Technical Information Brief – Test Fairness and Sensitivity

As part of the development process for CASAS assessments approved by the Office of Career, Technical, and Adult Education (OCTAE) for use in the National Reporting System (NRS) for Adult Education, CASAS follows an extensive review procedure to ensure that all test items and test forms are fair and sensitive to the population groups for whom the tests are intended. This includes fairness and sensitivity training, panel reviews, and analyses of empirical data.

Introduction

Test fairness is a fundamental part of the validation process for all CASAS test items and test forms. This focuses on the fairness of measurement quality to ensure the lack of measurement bias with respect to all population groups who take CASAS tests.

Each CASAS test series is comprised of sets of parallel forms that are constructed so that two forms at each test level can be used independently of each other and be considered equivalent. For example, the CASAS Reading STEPS test series is comprised of 354 test items distributed across ten test forms. These ten test forms consist of two parallel forms at each of five test levels (A through E). All items have undergone a multistage review process for test fairness and sensitivity that conforms to best practices outlined in the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014) and the Educational Testing Service Guidelines for Fair Tests and Communications (ETS, 2022).

Initial Item Revision and Review

All test items that are initially accepted from item writers are thoroughly reviewed by CASAS assessment specialists to ensure they meet fairness and sensitivity requirements. Both the item writers and the CASAS assessment specialists complete fairness and sensitivity training to ensure they incorporate all current fairness and sensitivity principles into their item writing and review. The fairness and sensitivity training is designed based on the Educational Testing Service (ETS) guidelines.

Subject-Matter Expert Review Panels

All items that pass the internal review are then reviewed by panels comprised of Subject-Matter Experts (SMEs). The panelists are SMEs in the area of fairness and sensitivity in education. In addition, panelists are carefully selected to represent diverse demographic groups (age, ethnicity, gender), and expertise in adult literacy, language instruction, mathematics, and/or assessment.

Among other considerations, when reviewing a test or an item, panelists consider whether the item avoids construct-irrelevant KSA barriers to success, construct-irrelevant emotional barriers to success, and construct-irrelevant physical barriers (ETS, 2022).

Fairness and Sensitivity Review using Empirical Data

The fairness and sensitivity of all CASAS test items is further analyzed by examining each item for Differential Item Functioning (DIF). The purpose of the DIF analyses is to examine whether test items are biased against examinees from the different demographic subgroups who take CASAS tests. DIF is said to occur when equally able examinees differ in their probabilities of answering a test item correctly as a function of group membership (AERA et al., 2014).

The DIF analyses involve the comparison of item responses from the different demographic subgroups while controlling for their overall ability level. If an item shows DIF it may not be measuring the intended construct (i.e., reading, math, or listening) the same way across all groups.

CASAS uses a DIF analysis procedure which adapts the Mantel-Haenszel statistic to compare the performance on an item of a *focal group* to that of a *reference group* matched on overall proficiency or ability. The Mantel-Haenszel statistic provides an estimation based on the common odds ratio (a_{MH}) as shown in Table 1 and the following formula (MacInnes, 2009).

This procedure provides a cumulative statistic of the log odds ratio of passing or failing an item for the two groups (focal and reference) matched on overall proficiency or ability. This odds ratio is converted to a Delta scale based on procedures outlined in Holland & Thayer (1988). CASAS then uses a DIF categorization method developed by ETS (Zieky, 1993) to quantify the degree of DIF and further examine items for bias.

Table 1 Mantel Haenszel Procedure

	Response to item *I*		
	Correct	Incorrect	Total
Reference	$n_{1r_{ij}}$	$n_{0r_{ij}}$	$n_{.r_{ij}}$
Focal	$n_{1f_{ij}}$	$n_{1f_{ij}}$	$n_{.r_{ij}}$
Total	$n_{1.ij}$	$n_{0.ij}$	$n_{..r_{ij}}$

Where

$n_{1r_{ij}}$ is the number of subjects in the reference group at trait or ability level *j*, which answered item *i* correctly

$n_{0r_{ij}}$ is the number of subjects in the reference group at trait or ability level *j*, which answered item *i* incorrectly

$n_{1f_{ij}}$ is the number of subjects in the reference group at trait or ability level *j*, which answered item *i* correctly

$n_{0f_{ij}}$ is the number of subjects in the reference group at trait or ability level *j*, which answered item *i* incorrectly

The formula is as follows:

$$\hat{\alpha}_{MH} = \frac{\sum_{j=1}^{k} n_{1r_{ij}} n_{0f_{ij}} / n_{..ij}}{\sum_{j=1}^{k} n_{0r_{ij}} n_{1f_{ij}} / n_{..ij}},$$

A logarithmic transformation is applied to α for interpretive purposes. This transformation is expressed as:

$$\Delta_{MH} = -2.35 \ln(\alpha_{MH}).$$

Summary

As stated in the *Standards for Educational and Psychological Testing* (AERA et al., 2014), a fair test "reflects the same construct(s) for all test takers, and scores from it have the same meaning for all individuals in the intended population; a fair test does not advantage or disadvantage some individuals because of characteristics irrelevant to the intended construct.

The items contained on CASAS tests have undergone a multistage review process consisting of a Subject-Matter Expert review and an examination of empirical data using Differential Item Functioning analyses. This internal and external review process ensures that all items contained on the CASAS test forms have been determined to be fair and sensitive to the diverse student populations served by adult education programs.

References

American Educational Research Association, American Psychological Association & National Council on Measurement in Education (2014). Standards for Educational and Psychological Testing. Washington, D.C.

Educational Testing Service. (2015). *ETS guidelines for fair tests and communications.* Princeton, NJ: Author.

Holland, P. W., & Thayer, D. T. (1988). Differential item functioning and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), Test validity (pp. 129–145). Hillsdale, NJ: Erlbaum.

MacInnes J.M. (2009). The Mantel-Haenszel Method for Detecting Differential Item Functioning in Dichotomously Scored Items: A Multilevel Approach.

Zekey, M. (1993). Practical questions in the use of DIF statistics in test development. In P. W. Holland & H. Wainer (Eds.), Differential item functioning (pp. 337–347). Lawrence Erlbaum Associates, Inc.

