

Automatically Scored Spoken Language Tests for Air Traffic Controllers and Pilots

Harcourt Assessment, Menlo Park, California, USA

1 November 2006
Bangkok, Thailand

Versant Tests may promote cooperation in global language training

Language training: Global challenges

Testing in ICAO LPR

Data-driven standardized tests

Versant Aviation English Test

Summary

Language Training: Global challenges

Limited training resources – financial & expertise

- Language training & testing for > 150K pilots & 50K ATCs

How to implement consistent proficiency levels across different regions of the world?

- Is level 4 in Asia the same as level 4 in Latin America?
- Auditing

How to comply with ICAO March 2008 deadline?

- About 320 working days remain
- Not enough qualified raters

Language Training: Overcoming challenges

Focus available training resources

- Allow trainer to train
- Training delivery & informing course development

Optimize instruction for efficiency & effectiveness

- Realize learning gains for each trainee in the quickest time
- Adopt training approaches that lower TCO

Evaluate and select valid and reliable test(s) & test systems

- Does the test meet ICAO standards & inform instruction?
- Is the test reliable in development and operation?

Versant Tests can help you meet ICAO standards

***Versant* Tests for English, Spanish & Dutch set a new standard in spoken language testing**

- consistent high precision in wide-range operation (standardized)
- NCES, Homeland Security, Dutch Gov't, ETS, and German Gov't

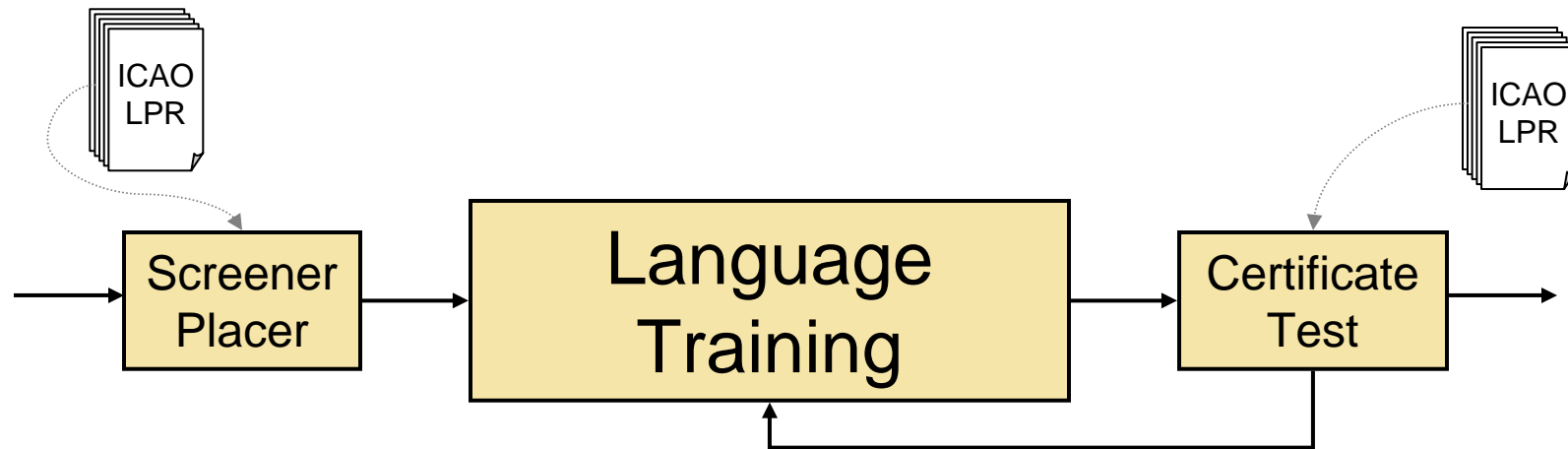
Versant Aviation English Test, CRDA with FAA Academy

- the very highest stakes testing
- demands accountability & high precision in world-wide operation

Describe the Versant tests: Their design, development & operation, and why they work

Importance of evaluating a test in operation

Testing in ICAO LPR



Testing gates training

Customized and proven training packages

Evaluating a Test

Evaluation Criteria

Valid: scores support warranted decisions

Reliable: consistent scoring

Secure: resistant to cheating or coaching

Practical: fits with local situation & budget

Validity and reliability should be demonstrated during development and in ongoing operation.

Data-driven standardized tests

Types of test and testing systems

- Standardized vs. ad hoc
- Closed loop vs. open loop
 - closed loop* ~ monitored in operation

Categories of scoring

- Measures vs. judgments
- Data driven vs. rubric referenced

Delivery-Scoring methods

- Human delivery – human judged
- Tech delivery – human judged
- Tech delivery – tech measurement

Versant Tests

Standardized tasks and scoring, delivered in a closed-loop system, scored with data-driven performance measurement.

- **Available world-wide, on demand**
- **Via telephone or via computer**
- **Scored with spoken language processing**
- **Scores returned within 5 minutes via Internet**

Proficiency & Automaticity

Spoken language proficiency

- Effective communication of relevant material

Automaticity measurement

- Supports voice communication under cognitive load
- Timed performance with limited contextual support

Advantages of automaticity measurement

- Efficient coverage in language skill space
- Multiple independent measures
- Produces stable, accurate estimates

Versant Base Measures

1. Response content

- Words & structures, as found in appropriate responses from high proficiency speakers (native & non-native)

2. Manner of speaking

- Latency, rate, intelligibility, rhythm, pausing, fluency

**Estimated with standard psychometric methods;
thus, producing consistent accuracy in operation.**

Enhanced Clarity of Construct

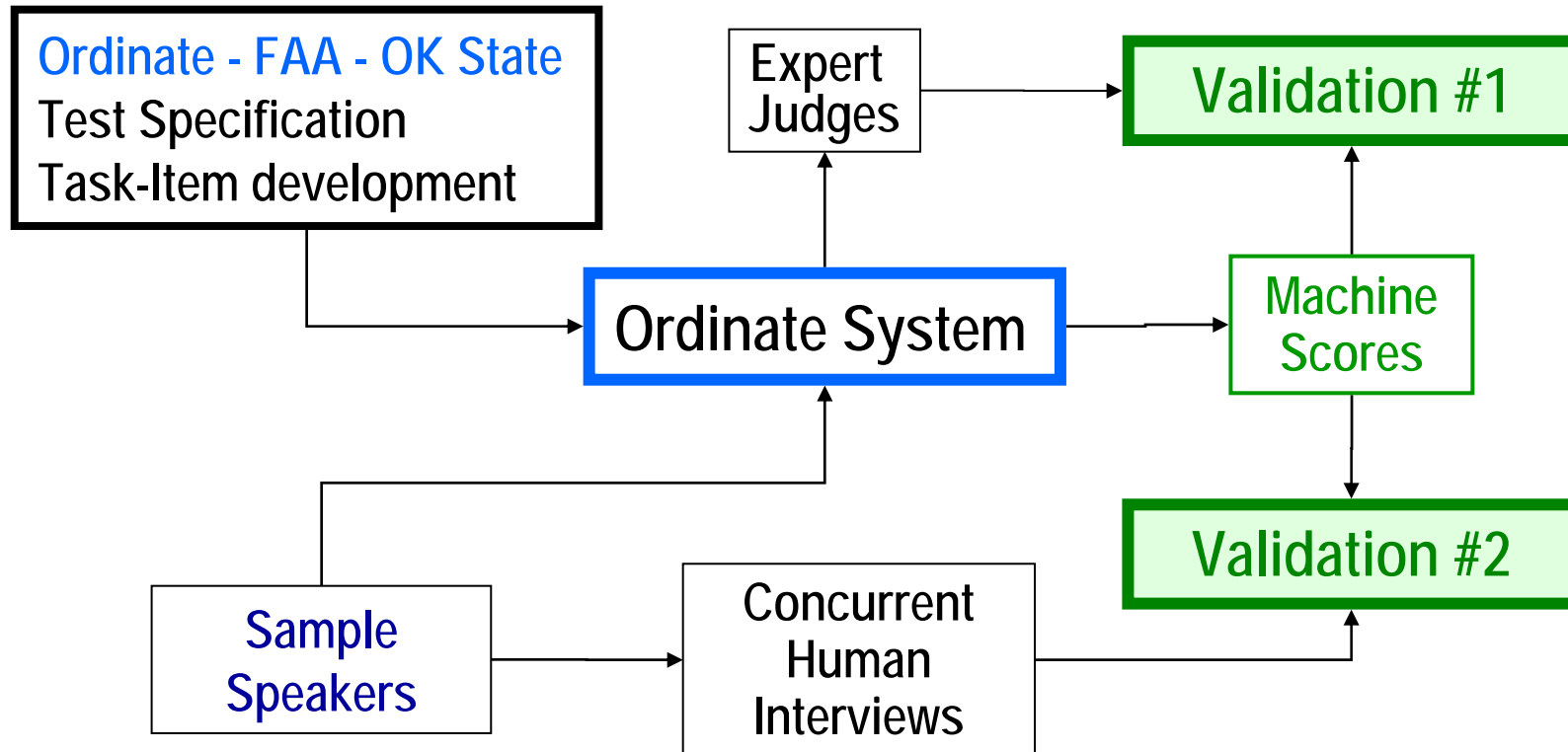
Example of ICAO proficiency rubric

- “Response time is slow and often inappropriate.
Interaction is limited to simple routine exchanges”

**Versant scoring provides a clear metric for
“slow” and “simple routine”.**

**Versant scoring implements a consistent, data-
driven decision on (in)appropriate responses.**

Versant Development and Validation



Versant Validation Outcomes

Validation #1

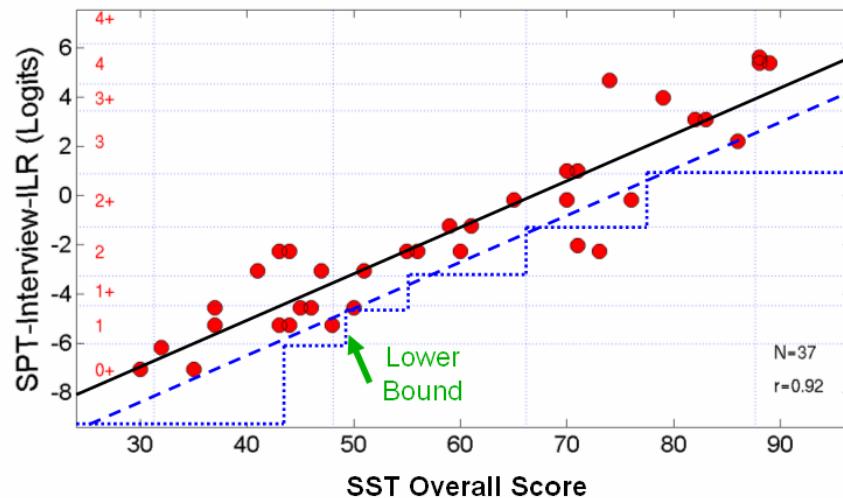
Machine measures tested skills at human accuracy levels.

Validation #2

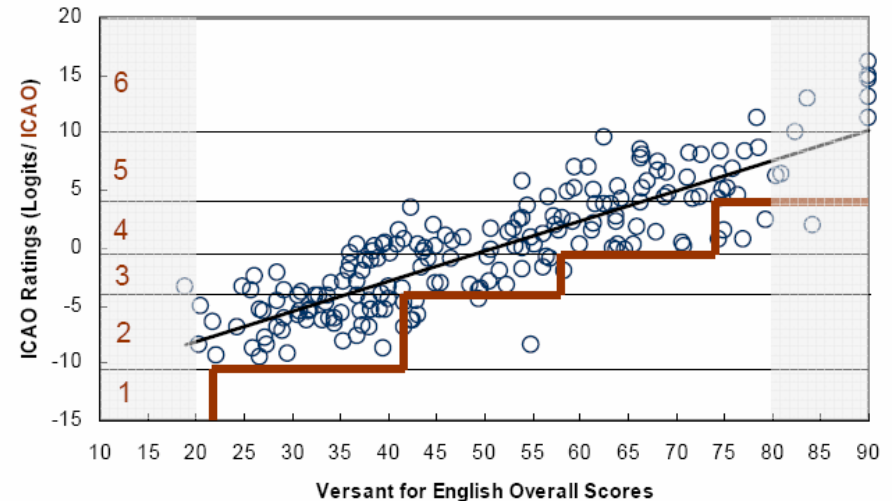
Versant task performance and Versant scoring will capture all the information produced by a human-human interactive test.

Validation Mappings

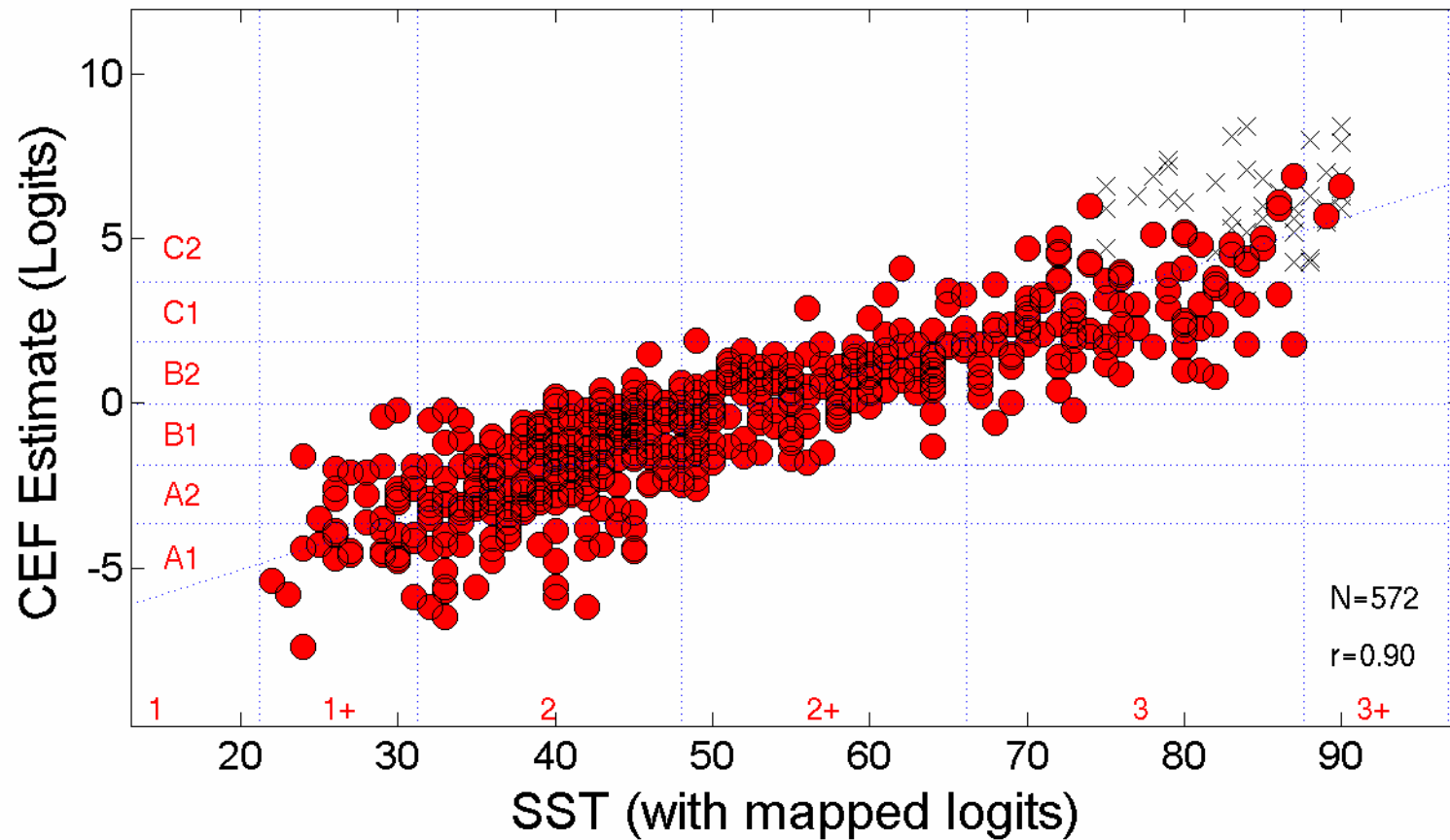
Versant Spanish test predicts human-human interactive test



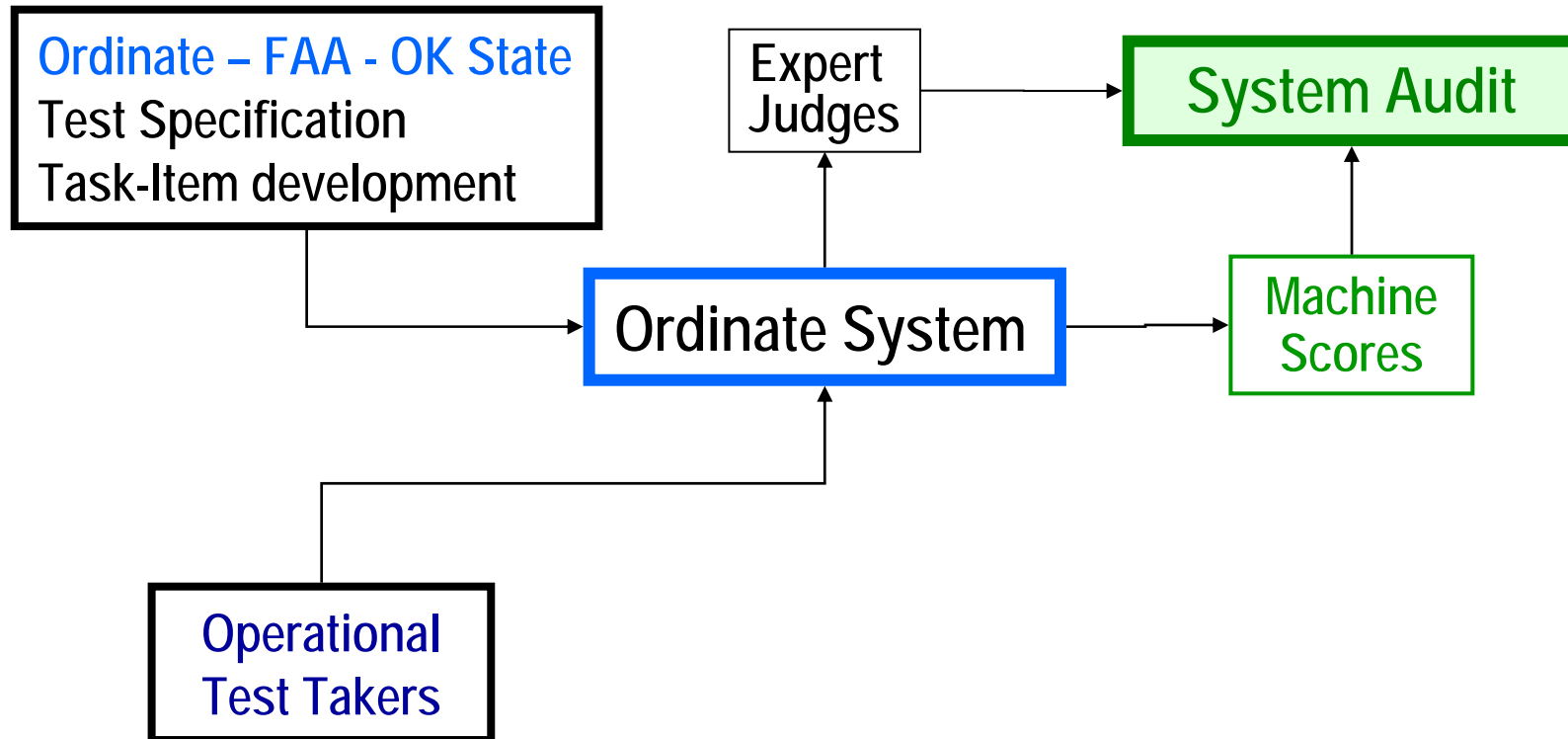
Versant English test mapped on human ICAO level judgments



Versant ~ CEF Estimates



Closed Loop Operational Audit



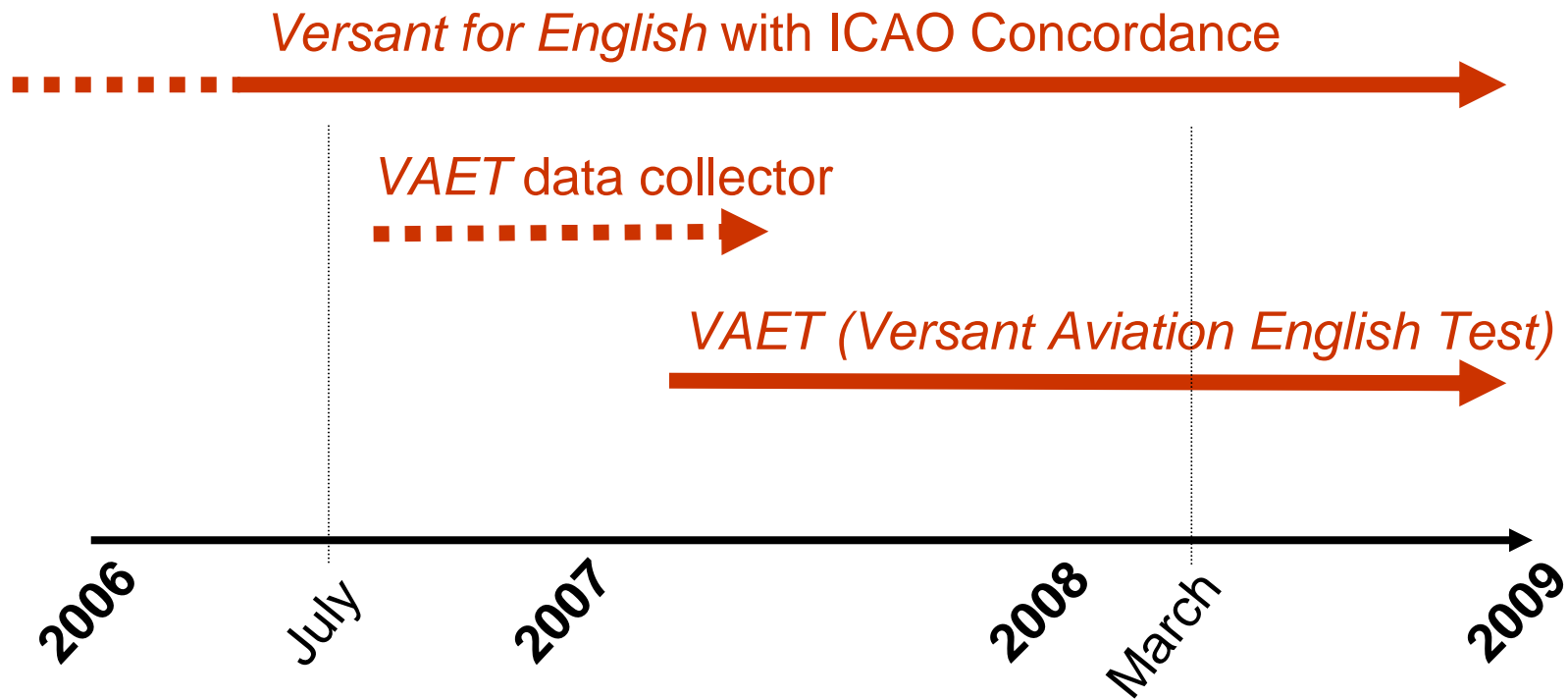
VAET Structure

Duration: 25 minutes

Tasks:

- | | |
|---------------------------------|----|
| – Aviation Readings | 2 |
| – Common English Readings | 2 |
| – Repeats | 16 |
| – Short-Answer Questions | 24 |
| – Readbacks | 10 |
| – Corrections and Confirmations | 10 |
| – Story Retellings | 3 |
| – Open Questions | 2 |

VAET Development Schedule



Global Partnerships

***Versant for English* with ICAO concordance is ready now for staff and training assessment.**

VAET development needs cooperative sites and organizations for review and data collection.

Sample test forms are available – please contact Masa Suzuki

MSuzuki @ ordinate.com

Summary: Reliable testing drives training quality

Repeatable, independent measurement of spoken language performance drives language training quality.

Data-rich, direct measures of listening/speaking performance can be applied for certification and auditing.

Versant for English (with ICAO concordance) is available now for placement/benchmarking.

Versant Aviation English Test in April, 2007 – demonstration version operational now.





THANK YOU

APPENDIX

The VAET Advantage – the 4 C's

- ▶ **Compliant** with ICAO standards
- ▶ **Credibility**
- ▶ **Convenience** and **Cost-effectiveness**

Compliance: VAET and ICAO Standards

ICAO Standard	VAET	Description
Both English and aviation content		Test has ICAO phraseology and plain English language tasks
Measure Level 4		Validation report will be provided – can measure ICAO levels 1 through 6
PSVFCI should be measured		Sub-scores showing performance on these skills will be available with score report
Test is secure and independently validated		Randomized item presentation, and rigorous test development process

Credibility

Test Design & Item Dev.

- FAA partnership
- Item review by experts in ICAO phraseology
- Stanford University 

Data Collection

- 2400 native and non-native voices
- Pilots and ATC responses
- **In progress**

Automated Scoring system development

- Scoring by multiple humans
- Patented speech processing technology

Validation

- Split-half, test-retest, Cronbach's alpha scores for reliability
- Data based validity evidence

10'th Global TRAINAIR, Oct. – Nov.2006

Convenient and cost-effective

Feature	Benefit	Outcome
Phone/PC based	No scheduling needed No raters needed	Almost no admin or logistical costs ON DEMAND testing
Fast assessment and score availability within a few minutes	Can do many concurrent assessments a day; Internet score reports	Reduces admin costs (no paper reports); eliminates lost or erroneous score reports
Sub-scores	Tailor training for each candidate to most needed areas	Reduces time spent in training; faster path to ICAO compliance
Hosted testing service	No software to install, no IT costs	Low Total Cost of Ownership (TCO)

Sample audio files

