 | 2 | 4.27 | .018 |
| | Within Groups | 6450.37 | 67 | | |
| | Total | 7271.94 | 69 | | |
| Sim3-Eff | | | | | |
| | Between Groups | 533.89 | 2 | 5.80 | .005 |
| | Within Groups | 3083.97 | 67 | | |
| | Total | 3617.84 | 69 | | |
| Sim3-Omiss | | | | | |
| | Between Groups | 704.82 | 2 | 9.60 | .000 |
| | Within Groups | 2459.48 | 67 | | |
| | Total | 3164.30 | 69 | | |

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 12

*ANOVA results of proficiency, efficiency, and omission scores for simulation #4*

| Variable | | Sum of Squares | df | F | Sig. |
|---|---|---|---|---|---|
| Sim4-Prof | | | | | |
| | Between Groups | 941.96 | 2 | 3.75 | .029 |
| | Within Groups | 8413.46 | 67 | | |
| | Total | 9355.44 | 69 | | |
| Sim4-Eff | | | | | |
| | Between Groups | 149.66 | 2 | 1.50 | .231 |
| | Within Groups | 3352.11 | 67 | | |
| | Total | 3501.77 | 69 | | |
| Sim4-Omiss | | | | | |
| | Between Groups | 375.94 | 2 | 4.27 | .018 |
| | Within Groups | 2946.35 | 67 | | |
| | Total | 3322.29 | 69 | | |

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 13

*ANOVA results of proficiency, efficiency, and omission scores for simulation #5*

| Variable | | Sum of Squares | df | F | Sig. |
|---|---|---|---|---|---|
| Sim5-Prof | | | | | |
| | Between Groups | 170.43 | 2 | .561 | .574 |
| | Within Groups | 10185.57 | 67 | | |
| | Total | 10356.00 | 69 | | |
| Sim5-Eff | | | | | |
| | Between Groups | 248.21 | 2 | 2.33 | .105 |
| | Within Groups | 3568.37 | 67 | | |
| | Total | 3816.58 | 69 | | |
| Sim5-Omiss | | | | | |
| | Between Groups | 199.28 | 2 | 1.95 | .151 |
| | Within Groups | 3431.06 | 67 | | |
| | Total | 3630.34 | 69 | | |

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 14

*ANOVA results of proficiency, efficiency, and omission scores for simulation #6*

| Variable | | Sum of Squares | df | F | Sig. |
|---|---|---|---|---|---|
| Sim6-Prof | | | | | |
| | Between Groups | 2682.91 | 2 | 11.97 | .000 |
| | Within Groups | 7509.68 | 67 | | |
| | Total | 10192.59 | 69 | | |
| Sim6-Eff | | | | | |
| | Between Groups | 281.28 | 2 | 2.60 | .082 |
| | Within Groups | 3629.71 | 67 | | |
| | Total | 3910.99 | 69 | | |
| Sim6-Omiss | | | | | |
| | Between Groups | 473.26 | 2 | 5.28 | .007 |
| | Within Groups | 3004.01 | 67 | | |
| | Total | 3477.27 | 69 | | |

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 15

*ANOVA results of proficiency, efficiency, and omission scores for simulation #7*

| Variable | | Sum of Squares | df | F | Sig. |
|---|---|---|---|---|---|
| Sim7-Prof | | | | | |
| | Between Groups | 1529.23 | 2 | 3.80 | .027 |
| | Within Groups | 13476.86 | 67 | | |
| | Total | 14997.09 | 69 | | |
| Sim7-Eff | | | | | |
| | Between Groups | 1006.82 | 2 | 3.68 | .025 |
| | Within Groups | 8690.96 | 67 | | |
| | Total | 9697.79 | 69 | | |
| Sim7-Omiss | | | | | |
| | Between Groups | 647.98 | 2 | 3.11 | .050 |
| | Within Groups | 6975.16 | 67 | | |
| | Total | 7623.14 | 69 | | |

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 16

*ANOVA results of proficiency, efficiency, and omission scores for simulation #8*

| Variable | | Sum of Squares | df | F | Sig. |
|---|---|---|---|---|---|
| Sim8-Prof | | | | | |
| | Between Groups | 464.47 | 2 | 3.50 | .036 |
| | Within Groups | 4382.75 | 67 | | |
| | Total | 4847.22 | 69 | | |
| Sim8-Eff | | | | | |
| | Between Groups | 544.68 | 2 | 2.13 | .127 |
| | Within Groups | 8434.19 | 67 | | |
| | Total | 8978.87 | 69 | | |
| Sim8-Omiss | | | | | |
| | Between Groups | 73.46 | 2 | 1.58 | .214 |
| | Within Groups | 1536.83 | 67 | | |
| | Total | 1610.29 | 69 | | |

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 17

*Significantly different groups by simulation*

| Sim | Proficiency Scores | Efficiency Scores | Omission Scores |
|---|---|---|---|
| #1 | — | — | — |
| #2 | — | Juniors-Seniors (.002) | Juniors-Seniors (.001) |
| | | Juniors-ATCs (.001) | Juniors-ATCs (.002) |
| #3 | Juniors-ATCs (.015) | Juniors-ATCs (.004) | Juniors-ATCs (.000) |
| | | | Seniors-ATCs (.033) |
| #4 | Seniors-ATCs (.046) | — | Juniors-ATCs (.025) |
| #5 | — | — | — |
| #6 | Juniors-Seniors (.000) | — | Juniors-ATCs (.005) |
| | Juniors-ATCs (.000) | | |
| #7 | Juniors-Seniors (.020) | Juniors-Seniors (.030) | Juniors-Seniors (.045) |
| | | Juniors-ATCs (.050) | |
| #8 | Juniors-Seniors (.046) | — | — |

Note. () = level of significance

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

## Criterion-related Validity Evidence

Ten subjects reported their NATABOC written simulation examination results. The Pearson product moment correlation coefficient determined the correlation between the senior subjects' written simulation score on the NATABOC certification examination and their overall proficiency, efficiency, and omission scores. Results revealed a correlation between proficiency and written simulation score of .231 (p=.522). The correlation between efficiency and written simulation score was .427 (p=.218). The correlation between omission and written simulation score was -.438 (p=.206).

Table 18 contains the raw scores for subjects that reported their NATABOC certification exam results. A passing score on the written simulation portion of the NATABOC is a score of 500 or higher. Of interest is subject #5, this subject scored well on the CSTAT examination but did not pass the written simulation portion of the NATABOC certification exam.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 18

*Raw scores for seniors reporting NATABOC certification results*

| Subject | Proficiency Score | Efficiency Score | Omission Score | Written Sim Score |
|---|---|---|---|---|
| 1 | 73 | 87 | 7 | 683 |
| 2 | 78 | 86 | 7 | 645 |
| 3 | 60 | 75 | 17 | 461 |
| 4 | 75 | 80 | 10 | 558 |
| 5 | 73 | 88 | 6 | 423 |
| 6 | 69 | 89 | 6 | 649 |
| 7 | 75 | 80 | 10 | 535 |
| 8 | 70 | 88 | 5 | 685 |
| 9 | 73 | 79 | 12 | 622 |
| 10 | 60 | 80 | 13 | 589 |

Table 19 contains the data for the Pearson Product Moment Correlation between overall proficiency, efficiency, omission, and written simulation scores. The only significant finding ($p<.001$) was the correlation between omission and efficiency scores of -.953. This significant negative correlation is expected since efficiency scores are the percentage of correct choices made and omission scores are the percentage of incorrect choices made in the CSTAT examination.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 19

*Correlation coefficients between proficiency, efficiency, omission, and written simulation score*

| | WS | Prof | Effic | Omiss |
|---|---|---|---|---|
| WS | | | | |
| Correlation | — | .231 | .427 | -.438 |
| Significance | | .522 | .218 | .206 |
| Prof | | | | |
| Correlation | .231 | — | .412 | -.619 |
| Significance | .522 | | .237 | .057 |
| Effic | | | | |
| Correlation | .427 | .412 | — | -.953 |
| Significance | .218 | .237 | | .000 |
| Omiss | | | | |
| Correlation | -.438 | -.619 | -.953 | — |
| Significance | .206 | .057 | .000 | |

Note: WS = written simulation; Prof = Proficiency; Effic = Efficiency; Omiss = Omission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 1 contains a scatterplot and line of best fit for the raw scores from the ten senior subjects. Of note, is the outlier on the far left of figure 1.

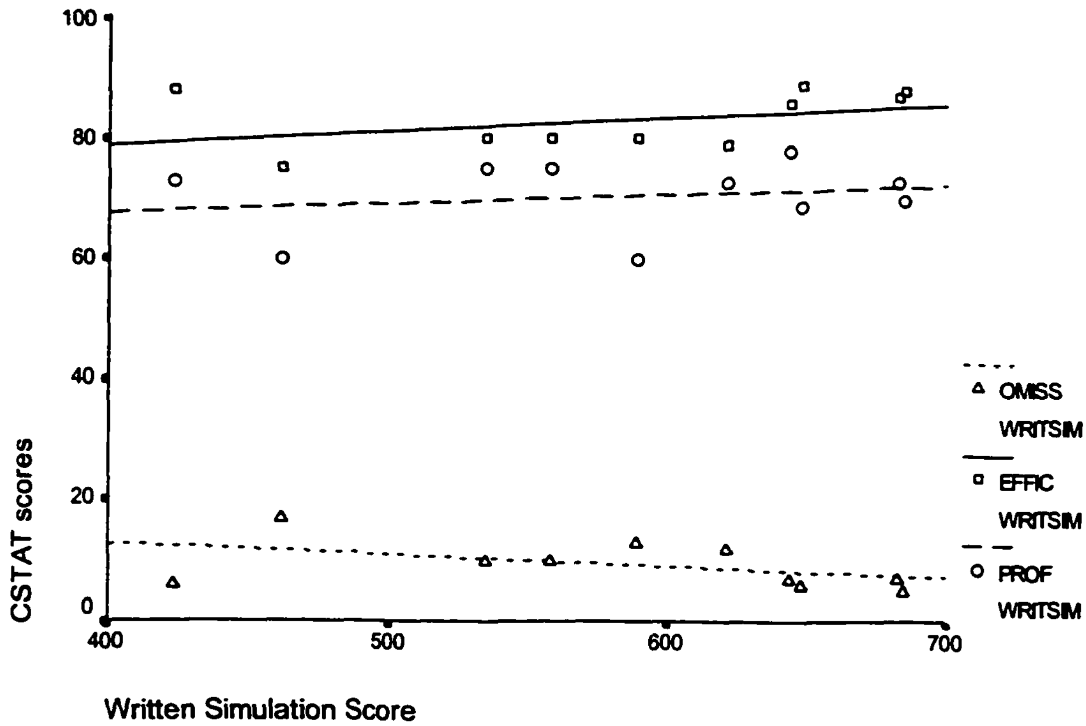



Figure 1. Scatterplot of proficiency, efficiency, omission, and NATABOC written-simulation scores.

This outlier subject scored well on the computerized simulation exam but did not score well on the NATABOC certification examination. Another Pearson product moment correlation was performed without this outlier data. Table 20 contains the correlation results without the outlier data.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 20

*Correlation coefficients between proficiency, efficiency, omission, and written simulation score without outlier data*

| | WS | Prof | Effic | Omiss |
|---|---|---|---|---|
| Written Simulation | | | | |
| Correlation | — | .412 | .882 | -.849 |
| Significance | | .270 | .002 | .004 |
| Proficiency | | | | |
| Correlation | .412 | — | .392 | -.611 |
| Significance | .270 | | .297 | .080 |
| Efficiency | | | | |
| Correlation | .882 | .392 | — | -.949 |
| Significance | .002 | .297 | | .000 |
| Omission | | | | |
| Correlation | -.849 | -.611 | -.949 | — |
| Significance | .004 | .080 | .000 | |

Note: WS = written simulation; Prof = Proficiency; Effic = Efficiency; Omiss = Omission.

Without the outlier data, significant correlations were found between efficiency and the written simulation exam ($p=.002$), and omission and the written simulation exam ($p=.004$).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Figure 2 contains the scatterplot and line of best fit for the correlational analysis without the outlier data. Without the outlier data, the line of best fit has a greater positive slope for proficiency and efficiency scores and a greater negative slope for omission scores than those in figure 1.

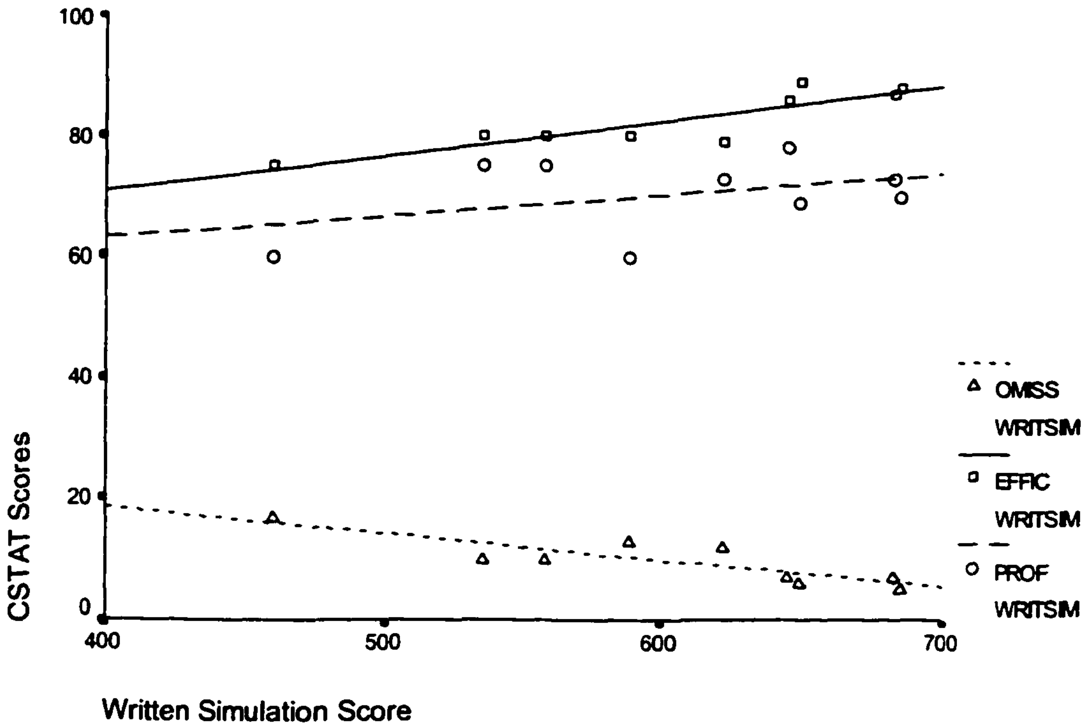



Figure 2. Scatterplot of proficiency, efficiency, omission, and NATABOC written-simulation scores without outlier data.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

# CHAPTER V

## SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

The purpose of the study was to collect evidence to validate a computer simulation program to measure decision-making skills in athletic training. The following research questions were investigated:

1. Does the CSTAT examination have a high level of content-related validity?
2. Do scores on the computer simulation program differ between known groups of users? (Construct-related validity evidence)
3. What is the correlation of senior-level subjects' scores on the computer simulation program and those same subjects' scores on the written simulation portion of the NATABOC certification examination? (Criterion-related validity evidence)

This chapter addresses the following areas: 1) summary, 2) conclusions, 3) discussion, and 4) recommendations.

### Summary

#### Content Validity

The twelve SMEs were program directors of CAAHEP-accredited athletic training education programs. These SMEs completed four forms for content validity - a content validation form, an NATA domain relevancy form, a

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

descriptive questionnaire, and a score form for three of the eight simulations. Yalow and Popham (1983) stated that there is no uniformity for content reviewers or in the method the resulting data has been analyzed, so recommendations from other authors were used to collect content-related validity evidence.

Content-related validity evidence collection followed the guidelines used for the NATABOC certification examination (NATABOC, 2002b). The item scoring for the CSTAT examination used the same 5-point scale as the NATABOC. The NATABOC used a separate panel of experts to validate the written simulation examination, similar to the 12 SMEs used in this study. In addition, the recommendations Sireci (1998) gave regarding SMEs were used as a guideline for this study. Sireci recommended using competent and representative SMEs, to use 15 or fewer SMEs, to minimize the burden on SMEs, to use rating scales of sufficient length, and to provide monetary compensation for SMEs. All of these recommendations were followed. Each SME received a completed copy of the CSTAT CD-ROM for his/her compensation.

Construct Validity

The CSTAT examination contained eight simulations and the three variables collected from the CSTAT examination

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

were proficiency, efficiency, and omission. The proficiency score is the sum of the points assigned to all selected options divided by the maximum score possible. The efficiency score is the proportion of an examinee's selections that are indicated options. The omission score is the proportion of contraindicated options that were selected (Swanson, 1987). These three variables were collected for each subjects' overall score from all eight simulations combined, and the subjects' score for each individual simulation.

One method of collecting construct-related evidence is to compare group differences on a test (Payne, 1997; Melnick, 1990). The groups in this study were ATCs (n=23), senior-level undergraduate students (n=31), and junior- or sophomore-level undergraduate students (n=16). Results of the ANOVA found significant group differences ($p<.05$) for overall proficiency, efficiency, and omission scores. Post-hoc testing revealed that for overall proficiency, there were only significant differences ($p<.05$) between the ATC group and the junior-level group, and the senior-level group and the junior-level group. For overall proficiency, there was no significant difference between ATCs and senior-level students. For overall efficiency and omission, only the junior-level and ATC group were significantly

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

different. ANOVA results for each simulation revealed significant group differences for all simulations except simulation #1 and #5 for proficiency, efficiency, and omission scores.

Criterion Validity

Ten senior-level subjects reported their NATABOC certification exam results. A Pearson product moment correlation determined the correlation between proficiency, efficiency, omission, and written simulation score. Results of the initial correlational analysis revealed low coefficients of .381 (proficiency-written simulation), .462 (efficiency-written simulation), and -.467 (omission-written simulation).

Upon further evaluation, it was noted that one data set was an outlier, and a Pearson correlation was conducted a second time without this outlier datum. Results of this correlation analysis without the outlier data revealed a correlational coefficient of .561 (proficiency-written simulation), .965 (efficiency-written simulation), and -.929 (omission-written simulation). Both correlations, between efficiency and NATABOC written simulation and omission and NATABOC written simulation, were significant at the .01 alpha-level.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

## Conclusions

Research question #1:

Does the CSTAT examination have a high level of content-related validity?

The results of the content validity form returned by the SMEs showed that all simulations had at least a mean validity score of 6.33 out of a maximum score of eight. The range of mean validity scores was from 6.33 to 7.10. Based on these results, and the methods of accumulating content validity data, the CSTAT contained a high level of content validity in measuring decision-making skills in athletic training.

Research question #2:

Do scores on the computer simulation program differ between known groups of users?

Due to the significant group differences between ATCs, senior-level subjects, and junior-level subjects, there is evidence of construct-related validity. However, ATCs and senior-level subjects were not significantly different in the overall scores of the CSTAT examination.

One weakness of this construct-related validity evidence is a relatively low number of subjects (n=70). In order to determine if the CSTAT examination is a true test

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

of decision-making skills in athletic training, more subjects are needed in future studies.

Research question #3:

What is the correlation of senior-level subjects' scores on the computer simulation program and those same subjects' scores on the written simulation portion of the NATABOC certification examination?

Efficiency and omission scores from CSTAT had high correlations ($p<.01$) with the written simulation portion of the NATABOC certification examination. Based on these high correlations, there is criterion-related validity evidence that efficiency and omission scores of the CSTAT are measures of decision-making skills in athletic training. In addition, the efficiency and omission scores are predictors of success on the written simulation portion of the NATABOC certification examination.

## Discussion

The only other study investigating computerized simulations in athletic training was conducted by Castle. Castle (2000) used 29 subjects and found a -.052 correlation between NATABOC written simulation score and his computerized simulation test. Castle's CTATSI (Computerized-Traditional Athletic Training Simulation Instrument) also contained eight simulations but did not

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

contain digital video or audio. Castle did not investigate group differences on the CTATSI, but rather investigated the effectiveness of CTATSI as a predictor of success on the written simulation portion of the NATABOC certification examination. Castle concluded that the CTATSI was not a significant predictor of success on the written simulation portion of the NATABOC certification examination.

There were very little measureable differences between senior-level subjects and ATCs in this study. This result is similar to results reported by several authors. Devitt, Kurrek, Cohen, and Cleave-Hogg (2001) found no significant difference between practicing anesthesiologists and final year resident students on a computer simulation program. Gates, Greene, and Kris-Etherton (1990) found little differences between dietitians and interns on a patient management problem. Both the dietitians and interns scored significantly higher than students ($p<.001$), but there were no significant differences between dietitians and interns. It has also been stated that "practicing physicians generally perform more poorly on knowledge-oriented test, reflecting both changes in medicine since training and natural decay of non-essential knowledge" (Swanson et al., 1987). This natural decay of non-essential knowledge could

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

be a possible reason for ATCs not earning statistically higher CSTAT scores than senior-level subjects.

There are several other possible explanations for a lack of significant differences between seniors and ATCs on the CSTAT examination. The main possibility was that eight of the ten senior-level subjects passed the written simulation portion of the NATABOC certification examination. Therefore, according to the NATABOC, those eight seniors have an acceptable level of decision-making skills to become a certified athletic trainer. A different possibility was that senior-level subjects were recruited at a conference designed to prepare them for the NATABOC certification examination. These subjects were actively engaged in studying for the NATABOC certification examination.

Another possible reason for the lack of significant difference between seniors and ATCs, was that some standard procedures have changed since the ATCs were in school. For example, the guidelines for CPR and removing an airway obstruction have changed in the last two years. In addition, standardized guidelines for spine boarding an athlete were created by the NATA and the Inter-Association Task Force for Appropriate Care of the Spine-Injured Athlete in 2001 (NATA, 2002). Senior-level subjects are

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

more likely to be up-to-date on this information, since they are preparing for the NATABOC certification examination.

Both the CSTAT and the NATABOC's written simulation examination measure decision-making skills in athletic training. The difference in these two examinations is that in the CSTAT examination, a user is unable to go to a previous section of the examination. In the written simulation, examinees may read ahead or go back and reveal answers based on what was found in later sections. One benefit of the CSTAT examination is that it enables users to see video and audio of a patient. This increased fidelity of the examination can enable users to practice more real-life situations without the fear of hurting a patient.

In addition, the CSTAT exam enables users to practice situations that are infrequent in their clinical rotation. A skill not frequently seen but of extreme importance is spine boarding a patient. The CSTAT examination evaluates the decision-making ability in this situation. This would enable the user to assess his/her current proficiency in the skill and identify weaknesses in the decision-making process of the skill. Athletic training students could use the CSTAT examination as a diagnostic evaluation of their

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

own decision-making skills in preparation for the NATABOC certification examination. Once the CSTAT examination identifies a student's strengths and weaknesses, the student can be more focused on areas of weakness in his/her preparation for the NATABOC certification examination.

An educational benefit of a computerized simulation examination would be in the preparation of students in athletic training. Teachers could incorporate computerized simulations into the athletic training curriculum as one method of assessing a student's decision-making ability. Fuller's (1997) study of critical thinking in athletic training found that only 14% of questions on a written examination were classified as critical thinking questions. Fuller noted that simulations would foster critical thinking skills and that none of the thirteen athletic training program directors had a simulation question on their exams. A computerized simulation examination would be a tool to use by educational faculty to measure decision-making skills.

Another educational benefit of the CSTAT examination would be as a teaching and assessment tool for continuing education for certified athletic trainers. As new guidelines and protocols are developed in the future for patient care, a computerized simulation could be created to

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

teach and then assess certified athletic trainers' knowledge in the subject-matter.

Due to these benefits of the CSTAT examination, one possible future use of the CSTAT is to augment or replace the written simulation portion of the NATABOC examination. The major disadvantage to replacing the written simulation portion of the NATABOC certification examination with a computerized simulation test is the cost. Each examinee would be required to use a personal computer with a CD-ROM drive and significant amounts of RAM to run the CSTAT program. In addition, there would need to be technological support in case of a problem with one of the computers during testing. All of these factors would increase the cost in administering the NATABOC certification examination.

It may not be beneficial to replace the NATABOC certification exam with the CSTAT examination at this time due to this cost increase. As computer costs decline, the incorporation of a computerized simulation could then be appropriate to augment or replace the written simulation portion of the NATABOC certification examination.

In a future version of CSTAT, rather than having eight simulations, it could have 4-5 simulations for each domain of athletic training. These simulations could be randomly

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

selected and this would allow the student preparing for the NATABOC certification examination to participate in up to 20-25 different simulations. A comprehensive CSTAT examination like that could reveal further weaknesses of the student information on decision-making skills. This information would not only help students prepare for the NATABOC certification exam, but would ideally help them become better athletic training professionals.

Future studies using the CSTAT examination should look at attitudes towards computers as a factor for success. Castle (2000) investigated attitudes towards computers on his CTATSI, and found significant correlations between attitudes towards computers and scores on the CTATSI. Another possibility is to investigate whether there are group differences on the CSTAT for subjects with different learning styles. It may be that students who are visual learners or individual learners may perform better on a computerized simulation examination. There are no studies that have investigated the effect of learning styles on computerized simulation examinations. Future studies should investigate this possible correlation.

Overall, the efficiency and omission scores on the CSTAT examination appear to measure decision-making skills in athletic training. However, more subjects are needed

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

for future research to establish the true measure of decision-making skills for the CSTAT examination.

**Recommendations**

As a result of this investigation, the following recommendations are given for future studies:

1. Increase the number of subjects submitting NATABOC certification examination results.
2. Obtain a more geographically diverse sample of subjects throughout the United States.
3. Obtain ATC subjects who are employed in a variety of settings.
4. Correlate scores on a computerized simulation examination to individual learning styles.
5. Determine if attitudes towards computers have an effect on computer simulation exam results.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

# Appendix A

## Content Validity Form

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

# Computer Simulation: Content Validation Form

**Scenario # ____:**

| Statements | Rating Scale (8 is highest value; 1 is lowest value) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| The problem flows in a logical manner | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| The opening scene is plausible | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| There is sufficient information available to the candidate to solve the problem | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| The sections are in proper order | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| There are appropriate tests, actions, etc. to cover all of the incorrect actions | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| How essential is mastery of the knowledge, skills, and abilities in this scenario? | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| How important are the knowledge, skills, and abilities assessed by this scenario to the assessment of performance as an entry-level athletic trainer? | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| How frequently would an entry-level athletic trainer be expected to encounter a case similar to that found in the scenario? | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |

How could the scenario reviewed be improved?

Total Score ________

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

# Appendix B

## Scenario Relevance to NATA Performance Domains

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.



**Scenario Relevance to NATA Performance Domains**

Please use the following scale to rate the relevance of the scenarios to the NATA Performance Domains.
**Provide six (6) relevance ratings for each scenario.**

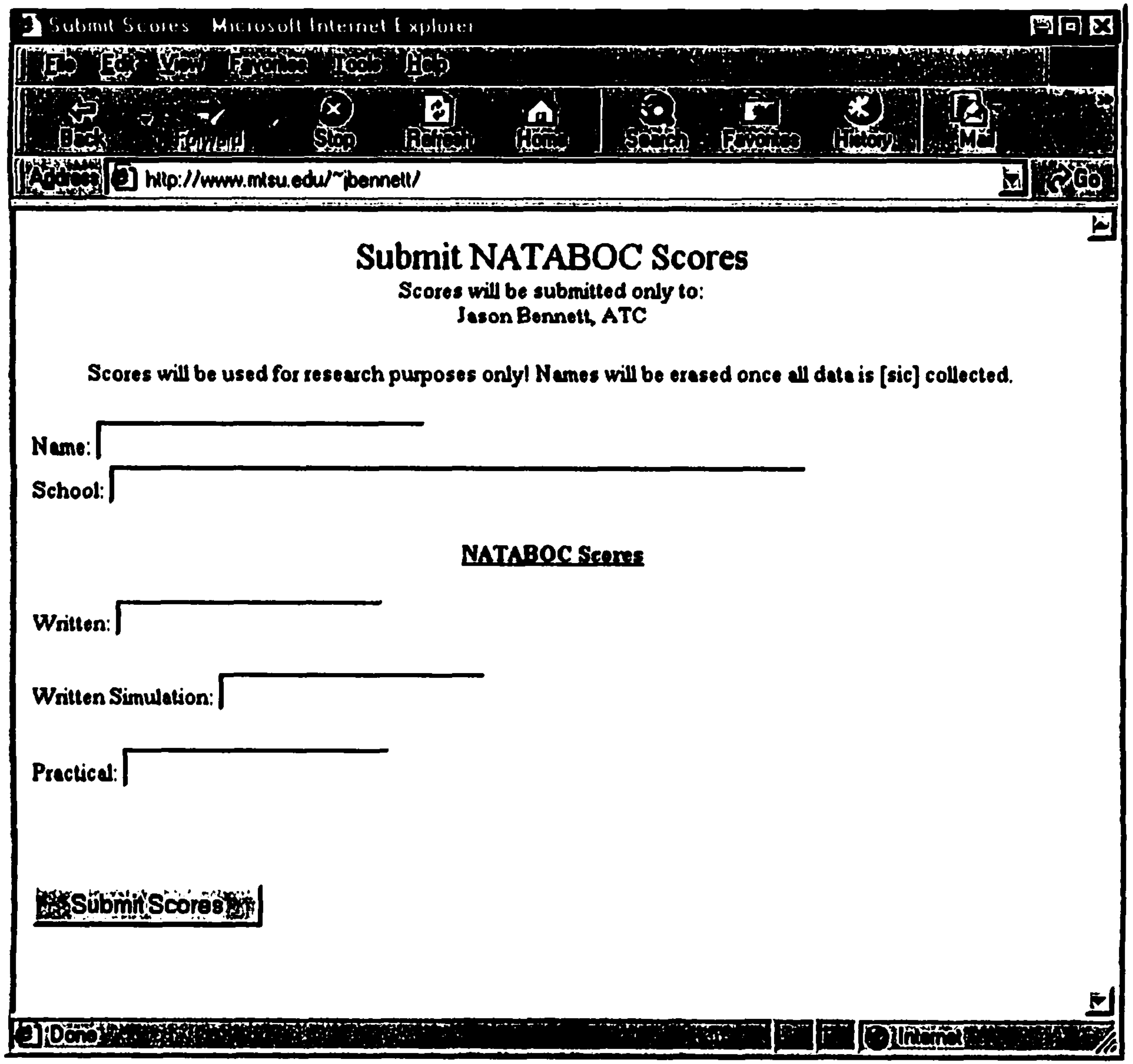
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| **Not at all Relevant** | | | | | | | | | **Highly Relevant** |

| | Prevention | Recognition, Evaluation and Assessment | Immediate Care | Treatment, Rehabilitation and Reconditioning | Organization and Administration | Professional Development and Responsibility |
|---|---|---|---|---|---|---|
| Scenario #1 | | | | | | |
| Scenario #2 | | | | | | |
| Scenario #3 | | | | | | |
| Scenario #4 | | | | | | |
| Scenario #5 | | | | | | |
| Scenario #6 | | | | | | |
| Scenario #7 | | | | | | |
| Scenario #8 | | | | | | |

# Appendix C

## Web Page for Submission of NATABOC Certification Examination Scores

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.




Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

# Appendix D

## IRB Approval

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Elementary and Special Education Department**
P.O. Box 69
Middle Tennessee State University
Murfreesboro, Tennessee 37132
(615) 898-2680

To: Jason Bennett

From: Nancy Bertrand, Chair *Nancy Bertrand*
MTSU Institutional Review Board

Re: "Development and Validation of a Computer Simulation Exam to Measure Decision-making Skills in Athletic Training"
Protocol #01-188

Date: April 3, 2001

The above named human subjects research proposal has been reviewed and approved. This approval is for one year only. Should the project extend beyond one year or should you desire to change the research protocol in any way, you must submit a memo describing the proposed changes or reasons for extensions to your college's IRB representative for review.

Best of luck in the successful completion of your research.

cc: Dr. Bill Whitehill



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

# Appendix E

## Letter of Consent for SEATA Conference Attendees

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

## Informed Consent Form

Title: Development and Validation of a Computer Simulation Exam to Measure Decision-making Skills in Athletic Training

Lead Researcher: Jason Bennett, ATC

Potential Subjects for Research,

My name is Jason Bennett and I am currently working on my doctorate at Middle Tennessee State University. I am collecting data on the validation of a computerized simulation program in athletic training. I am asking for volunteers to participate in this study. In this study you will receive a Computerized Simulation program that contains eight different simulations. You will be asked to submit the scores you received on the computer simulation program to the researcher. For those subjects that attempt the NATABOC certification exam for the first time in April or June, 2001 or 2002, you will be asked to submit your scores from the certification exam to the researcher. Participation in this study will not involve any financial cost to you.

Please note the following regarding your involvement in this study:

- Participation in this study is voluntary
- Your names and scores will kept in the strictest confidence by the researcher
- You may decide to withdraw from this study at any time

I, ______________________ (print your name), give consent to participate in the aforementioned study.

______________________________ ____________

Signature Date

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

REFERENCES

Alessi, S. (1988). Fidelity in the design of instructional simulations. *Journal of Computer-Based Instruction, 15*(2), 40-47.

Alessi, S. & Johnson, L. (1992). Simulations for dental licensure examinations: reliability and validity. *Simulation and Games for Learning, 22*(4), 286-307.

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (1985). Standards for educational and psychological testing. Washington D.C., American Psychological Association.

Barker, S. (1998). Computerized Athletic Training Simulations [computer program]. Developed for Southeast Athletic Trainers' Association's 1998 Student Meeting (Competencies in Athletic Training).

Barrett, R. (1992). Content validation form. *Public Personnel Management, 21*(1), 41-52.

Baumgartner, T & Jackson, A. (1999). Measurement for evaluation in physical education and exercise science ($6^{th}$ ed.). New York: McGraw-Hill.

Bersky, A. (1996). Computerized clinical simulation testing (CST): Frequently asked questions. *Issues, 17*(3), 6-7.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Bersky, A. & Yocom, C. (1994). Computerized clinical simulation testing: its use for competence assessment in nursing. *Nursing & Health Care, 15*(3), 120-127.

Castle, R. R. (2000). The effectiveness of a computer program and other factors in predicting students' success on the written simulation section of the N.A.T.A.B.O.C., Inc. Certification Examination. Unpublished doctoral dissertation, University of Southern Mississippi, Hattiesburg.

Clauser, B., Subhiyah, R., Piemme, T., Greenberg, L., Clyman, S., Ripkey, D., Nungester, R. (1993). Using clinician ratings to model score weights for a computer-based clinical-simulation examination. Academic Medicine, 68(10), S64-S65.

Crocker, L. (1997). Assessing content representativeness of performance assessment exercises. *Applied Measurement in Education, 10*(1), 83-95.

Croke, E. M. (1999). Role of reflective thinking in the development of clinical decision making in first-semester nursing students. Unpublished doctoral dissertation, Pepperdine University, Malibu.

Curry, B. & Moutinho, L. (1992). Using computer simulations in management education. *Management Education and Development, 23*(2), 155-167.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Devitt, J., Kurrek, M., Cohen, M., Cleave-Hogg, D. (2001). The validity of performance assessments using simulation. *Anesthesiology, 95* (1), 36-42.

Dick, W. & Hagerty, N. (1971). Topics in measurement; reliability and validity. New York: McGraw-Hill.

Education Council Web Site. (2002). Athletic Training Educational Competencies for the Health Care of the Physically Active. http://www.cewl.com/competencies/eddom2.html

Engberg, S. & White, J. (1991). Using interactive video simulations to teach and evaluate clinical decision-making. *Nuring Education Microworld, 5*(5), 36.

Forker, J. & McDonald, M. (1996). Methodologic trends in the healthcare professions: computer adaptive and computer simulation testing. *Nurse Educator, 21*(4), 13-14.

Friedman, C. (1995). Anatomy of the clinical simulation. *Academic Medicine, 70*(3), 205-09.

Fuller, D. (1997). Critical thinking in undergraduate athletic training education. *Journal of Athletic Training, 32*(3), 242-247.

Gagne, R. (1954). Training devices and simulators: some research issues. *American Psychologist, 9*, 95-107

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Gates, G., Greene, G., & Kris-Etherton, P. (1990). Nutrition care planning: Comparison of the skills of dietitians, interns, and students. *Journal of the American Dietetic Association, 90*(10), 1393-1397.

Gill, D. & Deeter, T. (1988). Development of the sport orientation questionnaire. *Research Quarterly for Exercise and Sport, 59*(3), 191-202.

Grace, P. (1990). Milestones in athletic trainer certification. *Journal of Athletic Training, 34*(3), 285-291.

Henry, S. & Holzemer, W. (1993). The relationship between performance on computer-based clincial simulations and two written methods of evaluation: Cognitive examination and self-evaluation of expertise. *Computers in Nursing, 11*(1), 29-34.

Henry, S. & Waltmire, D. (1992). Computerized clinical simulations: a strategy for staff development in critical care. *American Journal of Critical Care, 1*(2), 99-107.

Iserson, K. (1999). Simulating our future: Real changes in medical education. *Academic Medicine, 74*(7), 752-754.

Johnson, L., Wohlgemuth, B., Cameron, C., Caughman, F., Koertge, T., Barna, J., Schulz, J. (1998). Dental Interactive Simulations Corporation (DISC):

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Simulations for education, continuing education, and assessment. *Journal of Dental Education, 62*(11), 919-928.

Kerman, P. (2000). Director 8 Shockwave Studio: definitive multimedia authoring tool gets usability boost. *Macworld, 17*(7), 36-37.

Kulinna, P. & Silverman, S. (1999). The development and validation of scores on a measure of teachers' attitudes toward teaching physical activity and fitness. *Educational and Psychological Measurement, 59*(3), 507-517.

Leighton, R. & Sheldon M. (1997). Model for teaching clinical decision making in a physical therapy professional curriculum. *Journal of Physical Therapy Education, 11*(2), 23-30.

Lowdermilk, D. & Fishel, A. (1991). Computer simulations as a measure of nursing students' decision-making skills. *Journal of Nursing Education, 30*(1), 34-39.

Lyznicki, J., Riggs, J., & Champion, H. (1999). Certified athletic trainers in secondary schools: report of the Council of Scientific Affairs, American Medical Association. *Journal of Athletic Training, 34*(3), 272-276.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Mansen, T. & Haak, S. (1996). Evaluation of health assessment skills using a computer videodisc interactive program. Journal *of Nursing Education, 35*(8), 382-384

McCarthy, W. (1966). An assessment of the influence of cueing items in objective examinations. *Journal of Medical Education, 41*, 263-266.

McKenzie, J., Wood, M., Kotecki, J., Clark, J., & Brey, R. (1999). Establishing content validity: Using qualitative and quantitative steps. *American Journal of Health Behavior, 23*(4), 311-318.

McGuire, C. & Babbott, D. (1967). Simulation technique in the measurement of problem-solving skills. *Journal of Educational Measurement, 4*(1), 1-9.

Melnick, D. (1990). Computer-based clinical simulation: state of the art. *Evaluation & the Health Professions, 13*(1), 104-120.

Melnick, S. & Henk, W. (1997). *Content validation: A comparison of methodologies*. Proceedings of the Eastern Educational Research Asssociation, Hilton Head, SC. (ERIC Document Reproduction Service No. ED414329).

National Athletic Trainers' Association. (2002). Prehospital Care of the Spine-Injured Athlete.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Retrieved April 22 , 2002, from http://www.nata.org/downloads/0684602NATAPreHospital.pdf

National Athletic Trainers' Association. (1999). Athletic Training Clinical Proficiences. Retrieved May, 10, 2002 from http://www.cewl.com/clined/clinprof.html

National Athletic Trainers' Association Board of Certification. (2002a). Retrieved April 22 , 2002 from http://www.nataboc.org

National Athletic Trainers' Association Board of Certification. (2002b). Frequently Asked Questions. Retrieved April 22 , 2002 from http://www.nataboc.org/fut_ATC/links/faq/

National Athletic Trainers' Association Board of Certification. (2001). NATA Nashville Presentation. Retrieved March 22 , 2002 from http://www.nataboc.org/ATC/links/_nashville/

National Athletic Trainers' Association Board of Certification. (2000). 1999 Exam Report. Retrieved April 22 , 2002 from http://www.nataboc.org/ATC/docs/other/reporting99.pdf

National Center for Catastrophic Sports Injury Research. (2002). Eighteenth annual report fall 1982 - spring

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

2000. Retrieved April 22 , 2002 from http://www.unc.edu/depts/nccsi/AllSport.htm

National Federation of State High School Associations. (1997). NFHS Handbook.

Newble, D., Hoare, J., & Baxter, A. (1982). Patient Management Problems: issues of validity. *Journal of Medical Education, 16*, 137-142.

Norcini, J., Meskauskas, J., Langdon, L., & Webster, G. (1986). An evaluation of a computer simulation in the assessment of physician competence. *Evaluation & the Health* Professions, 9(3), 286-304.

Orr, N. (1988). Non-cognitive correlates of performance on conventional and interactive components of a computerized examination. *Proceedings of the Annual Conference on Research in Medical Education, 27*, 284-9.

Payne, D. (1997). *Applied Educational Assessment*. Belmont, CA: Wadsworth Publishing.

Peitzman, S., Nieman, L., & Gracely E. (1990). Comparison of "fact-recall" with "higher-order" questions in multiple-choice examinations as predictors of clinical performance of medical students. *Academic Medicine, 65*(9), S59-S60.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Scheuneman, J., Van Fan, Y., & Clyman, S. (1998). An investigation of the difficulty of computer-based case simulations. *Medical Education, 32*, 150-158.

Shearer, F.J., & Breitbach, T. (1997). Sims-Master [computer program]. Gardner, KS: Cramer Products, Inc.

Sireci, S. (1998). Gathering and analyzing content validity data. *Educational Assessment, 5*(4), 299-321.

Sireci, S. & Geisinger, K. (1995). Using subject-matter experts to assess content representation: an MDS analysis. *Applied Psychological Measurement, 19*(3), 241-255.

Smith, L. & Hambleton, R. (1990). Content validity studies of licensing examinations. *Educational Measurement: Issues and Practice, 9*(4), 7-10.

Swanson, D., Norcini, J., & Grosso, L. (1987). Assessment of clinical competence: written and computer-based simulations. *Assessment and Evaluation in Higher Education, 12*(3), 220-246.

Swanson, D., Norman, G., & Linn, R. (1995). Performance-based assessment: lessons from the health professions. *Educational Researcher, 24*(5), 5-11, 35.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Tharp, C. (1990). Computerized clinical simulation testing: the future is happening today. *Issues, 11*(1), 1, 10-11.

Van Camp, S., Bloor, C., Mueller, F., Cantu, R., & Olson, H. (1995). Nontraumatic sports death in high school and college athletes. *Medicine and Science in Sports and Exercise, 27*(5), 641-647.

Wong, J., Wong, S., & Richard, J. (1992). Implementing computer simulations as a strategy for evaluating decision-making skills of nursing students. *Computers in Nursing, 10*(6), 264-269.

Yalow, E. & Popham, W. J. (1983). Content validity at the crossroads. *Educational Researcher, 12*(8), 10-14, 21.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.