National Athletic Trainers' Association Board of Certification, Inc.

Annual Examination Report: 2001
CASTLE Worldwide, Inc.
April, 2002

The testing year beginning in April 2001 and ending in March 2002 continued the tradition of strong performance for the examination program of the National Athletic Trainers Association Board of Certification (NATABOC). Candidates take the NATABOC examination after meeting eligibility requirements as graduates of either approved curriculum programs or internship programs in order to achieve national certification in athletic training. The NATABOC credential (Certified Athletic Trainer: ATC) is accepted in lieu of licensure in many states and is a requirement for employment virtually throughout the United States.

Because of the role played by the examination in making high stakes decisions that affect athletes, candidates for certification, and other stakeholders, the examination is developed, administered, and scored in accordance with rigorous psychometric standards. Key among the standards are the *Joint Technical Standards for Educational and Psychological Tests* (AERA, APA, and NCME, 1999), *Guidelines for Employee Selection Criteria* (EEOC, 1978 and updated), and *Standards for Accreditation of National Certifying Agencies* (NCCA, 1995 and 2002). These standards address important matters of quality and fairness.

The NATABOC Examination

The certification examination comprises three parts: a multiple-choice test of 140 scored items and 10 non-scored, experimental items (per form and test date, allowing 100 questions to be pretested each year), a written simulation test of eight multi-part problems, and a practical examination with about 16 problems, depending on the form. Each part contributes to the assessment of athletic trainer competence by emphasizing knowledge, professional judgment, and skill. All parts of the NATABOC examination are based on the Role Delineation Study completed in 1999.

The multiple-choice examination is primarily a test of knowledge, with items classified in accordance with three levels of Bloom's taxonomy: recall and understanding, application, and analysis. Qualified subject matter experts write, review, revise, reference, and validate the items in accordance with best practice as identified through research literature on multiple-choice items. Item analysis supplements expert judgment in evaluating the quality of each question. The item bank includes 1052 questions that have been approved for use on the test.

NATABOC employed two equivalent versions, or forms, of the multiple-choice examination in 2001. Test assembly follows a structure that permits the use of linear equating procedures (Tucker model), with a pool of equator items, items that are unique to the form, and experimental items that do not count in the scoring formula but that allow NATABOC to evaluate the performance of the question from a statistical perspective before using it as a scored item. The multiple-choice committee, which writes and revises items and assembles this part of the test, selected a new pool of equator items in July, 2001, in order to re-anchor the examination in April, 2002.

The written simulation part of the examination assesses professional judgment and decision making. The written simulation committee writes, reviews, and validates new problems each year, and uses item statistics to modify problems to improve their performance. The item bank of simulation problems includes 30 problems.

NATABOC administered two versions of the written simulation in 2001. Scoring is based on the responses selected by the candidates when they take the test. Each response option is classified as Clearly Indicated, Indicated, Neutral, Contraindicated, or Clearly Contraindicated and weighted according to the importance of the section of the problem in which it is found. Scores are then calculated in accordance with the scoring formula, equated, and reported on a scale ranging from 200 to 800.

The practical portion of the examination is a skills assessment, focusing on evaluation and rehabilitation techniques for different sections and systems of the body. Test assembly uses problems that have been developed, reviewed, and revised by the practical committee and pretested by the test administration committee. Various athletic training supplies are required for the administration of the test, which also includes a live model (an ATC). The practical item bank includes 63 problems.

Scoring of the practical examination is based on the observations of two trained examiners who observe candidate performance in a live, standardized setting. For each problem there are scoring criteria written so that examiners mark Yes or No, depending on whether or not they observed the targeted behavior. Each criterion is weighted, and total scores represent the arithmetic average of the two examiners. Scores are equated and scaled to range from zero to 50 points, although the number of scoring criteria typically ranges from 70 to 80.

Reliability

NATABOC computes reliability coefficients to evaluate the consistency of scores on each part of the examination, in accordance with the purpose and nature of the parts of the test. The analysis is conducted on each test date and form. 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 candidate's 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. The decision consistency reliability, computed with the Livingston formulation, estimates the accuracy of pass/fail decisions, and should be interpreted in a manner that is similar to the KR(20).

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

	Written		Practical				Simulation	
Form	290	291	294	295	296	297	292	293
Int'l Consistency KR(20)	.78	.79	.95	.94	.93	.95	.96	.95
Interrater Rel. Percent Agreement			.93	.92	.94	.92		
Interrater Rel. Pearson's ?			.92	.89	.91	.90		
S.E.M.	5.02	5.06	1.24	1.19	1.25	1.17	20.46	21.07
Dec. Consistency	.79	.81	.94	.92	.92	.90	.91	.90

**Reliability estimates are reported as coefficient based on scaled scores*

Scoring of the practical portion of the examination relies on the observation of examiners. Because this is the case, it is important to assess the consistency of examiner observations as the best indicator that scores are valid indications of actual candidate performance. Interrater reliability analysis is used for this purpose.

Candidate Performance on the Certification Examination

A total of 5285 candidates took the multiple-choice portion of the NATABOC examination in the testing year 2001. This number includes both first-time and repeat candidates. In addition, 3936 candidates took the practical portion of the test, again including both first and repeat candidates. Finally, 4441 first-time and repeat candidates sat for the written simulation during the year. Specific pass/fail information is provided in Table II for all candidates, and Tables III and IV summarize the performance of first-time and repeat candidates separately. Given that the forms of the examination are equivalent, the information below is reported across forms. Overall, a total of 878 candidates passed all three parts of the examination on their first attempt, which represents 33.94%. 487 curriculum candidates (42.57%) and 391 (27.10%) internship candidates passed all three parts on their first attempt.

Table II: All Candidates Taking the Examination

	Total Sitting	**# Pass**	**% Pass**	**# Fail**	**% Fail**
Written	5285	2316	43.82	2969	56.18
Practical	3936	2384	60.57	1552	39.43
Simulation	4441	2431	54.74	2010	45.26

Candidates taking the test for the first time must sit for all three parts on the same test administration date. Curriculum candidates achieve eligibility for ATC certification by completing a course of study that has been approved by the National Athletic Trainers Association Education Committee. Internship candidates complete a set of seven required courses, a supervised internship, and various other requirements as part of meeting eligibility requirements.

Table III: Candidates Taking the Examination for the First Time

	Total Sitting	Total Pass/%	Total Fail/%	Total Intern	#Intern Pass/%	#Intern Fail/%	Total Curric.	#Curric. Pass/%	#Curric. Fail/%
Written	2595	1128 43.47	1467 56.63	1449	396 27.33	1053 72.67	1146	732 63.87	414 36.13
Practical	2590	1650 63.71	940 36.29	1446	806 55.74	640 44.26	1144	844 73.78	300 26.22
Simulation	2592	1482 57.18	1110 42.82	1446	739 51.11	707 48.89	1146	743 64.83	403 35.17

Table IV: Candidates Retaking the Examination

	Total Sitting	Total Pass/%	Total Fail/%	Total Intern	#Intern Pass/%	#Intern Fail/%	Total Curric.	#Curric. Pass/%	#Curric. Fail/%
Written	2690	890 33.09	1800 66.91	1866	550 29.47	1316 70.53	824	340 41.26	484 58.74
Practical	1346	732 54.38	614 45.62	970	476 49.07	494 50.93	376	256 68.09	120 31.91
Simulation	1849	947 51.22	902 48.73	1237	591 47.78	646 52.22	612	356 58.17	256 41.83

Descriptive Statistics of Candidate Performance

There are 140 scored items on each form of the multiple-choice examination, but scores are scaled and reported on a scale ranging form 0 to 150, with the criterion-referenced passing standard set at 106. Scores on the practical test are scaled and reported to range from 0 to 50, with the criterion-referenced passing standard established at 34. The written simulation part of the examination has a score range from 200 to 800, with the criterion-referenced passing standard anchored at 500. Table V provides descriptive statistics on candidate performance for each part of the test. A one-way Analysis of Variance (ANOVA) was used to test differences between curriculum and internship candidates. As seen below, scores for curriculum candidates were significantly higher than those for candidates from the internship route of entry.

Table V: Candidate Score Distributions

Multiple Choice Examination

	Written 290	**Written 291**
High Score	140	135
Low Score	32	53
Average Score	Overall 104.33 Internship 102.59 Curriculum 107.26*	Overall 101.70 Internship 99.95 Curriculum 104.64*
Standard Deviation	10.76	10.98

Written By Domain	**Domain I n items = 21**		**Domain II n items = 33**		**Domain III n items = 28**		**Domain IV n items = 31**		**Domain V n items = 15**		**Domain VI n items = 12**	
	290	**291**	**290**	**291**	**290**	**291**	**290**	**291**	**290**	**291**	**290**	**291**
High Score	20	21	32	31	28	28	29	29	15	15	12	12
Low Score	2	4	5	8	6	11	9	7	2	2	2	2
Avg. Score	15.02	13.76	22.02	20.03	21.15	21.26	19.02	20.38	10.20	11.65	8.17	8.13
Std. Dev.	2.11	2.33	3.32	3.51	2.68	2.65	2.98	3.02	1.83	1.70	1.65	1.61

Table V: Candidate Score Distribution (Continued)

Practical Examination

	294	295	296	297
High Score	49	49	49	50
Low Score	2	21	21	4
Average Score	Overall 38.78 Internship 37.21 Curriculum 40.72*	Overall 39.37 Internship 38.25 Curriculum 41.07*	Overall 38.72 Internship 37.80 Curriculum 40.60*	Overall 38.31 Internship 37.63 Curriculum 40.02*
Std. Dev.	7.10	6.34	6.64	6.77

Written Simulation

	Written 292	Written 293
High Score	737	729
Low Score	200	200
Average Score	Overall 500.81 Internship 486.24 Curriculum 520.89*	Overall 500.99 Internship 488.16 Curriculum 523.30*
Standard Deviation	98.67	98.51

NOTE: Curriculum Scores were significantly higher than internship scores (*=$p<.01$)

Conclusion

NATABOC applies accepted psychometric procedures to ensure the fairness of its examination program, addressing such matters as validity, clarity, reliability, and appropriateness in a systematic manner. It adheres to pertinent standards for certification and licensure tests and continues to achieve a desirable level of quality.