

Interim Report on the Pilot Trainee

Selection Examination

By

Progeny Systems Corporation

Prepared for the Board of Pilot Commissioners for the Bays of San Francisco, San
Pablo, and Suisun

Submitted to Captain Allen Garfinkle

Executive Director

July 23, 2014

DRAFT

San Francisco Pilots Examination Report

Overview

The selection program applied by the Board of Pilot Commissioners (BOPC) for the Bays of San Francisco, San Pablo, and Suisun to select pilot trainees consists of three equally weighted components. All three components must be passed by the candidate. The values for each of the components are equally weighted. The scores for the three components, experience, written, and simulator are totaled with the candidates ranked by the sum of the three scores.

The process begins with an application by the candidate to establish that they have sufficient experience to meet the requirements of 7 CCR, Division 2, Article 4, Subsection 213(e). The application is evaluated and those candidates who can establish proof of sufficient experience and meet the other requirements are able to sit for the written examination.

The written examination consists of 150 multiple-choice questions. Each question is weighted one point. The questions are distributed according to the weight of the subject matter areas. The weights were established according to the results from a job analysis. There are four subject matter areas: (a) pre-transit planning, (b) master/pilot transition, (c) route piloting, and (d) mooring and unmooring. Multiple focus groups reviewed all of the questions in BOPC's item bank. Every item was renewed/refreshed by reformulating the stem and/or the distracters. Also, every item was referenced to a current reference. Additionally, new questions were written for the new editions or new references. Every item on the examination had been edited or was new so none of the candidates had seen the items previously to sitting for the examination.

The simulation examination evaluates candidates on the following seven areas: (a) situational awareness, (b) appropriate response, (c) ability to respond correctly under stress, (d) communication and bridge presence, (e) fundamental shiphandling, (f) bridge resource management, and (g) rules of the road. The development of the simulation examination was an intensive process involving a computer programmer, a coordinator who was a retired pilot, multiple active pilots, the BOPC staff and a psychometrician. The development was a process of multiple iterations of developing, live testing with pilots, and revising. Multiple revisions were necessary to ensure that virtually all of the actions that could be taken by pilots were identified and built into the system. The scoring system was designed around the metrics of +2 for highly effective, +1 for acceptable, 0 for ineffective, -1 for touching, and -2 for a collision or alison. Evaluation forms were designed such that the candidates' evaluations were standardized. Numerous iterations of the evaluation form were necessary in order to achieve consensus about the wording, rating scales, and order of presentation of the measurement opportunities. A great deal of planning and effort were devoted to ensuring that the examination experience was the same for each candidate. To enhance the measurement accuracy, 53 measurement opportunities were developed.

Results

Experience Points. The BOPC evaluated the applications and determined that 38 individuals met or exceeded the minimum requirements to sit for the written examination. Individuals were awarded from 10 to 35 points. The points awarded would be combined to the points earned from the written examination if they passed the written examination.

Written Examination. Candidates were afforded 4 ½ hours to respond to the 150 multiple-choice items. The candidates were seated in the California Maritime Academy's cafeteria at large round tables, two candidates to a table. The arrangement ensured that candidates were not able to observe other candidate's responses. Of the 38 qualified candidates, 33 sat for the written examination.

The measurement properties of the examination were very strong. Of the thirty-three persons who sat for the examination, 25 achieved a passing score. The lowest score was 60 and the highest was 139. The mean score was 109.58 and the median score was 113.00. Of particular note was the excellent reliability (coefficient alpha) of .94. For a perspective, the maximum positive reliability coefficient possible is 1.00. The examination functioned very effectively and the items discriminated consistently between the high performing candidates and the lower performing candidates. In other words, the higher performing candidates consistently answered the items correctly while the lower performing candidates were much less consistent. The written examination performed effectively in identifying the candidates who possessed the job knowledge required of a pilot trainee.

For licensing examinations, it is necessary to establish a passing score (cut score) based on the concept of minimal acceptable competence. The methodology most frequently applied and which was used for the written examination was the modified Angoff. For this examination seven pilots, referred to as subject matter experts (SMEs), evaluated the performance expected of minimally competent trainees. After the SMEs received training in the process, they responded to the following question. "What percentage of minimally competent candidates (pilot trainees) would answer the item correctly." The data was aggregated across all items and all raters and divided by seven, the number of SMEs. The results from the workshop established the passing score (cut score) at 101. With the passing score at 101, 25 candidates passed the examination and eight failed. The passing score was optimally established—the closest score above was 104 and the closest below was 96 which left clear gaps on both the pass and fail sides of the passing score.

Simulation Examination. The evaluators engaged in examination development received extensive training in the procedures for evaluating and scoring the candidates. For the training, the evaluators proceeded as if the person piloting the simulator was an actual pilot. Each evaluator conducted their evaluation independently. After each simulation experience, after the test pilot had been

evaluated, the three evaluators convened to discuss the candidates performance to ensure that nothing in the design of the simulation negatively impacted the try-out candidate's performance

The processes applied for training the actual evaluators proceeded along the same lines. The evaluators observed and rated pilots during numerous dry runs until the evaluators were consistent in their evaluations of the level of performance exhibited by the test pilots. The pilots were knowledgeable regarding the consequences of their ratings. Two perspectives were weighed by the evaluators. The first is the necessity that the pilot trainees are able to protect the public's health safety and welfare. The second is the interest to ensure that the candidates' are assessed fairly and provided opportunities to succeed in the examination program and selection process. The evaluators represented the BOPC's Pilot Evaluation Committee, state licensed pilots from another jurisdiction, and industry representatives with command experience on deep draft vessels.

The simulation examination was offered at California Maritime Academy on a full bridge simulator over three days, June 26th, 27th, and 28th, 2014. On the 25th, the candidates who were successful on the written examination were fully briefed on the process, received materials designed to assist in preparing for the examination which they were allowed to take from the site, received an hands-on orientation to the bridge, and observed the vessel's track through the simulation exercise.

At the scheduled examination time, the candidates were required to arrive in time to return study materials and to make final preparations prior to entering the simulator. There were six evaluators who observed the candidates performance and participated in the discussion of the candidate's performance as based on the scoring sheets. Participation by all six evaluators assisted in ensuring that the performance of the candidates was accurately recorded. From those six, three primary evaluator's ratings were recorded for each candidate to compute the candidate's ratings. To obtain a final score for each candidate the ratings were summed across the three evaluators and the 53 measurement opportunities. This means that the candidate's performance was measured 159 times resulting in an accurate scoring process. In psychometric measurement as with taking measurements with a ruler, multiple averaged ratings results in accurate measurement.

Similarly to the written examination, a modified Angoff passing score workshop was conducted. There were 12 SMEs who participated in the workshop. Included were the six evaluators, three PEC members, and three pilots who participated in the development of the simulation exercise. All of the evaluators were intimately familiar with the content of the examination and the evaluation process. All had been engaged in training pilots so they had an in-depth appreciation of the skill set needed by pilot trainees. They was asked to evaluate the level of performance that a minimally competent candidate would perform on each of the 53 measurement opportunities. They were asked to judge the level of performance based on the examination rating system, +2, +1, 0, -1, and -2. The passing score was a scaled 52.

Scaling. The three components of the assessment process were equally weighted. Because, the possible number of points that could be obtained varied by examination, it was necessary to scale the experience points and the simulation exercise. The maximum score achievable on the written examination was 150 so the maximum experience points and the maximum simulator scores were scaled to equal 150. The maximum experience points was 90 which was scaled to equal 150 and the maximum score achievable on the simulation exercise was 106 which was also scaled to equal 150.

Data: The candidates are ranked based on the sum of their performance scores for the three components. The passing candidate results and their rank on the list is provided in Table 1.

Table 1

Passing Candidates Ranked by Total Score

ID Num	Experience Scaled	Written	Simulation Scaled	Total Score
33	33	139	74	246
32	33	122	75	230
17	42	127	59	228
14	33	122	71	226
3	17	133	74	224
11	33	121	68	222
27	33	115	67	215
21	33	121	60	214
25	33	109	66	208
24	25	108	64	197
4	33	109	53	195
6	33	104	55	192
15	17	113	55	185

Table 2

List of Candidates who Failed Simulation

Written Passing = 101 Simulation Passing = 52

ID Num	Experience Scaled	Written	Simulation Scaled
1	33	122	9
5	25	118	26
9	33	129	42
13	25	117	30
16	33	104	32
18	33	120	20
19	33	109	44
20	33	128	42
22	25	132	43
23	17	112	47
26	25	115	36
28	58	106	13