



# **“Operationalizing” Test Quality and Fairness**

**Cathy Wendler, Ph.D.**  
**Principal Director, Research Management**  
**Foundational & Validity Research**  
**Educational Testing Service**



# Overview

- Existing standards
  - *Standards for Educational and Psychological Testing*
  - *ETS Standards for Quality and Fairness*
- ETS audit process for ensuring quality and fairness
- ETS Standards and examples
  - TOEFL™ iBT
  - TOEIC®
  - TOEIC Bridge™



# Standards for Educational and Psychological Testing

- Jointly developed by three professional organizations
  - American Educational Research Association (AERA)
  - American Psychological Association (APA)
  - National Council on Measurement in Education (NCME)
- Standards provide criteria by for evaluating tests, testing practices, and test use
  - Educational assessments
  - Psychological assessments



# Standards for Educational and Psychological Testing

## Brief history:

- 1954: *Technical Recommendations for Psychological Tests and Diagnostics Techniques* (APA)
- 1955: *Technical Recommendations for Achievement Tests* (AERA & NCME)
- 1966: *Standards for Educational and Psychological Tests and Manuals* (APA, AERA & NCME)
- 1974: *Standards for Educational and Psychological Tests* (AERA, APA & NCME)
- 1985: *Standards for Educational and Psychological Testing* (AERA, APA & NCME)



# Standards for Educational and Psychological Testing

- Current version of standards
  - Guided by the five earlier documents
  - *Standards for Educational and Psychological Testing* (AERA, APA & NCME, 1991)
- 1991 version undergoing revision
  - Committees representing AERA, APA, and NCME created
  - First draft issued
  - Comments from members of AERA, APA, and NCME gathered
  - Next draft underway

# Standards for Educational and Psychological Testing

- These standards provide *guidance only* – they do not specify
  - Actions that might result from test misuse
  - Actions that might result from failure of test publisher to follow standards
  - How standards can be applied and used operationally



# ETS Standards for Quality and Fairness

- ETS mission
  - Advance quality and equity in education by providing fair and valid assessments
- Why ETS Standards?
  - Reflect the specific nature of ETS product and service offerings
  - Define benchmarks that are used to develop tests, report scores, and maintain data



# ETS Standards for Quality and Fairness

- Developed and adopted as corporate policy by the ETS Board of Trustees in 1981
- Revised in 1987 and again in 2000
- Two sets of standards are comparable but the *ETS Standards* focus on the types of assessments and products that ETS offers
- *ETS Standards* are used by ETS staff on a daily basis





# ETS Quality and Fairness Process: Program Audit

- Program audit
  - ETS Office of Corporate Quality Assurance
  - Completed every three years
- Every assessment and product is evaluated for compliance with the *ETS Standards*



## **ETS Audit Process**

- (1) Program prepares documentation for audit
  - Template used to capture documentation for each standard
  - Documentation varies
    - Research reports
    - Statistical reports
    - Test user manuals
    - Program brochures and marketing materials
    - Score reports
    - Security procedures



## **ETS Audit Process**

### **(2) Audit teams formed**

- Selected from ETS professional staff
- Receive special training
- Reviews documentation and program practices for compliance with each standard, using professional judgment



## **ETS Audit Process**

### **(3) Audit results**

#### **– Action Plan**

- Program responds to “not sufficiently compliant” ratings
- Reviewed and approved by officers, Office of Corporate Quality Assurance, and audit team
- Progress on Action Plan reviewed annually

#### **– ETS Board of Trustees**

- Review audit findings



# **ETS Standards for Quality and Fairness**

- Thirteen chapters, some of which
  - Apply only to non-assessment products
  - Apply to all products
  - Apply primarily to assessments



# **ETS Standards for Quality and Fairness**

- **Chapter 1: Developmental Procedures**
  - Non-assessment products are developed and revised using planned, documented processes
- **Chapter 2: Suitability for Use**
  - Non-assessment products are capable of meeting their intended purpose for the intended audience



# ETS Standards for Quality and Fairness

- **Chapter 3: Customer Service**
  - Customer needs are met and high quality maintained
- **Chapter 4: Fairness**
  - All products treat all people equally and fairly
- **Chapter 5: Uses and Protection of Information**
  - All critical and confidential information is safeguarded



# ETS Standards for Quality and Fairness

- **Chapter 3: Customer Service**
  - Customer needs are met and high quality maintained
- **Chapter 4: Fairness**
  - All products treat all people equally and fairly
- **Chapter 5: Uses and Protection of Information**
  - All critical and confidential information is safeguarded



## Fairness Example: TOEFL

- **Standard 4.2:** Obtain and document judgmental and, when possible, empirical evidence of fairness for studied groups.
- **Auditors' Comments:** *Need to perform analyses for separate groups in the test taking population.*
- **Action Plan:** Program will request assistance from Research to determine which approach would be best suited for meeting the fairness-oriented needs of the program.
- **Result:** *New research plan includes studies specifically aimed at understanding fairness in the context of international test takers.*



# ETS Standards for Quality and Fairness

- **Chapter 6: Validity**
  - Appropriate evidence is gathered and documented to support the intended inferences of the assessment results
- **Chapter 7: Assessment Development**
  - Assessments are constructed using planned, documented processes
- **Chapter 8: Reliability**
  - Scores and assessment results are sufficiently reliable for their intended purposes

# ETS Standards for Quality and Fairness

- **Chapter 6: Validity**
  - Appropriate evidence is gathered and documented to support the intended inferences of the assessment results
- **Chapter 7: Assessment Development**
  - Assessments are constructed using planned, documented processes
- **Chapter 8: Reliability**
  - Scores and assessment results are sufficiently reliable for their intended purposes

## Validity Example: TOEIC Bridge

- **Standard 6.4:** Obtain and document the logical and/or empirical evidence that the assessment will meet its intended purpose.
- **Auditors' Comments:** *Reliability of one subscore is low. Either subscores should be dropped or warning given to guard against misinterpretations.*
- **Action Plan:** Program will update external materials
- **Result:** *Examinee Handbook now includes statement: "Subscores are less reliable than section scores...it is not recommended subscores be used in decision making."*



# **ETS Standards for Quality and Fairness**

- **Chapter 9: Cut Scores, Scaling, and Equating**
  - Score reporting scales are meaningful and procedures for setting cut scores or linking scores are rational and clearly described
- **Chapter 10: Assessment Administration**
  - Assessments are administered appropriately



# ETS Standards for Quality and Fairness

- **Chapter 9: Cut Scores, Scaling, and Equating**
  - Score reporting scales are meaningful and procedures for setting cut scores or linking scores are rational and clearly described
- **Chapter 10: Assessment Administration**
  - Assessments are administered appropriately



## Cut Score Example: TOEIC Bridge

- **Standard 9.3:** Describe the equating or other linking studies that were done in sufficient detail to allow knowledgeable people to evaluate and replicate the studies.
- **Auditors' Comments:** *Equating memo, attached to audit report, is outdated.*
- **Action Plan:** Program will update documentation.
- **Result:** *New equating memo provided to auditors.*



# ETS Standards for Quality and Fairness

- **Chapter 11: Reporting Assessment Results**
  - Accurate, understandable scores are provided
- **Chapter 12: Assessment Use**
  - Encouraging proper use of assessments
- **Chapter 13: Test Takers' Rights and Responsibilities**
  - Examinees rights and responsibilities during the assessment process





# ETS Standards for Quality and Fairness

- **Chapter 11: Reporting Assessment Results**
  - Accurate, understandable scores are provided
- **Chapter 12: Assessment Use**
  - Encouraging proper use of assessments
- **Chapter 13: Test Takers' Rights and Responsibilities**
  - Examinees rights and responsibilities during the assessment process



## Assessment Use Example: TOEIC

- **Standard 12.1:** Provide intended users of assessments with the information they need to evaluate the appropriateness of assessments.
- **Auditors' Comments:** *The TOEIC User Guide needs to be finalized. Ensure that the section titled "Appropriate Uses of TOEIC Speaking & Writing Scores" is easily accessible by score users and clients.*
- **Action Plan:** Program will update User Guide.
- **Result:** *Finalized User Guide containing user-friendly, accessible language published.*



- <http://www.ets.org/about/fairness/>