Annual Report for the 2000 Testing Year
CASTLE Worldwide, Inc.
Research Triangle Park, NC
March 2001

The National Athletic Trainers' Association Board of Certification (NATABOC) offers the designation Certified Athletic Trainer (ATC) for individuals who meet clearly defined criteria. This designation protects the public by establishing and evaluating individual compliance with entry-level standards for the athletic training profession. The NATABOC is the only accredited certification program for athletic trainers in the United States. Every five years, the NATABOC must undergo review and reaccreditation by the National Commission for Certifying Agencies.

As part of the requirements for being certified, candidates must pass NATABOC's three-part criterion-referenced examination consisting of a written multiple-choice test, a written simulation, and a practical test. The value of the credential rests in large measure on the quality of the NATABOC examinations. Accordingly, NATABOC works with CASTLE Worldwide, Inc., (CASTLE), a professional testing service specializing in certification and licensure examinations, to ensure that all guidelines and standards pertaining to the examination are fully satisfied. NATABOC monitors the performance of the examinations regularly to ensure that they function at the highest possible level of quality.

The Written, Simulation, and Practical Examinations

NATABOC's three-part examination is designed to provide a thorough assessment of a candidate's knowledge and abilities concerning the practice of athletic training. The written examination assesses whether a particular candidate has sufficient understanding of the principles, practices, and science underlying the practice of athletic training. The simulation examination assesses whether or not a candidate for certification makes decisions appropriately, while the practical examination assesses whether a candidate demonstrates the necessary level of skill in employing a wide variety of procedures and techniques.

While each of these examinations is unique, they must adhere to established psychometric principles. The purpose of these principles is to ensure that the NATABOC examinations are sufficiently valid and reliable.

Validity refers to whether or not the examinations actually assess what they are designed to assess. For NATABOC, the tests are designed to assess knowledge and skills in the area of athletic training. To achieve this purpose, all of the examinations are linked to the 1999 Role Delineation Study. The 1999 Role Delineation Study defines the current state of the profession, by describing the profession of athletic training in terms of domains, tasks, and knowledge and skill statements. Each item on a NATABOC certification examination must be linked to a domain, a task, and a knowledge or skill statement, as defined by the 1999 Role Delineation Study.

NATABOC employs other procedures to ensure that its examinations are valid. Each item on an examination is referenced to a published resource in athletic training. In addition, each item is

validated by a group of subject matter experts in terms of its importance, criticality, and relevance to the profession of athletic training.

Reliability refers to consistency in test scores. Every examination needs to be reliable; however, different types of examination require different reliability measures. For the written and simulation examinations, the most important type of reliability is referred to as internal consistency. For the practical examination, the most important type of reliability is interrater reliability.

Internal consistency refers to whether all the items on an examination measure the same characteristic. For NATABOC, the characteristic being tested is knowledge and ability in athletic training. While this may seem self-evident, a poorly designed examination may provide a measure of many things that it is not designed to examine. For example, an examination item that employs unusual words may be more a test of English than athletic training, while an examination item that uses double negatives and complex grammatical structures may be more a test of logical analysis.

Interrater reliability refers to whether independent judges are coming to the same conclusion concerning a candidate's performance. During the practical examination, the candidate is required to perform many complex procedures and techniques. It is extremely important that judges, working independently of each other, agree that the candidate either did or did not perform the procedure or technique correctly. If judges are not in agreement, then a candidate's score is less dependent on his or her skill, and more dependent on other factors such as the judges' personalities, experiences, etc.

While reliability is calculated in different ways for different types of examinations, many of the methods used to achieve high reliability are the same. Constant and consistent scrutiny of examination items by subject matter experts, psychometricians, and editors provides the best assurance of developing examinations that are highly reliable. For the NATABOC examinations, each item is written by experts in athletic training. These experts are trained in the characteristics of high quality examination items. Each question is then reviewed and edited by other experts in athletic training, who focus on characteristics that are important for the particular type of examination that the item will appear upon. For example, important characteristics of a multiple-choice item include the accuracy of the item, the correctness of the keyed response, the plausibility but incorrectness of the distractors, and the clarity and fairness of the item. Then psychometricians and editors from CASTLE conduct a review to ensure that standards governing the development of items have been met. Examination assembly then occurs, including a careful review of an analysis of the statistical performance of each item on an examination. The purpose of this review is to verify that the questions are fair and appropriate for candidates and to determine if modifications are necessary to enhance the psychometric quality of the items. For the practical examination, it is also important that the judges receive adequate training in how to score each item. Only after this extensive process is an item put on a NATABOC examination.

Reliability Ratings for Examinations Administered Between April 2000 and February 2001

Reliability computation for the NATABOC examination is an estimate of the consistency of the scores as a measure of competence in athletic training. While NATABOC computes reliability coefficients for each form and test date, the following report is based on the annual statistics for each form and part of the examination. Internal consistency reliability is reported as the Kuder Richardson [KR(20)] coefficient and accounts for the degree to which items on the tests

contribute consistently to candidates' scores. The KR(20) statistic ranges from 0 to 1, with coefficients above .70 meeting minimum standards. The standard error of measurement is the range within which candidates' true scores lie.

Table I. Internal Consistency, Standard Error of Measurement (S.E.M.), and Decision Consistency

	Written		Practical				Simulation*	
Form	282	283	286	287	288	289	284	285
Reliability KR20	.78	.79	.94	.95	.95	.94	.89	.89
Interrater Reliability % Agreement			.92	.91	.91	.91		
S.E.M.	5.12	5.08	1.58	1.42	1.28	1.37	33.06	32.92
Decision Consistency	.79	.81	.94	.96	.95	.95	.89	.89

**Reliability coefficients based on raw scores*

Since the purpose of the certification is to make valid and reliable pass/fail decisions, NATABOC also calculates decision consistency estimates using the Livingston formulation. The result of this analysis indicates that the classification of candidates in pass/fail categories exhibits reasonable consistency for all parts of the test. (The decision consistency coefficient is interpreted in a manner comparable to internal consistency.)

The accuracy of scores on the practical portion of the examination is a function of the interrater reliability. Examiners observe candidate performance on the practical examination and, using the scoring criteria defined for each problem, record whether or not the candidate performs required tasks at a level that is at least minimally competent. Interrater reliability is the measure of agreement between the examiners. A high level of agreement indicates that the scores are highly accurate.

How Candidates Perform on the Certification Examination

During the 2000 testing year, a total of 4976 candidates (including both first-time candidates and those retaking parts of the examination) took the written multiple-choice part of the NATABOC examination. In addition, a total of 4161 candidates (including both first-time candidates and those retaking parts of the examination) took the practical part of the NATABOC examination, and 4403 candidates (including both first-time candidates and those retaking parts of the examination) took the written simulation part of the NATABOC examination. A total of 36.58% of the candidates passed all three parts of the examination on the first attempt. Curriculum candidates passing all three sections of the exam on the first attempt was 45.42%. First time passing rate for internship candidates was 31.03%. The performance of the total group of candidates is presented in Table II. Given that the various forms are equivalent, the data below are presented across forms.

Table II. All Candidates Taking the Examination

	Total Sitting	# Pass	% Pass	# Fail	% Fail
Written	4976	2200	44.21%	2776	55.79%
Practical	4161	2577	61.93%	1584	38.07%
Simulation	4403	2429	55.17%	1974	44.83%

For those candidates who are taking the test for the first time, it is required that they take all three parts of the test on the same test date. Before taking the examination, they must meet eligibility requirements as curriculum or internship candidates. In following table (Table III), the analysis of first-time candidates is shown according to their route to eligibility.

Table III. Candidates Taking the Examination for the First Time

	Total Sitting	Total Pass / %	Total Fail / %	Total Intern	# Intern Pass / %	# Intern Fail / %	Total Curr	# Curr Pass / %	# Curr Fail / %
Written	2718	1481 / 54.49	1237 / 45.51	1585	755 / 47.63	830 / 52.37	1133	726 / 64.08	407 / 35.92
Practical	2716	1757 / 64.69	959 / 35.31	1584	897 / 56.63	687 / 43.37	1132	860 / 75.97	272 / 24.03
Simulation	2718	1589 / 58.46	1129 / 41.54	1584	826 / 52.15	758 / 47.85	1134	763 / 67.28	371 / 32.72

Table IV. Candidates Retaking the Examination

	Total Sitting	Total Pass / %	Total Fail / %	Total Intern	# Intern Pass / %	# Intern Fail / %	Total Curr	# Curr Pass / %	# Curr Fail / %
Written	2258	719 / 31.84	1539 / 68.16	1570	447 / 28.47	1123 / 71.53	688	272 / 39.54	416 / 60.46
Practical	1445	820 / 56.75	625 / 43.25	1063	560 / 52.68	503 / 47.32	382	260 / 68.06	122 / 31.94
Simulation	1685	840 / 49.85	845 / 50.15	1170	540 / 46.15	630 / 53.85	515	300 / 58.25	215 / 41.75

Descriptive Statistics of Candidate Performance

Scores on the written multiple-choice part of the examination are scaled from 0 to 150, with the criterion-referenced cut score established at 106. Scores on the practical are scaled from 0 to 50, with the criterion-referenced score required to pass established at 35. Scores on the written simulation range from 200 to 800, with the criterion-referenced cut score set at 500. Table V provides descriptive statistics on candidate performance for each part of the test. One-way Analysis of Variance (ANOVA) was used to test differences between internship and curriculum candidates. As seen below, scores for all curriculum candidates were significantly higher than those for internship candidates.

Table V. Candidate Score Distributions

	Written 282		Written 283	
High Score	134		134	
Low Score	56		67	
Average Score	Overall	103.21	Overall	106.19
	Internship	101.54	Internship	104.68
	Curriculum	105.97	Curriculum	108.99
Standard Deviation	10.92		10.85	

Written By Domain	**Domain I n items = 31**		**Domain II n items = 59**		**Domain III n items = 42**		**Domain IV n items = 9**		**Domain V n items = 9**	
	282	**283**	**282**	**283**	**282**	**283**	**282**	**283**	**282**	**283**
High Score	28	28	55	55	37	36	9	9	8	8
Low Score	8	7	18	22	8	8	2	1	0	1
Avg. Score	19.74	19.59	39.79	39.81	24.48	23.01	7.01	6.91	5.29	5.04
Std. Dev.	3.11	2.89	5.78	5.39	4.27	4.34	1.3	1.34	1.4	1.34

Practical	**286**		**287**		**288**		**289**	
High Score	49		47		49		50	
Low Score	8		3		10		4	
Avg. Score	Overall	35.06	Overall	35.15	Overall	35.32	Overall	37.04
	Internship	33.47	Internship	34.18	Internship	34.26	Internship	36.33
	Curriculum	37.29	Curriculum	36.67	Curriculum	37.82	Curriculum	38.82
Std. Dev.	6.49		6.55		7.52		6.53	

Simulation	**Simulation 284**		**Simulation 285**	
High Score	755		748	
Low Score	200		200	
Avg. Score	Overall	501.12	Overall	501.26
	Internship	488.89	Internship	489.59
	Curriculum	521.40	Curriculum	520.94
Std. Dev.	97.49		99.26	

NOTE: Curriculum Scores were significantly higher for all tests: $p < .01$

Conclusion

NATABOC works hard to ensure that the certification examination and all aspects of its development and administration are fair and of high quality. In addition, the testing program has a strong foundation for content validity and functioning well to protect the public. The organization adheres to pertinent standards governing certification tests and implements an examination program that continues to be a valid and reliable measure of entry-level competence in the professional practice of athletic training. Only candidates who truly demonstrate competence are successful in achieving certification.