SkillGrader

Simulation Assessment Technology to Improve Objectivity, Evaluation Granularity and Opportunities for Data Analysis

Murray W. Goldberg, Marine Learning Systems
Arvinder Aujla, Royal Canadian Navy

INTRODUCTION

Despite advances in training technology, one significant gap remains: how to assess the performance of observable skills in a way that eliminates bias and delivers actionable metrics. Skills assessment is typically paper-based and relies on the viewpoint of expert assessors, limiting our ability to extract metrics to drive improvement. Within the Canadian Naval Fleet School Pacific, the Navigation and Bridge Simulator (NABS) supports Naval Warfare Officer training for the Royal Canadian Navy. NABS employed SkillGrader in a pilot program to support their simulation assessment, improve the process and draw out deeper metrics on Officer skills and competency.

LIMITATIONS OF CURRENT ASSESSMENTS

- Require significant degree of assessor expertise to interpret observations
- Natural variability from one assessor to another, impacting reliability
- Assessor must 'compute' score on paper
- Limited value in data collected beyond immediate debrief

Survey of NABS CTOs and mentors:
- < 50% could identify trends and gaps over many assessments
- < 30% could say whether their training practices were more/less effective than past and why
- < 40% could easily rank trainees, past and present, for a given role
- Only 40% agreed that current practices were impartial and data-driven; even fewer felt that trainees saw them as such

TECHNOLOGY AND APPROACH

- Digitization of five existing paper assessment sheets that contain a scoring rubric into digital forms that use Performance Indicators:
  - Performance Indicators are binary (yes/no) observations
  - Intended to embody the rubric and derive the score electronically
  - Tagged according to the competencies they measure
  - Assigned their level of importance to the competency
  - Weighted according to how much they contribute to performance
- Assessors uses tablet-based application during simulation assessment
- SkillGrader algorithm displays performance outcome to facilitate debrief:
  - Detailed measurement of overall team performance score
  - Individual contribution to team and score for each competency
- Performance indicator data and calculated scores are collected in back-end server in support of further analysis

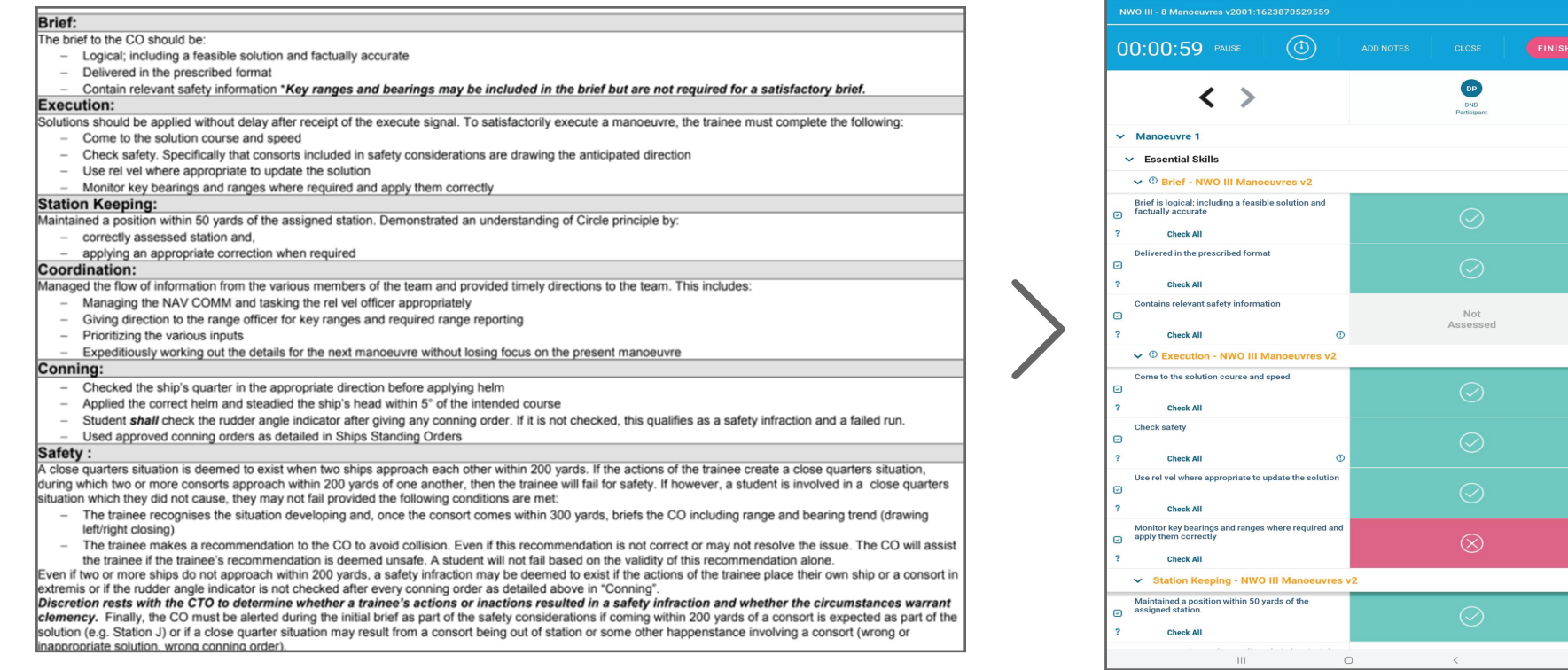FIGURE 1: Conversion of Paper Assessment to Digital Form on Application

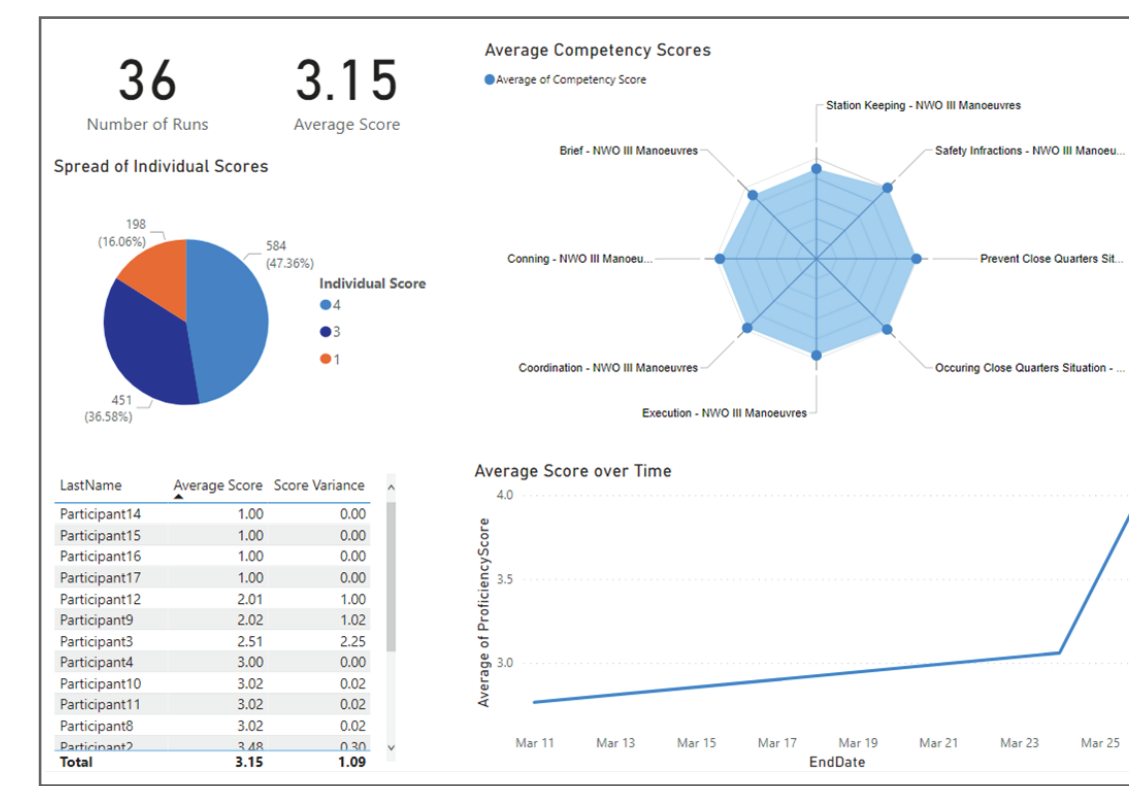FIGURE 2: Learning Analytics Example - Average Competency Scores

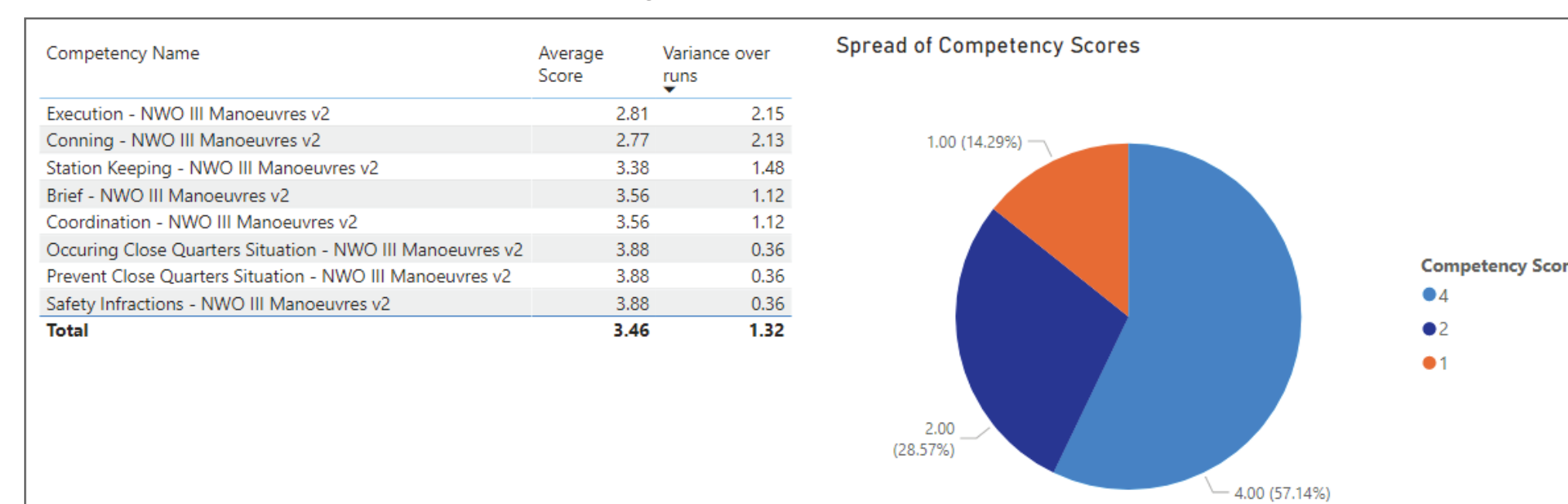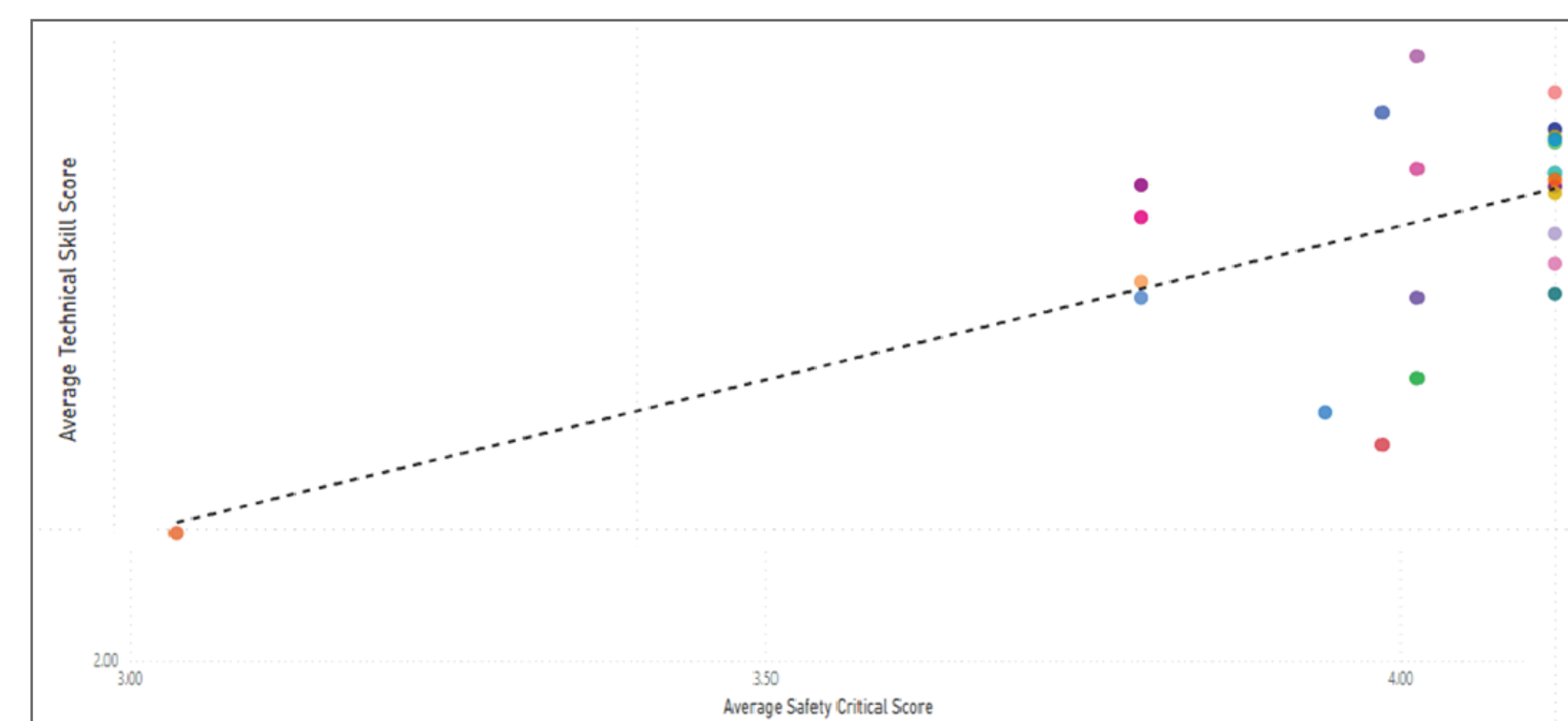FIGURE 3: Learning Analytics Example - Cohort Combined Report

FIGURE 4: Learning Analytics Example - Technical vs. Safety Score Plot

FINDINGS

- Existing paper forms could determine whether the student met the standard, however they are generally unable to identify a gradient of performance for those that are above the standard.
- Conversion process stimulated general discussion of objectivity vs. subjectivity, and forces a conscious choice to be made between the two.
- Early application analytics created hard data which backed up what assessors could feel intuitively, and allow concrete changes such as the creation of part-task scenario training to address areas of weakness
- Analytics revealed there was greater variance in performance of some competencies than others.
- Understanding degree of variance presents opportunities for streaming trainees into different operational paths, to measure the success of training to close gaps for the individual and for the program as a whole, and to support trainees throught their career in the armed forces.
- Assembly of data over time can generate insights and identify trends:
  - Individual strengths and weaknesses, from course to course and within their time at school, to develop individual learning plans
  - Intra-cohort analysis to detect CTO differences, mentor contributions, time-of-day effects, etc.
  - Inter-cohort analysis to reveal program-wide insights such as common skill gaps and strengths, timing of skill fade, etc.
  - Discovery of correlations between inputs (trainee background, grades, instructors, sea time) and outputs (skill assessment data)

CONCLUSION

- Opportunities for improvement:
  - As assessments can be dynamic, further work required to vary the number of assessment phases supported in situ.
  - There is value in subjective data, such as the ability to add notes, for more complex assessment.
- The work to introduce SkillGrader to NABS has been highly valuable in improving the utility and value of the application.
- The hope and expectation is that the pilot will result in a new model for assessing students beyond NWO courses for the school.
- The SkillGrader application has potential to be used in other operational contexts: to assess individual training events and team training events in the army and air force, as well as force generation events and force employment events.

APPROVED FOR PUBLIC RELEASE