ICT Literacy Assessment Technical Manual

Table of Contents

Preface

Purpose of This Manual

The purpose of the ICT literacy Technical Manual is to explain:

- The purpose of the ICT literacy test;
- How to use the ICT literacy test;
- The approach that was taken in developing ICT literacy tests;
- The validity evidence supporting ICT literacy test score use;
- The statistical processes supporting the psychometric quality of the ICT literacy assessment;
- The score reporting process.

Audience

This manual was written for technical experts, educators, and policy-makers and who are:

- Interested in knowing more about the ICT literacy program;
- Interested in how ICT literacy assessment relates to school programs;
- Interested in understanding how ICT literacy assessment is developed and scored;
- Interested in the statistical characteristics of ICT literacy assessment.

Purpose of the ICT Literacy Assessment Overview

The ICT literacy assessment supports decisions about individual students. The target population for the assessment is the 9th grade students. Students taking the assessment should function in a language that assessment is administered at the 10th grade level or higher and possess basic technical competence, such as the ability to use a mouse and keyboard.

Appropriate Use of Assessment Results: Low- to Medium-Stakes Decision-Making
Assessment scores may inform several types of decisions for individuals as well as policy and administrative decision-making based on appropriate aggregations of examinees. Potentially appropriate uses for the assessment include:

- **Informing policy makers.** By appropriately measuring what 9th grade students know and can do; and to measure this as a whole and subgroups – to see how subgroups of students as defined by various background and contextual characteristics performed over time.
- **Coursework or training guidance.** The assessment provides information that can contribute to helping guide 9th grade students in their learning of ICT literacy skills.
- **Fundamental ICT Literacy instruction.** The assessment scores may contribute to the identification of individuals who would benefit from basic ICT Literacy training or should be exempt from otherwise required training.
- **Resource allocations.** By appropriately summarizing examinees scores, administrative decision-makers can collect evidence that may reflect how many individuals need fundamental ICT Literacy training at their location.
- **Training evaluation.** Decision-makers can examine assessment results as part of an evaluation program to estimate the efficacy of ICT literacy training.

These potential uses require validation. Current research seeks to validate these uses of assessment results. In addition, we advise institutions to validate their own use of assessment results for their own institution and examinees.

Inappropriate Use of Assessment Results

- **High-stakes decision-making.** ICT literacy assessment scores should not be used to make decisions that have severe consequences for individuals, such as employment, graduation requirements, requirements for entering a major, or entrance requirements for an institution.
- **Technical certification.** The ICT literacy assessment focuses on problem-solving and critical thinking skills in the context of technology. No inferences should be made about test takers' ability to use specific software products.

Assessment Development

Fairness in ICT Test Development

ICT assessment is created to be of the highest quality and as free from bias as possible and ICT literacy assessment – including individual test items, tests, instructional materials, and publications – are evaluated during development, according to the industry standards (i.e. *ETS Fairness Review Guidelines)*, so that they are not offensive or controversial; do not reinforce stereotypical views of any group; are free of racial, ethnic, gender, socioeconomic, or other forms of bias; and are free of content believed to be inappropriate or derogatory toward any group

ICT Test Development Standards

During the ICT literacy test development process, the program follows the strict guidelines detailed in the industry standards (i.e. *Standards for Educational and Psychological Testing)*:

- Define clearly the purpose of the test and the claims one wants to make about the test takers;
- Develop and conduct content validation surveys to confirm domains of knowledge to be tested;
- Develop test specifications and test blueprints consistent with the purpose of the test and the domains of knowledge defined by the content validation;
- Develop specifications for task and item types and numbers of tasks and items needed to adequately sample the domains of interest validated by the content analysis survey;
- Develop test tasks and items that provide evidence of the measurable-behavior indicators detailed in the test specifications;
- Review tasks and items and assembled test forms so that each item has a single best defensible answer and assesses content that is relevant;
- Review test tasks and items and assembled forms for potential fairness or bias concerns, overlap, and cueing, revising or replacing items as needed to meet standards.

Validity

A test is developed to fulfill one or more intended uses. The reason for developing a test is fueled, in part, by the expectation that the test will provide information about the test taker's knowledge and/or skill that:

- May not be readily available from other sources;
- May be too difficult or expensive to obtain from other sources;
- May not be determined as accurately or equitably from other sources.

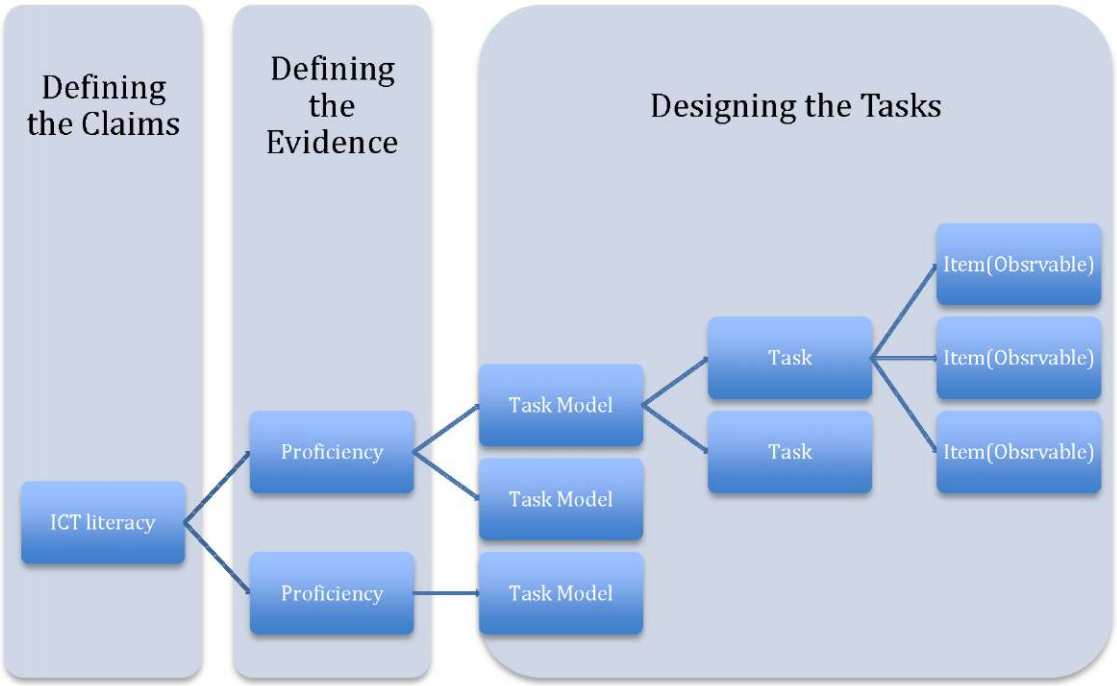
But regardless of why a test is developed, evidence must show that the test measures what it was intended to measure and that the meaning and interpretation of the test scores are consistent with each intended use. Herein lays the basic concept of validity: the

degree to which evidence (rational, logical, and/or empirical) supports the intended interpretation of test scores for the proposed purpose (*Standards for Educational and Psychological Testing, AERA, APA, NCME, 1999*).

The main source of validity evidence for ICT literacy comes from the alignment between what the profession defines as skills important to be ICT literate and effective practice and the content included on the test (*Standards for Educational and Psychological Testing, 1999*).

Within the test development cycle, the tasks and items in the ICT literacy assessment are developed using an evidence-centered design process (ECD) that adds to the validity of the tests. Evidence-centered design is a construct-centered approach to developing tests begins by identifying the knowledge and skills to be assessed through a job/content analysis. Building on this information, test developers then work asking what factors would reveal those constructs and, finally, what tasks elicit those behaviors. This design framework, by its very nature, makes clear the relationships among the inferences that the assessor wants to make, the knowledge and behaviors that need to be observed to provide evidence for those inferences, and the features of situations or tasks that evoke that evidence. Thus, the nature of the construct guides not only the selection or construction of relevant tasks and items but also the development of scoring criteria and rubrics. In sum, test items follow these three ECD stages: (a) defining the claims to be made, (b) defining the evidence to be collected, and (c) designing the tasks to be administered (see Figure 1).

Figure 1: ECD Validity Chain

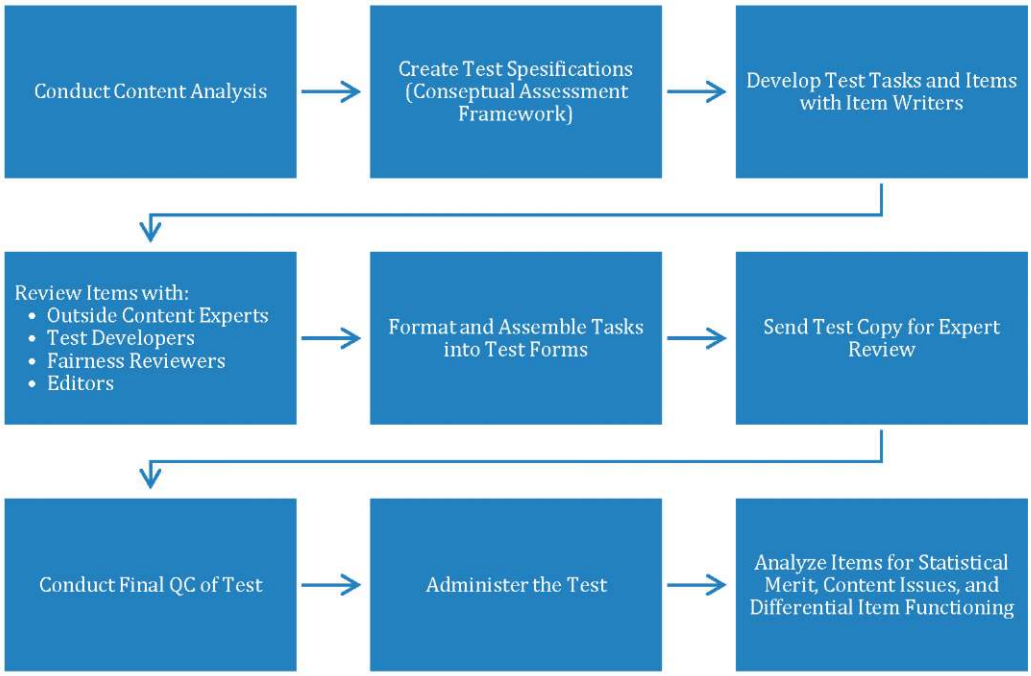


Test Development Process

The ICT literacy test and related materials follow a rigorous development process, as outlined below and in Figure 2:

- Research national standards and curricula to verify alignment with the claims made for the test and the test takers;
- Develop the content (job) analysis claims;
- Conduct content validation survey;
- Develop test specifications and blueprints, using the results of the content (job) analysis survey;
- Recruit and train item developers to write tasks and items (observables. See Figure 1) for the test;
- Develop sufficient numbers of tasks and items to form a pool from which parallel forms can be assembled;
- Review the tasks and items developed by trained writers, applying the industry standards (i.e. *ETS Standards for Fairness and Quality)* and editorial guidelines;
- Task and item reviews also are done by practitioners in the field who may not be trained writers but who have the content expertise to judge the accuracy of the items;
- Prepare the approved tasks for administration and assemble them into operational forms;
- Send assembled test(s) to appropriate content experts for a final validation of the match to specifications, and accuracy of the correct response;
- Perform final quality-control checks, according to the program's standard operating procedures;
- Administer a pilot test;
- Analyze and review test data from the pilot or first administration to verify that items are functioning as intended and present no concerns about the intended answers or impact on subgroups.

Figure 2: Test Development Process



Create Test Specifications

Test specifications are documents that inform stakeholders of the essential features of tests. These features include:

- A statement of the purpose of the test and a description of the test takers;
- The major categories of ICT proficiency or skills covered by the test and a description of the specific ICT proficiency or skills that define each category; the proportion that each major category contributes to the overall test; and the length of the test;
- The kinds of tasks and items (observables) on the test;
- How the test will comply with the industry standards (i.e. *ETS Standards for Fairness and Quality).*

The test specifications also are used to direct the work of item writers by providing explicit guidelines about the types of tasks and items needed and the specific ICT literacy proficiency that each item needs to measure.

Develop Tasks and Test Items

Experts are recruited to develop tasks and test items. The experts are people who know the domains of knowledge to be tested and are adept at using the complexities and nuances of language to write tasks and items at various difficulty levels. They write tasks and items that match the behavioral objectives stated in the test specifications and their items are written to provide sufficient evidence that the test taker is competent to begin practice.

Review Tasks and Items (by Test Developers, Fairness Reviewers, Editors)
The review of tasks and items is an essential step in the validity chain of evidence required by good test development practice. All tasks and items for use on the ICT literacy test are vetted for importance and relevance for match to specifications and correctness of intended response.

In addition all tasks and items used on a test are reviewed by fairness reviewers and editors. Changes to items are documented and discussed often before the final version is used on a test.

Format and Assemble Tasks into Test Forms
When tasks are ready to be used in a test form, they are formatted. Formatted tasks selected to be used in a form are assembled and automatically laid out for computer based testing (CBT) administration. The computer test layouts are modified as needed and checked for quality. These layouts are then checked for quality by test developers. Changes to each version of the computer layout are documented in an electronic record.

Send Test Copy for Expert Review
Before a test is signed off by test developers, it receives a content review to verify that every item has a single best answer, which can be defended. The reviewer must understand the purpose of the test and be prepared to challenge the use of any task and item that is not important to test or is not a match to the test specifications.

Conduct Final QC of Test
When the reviews of a particular test form have been examined, test developers perform multiple checks of the reviewers' keys against the official key. They must revise the layout, if necessary, and document the revisions. They must certify the test as ready for delivering. They do this by initialing each page of an electronic copy of the most recent version, and by signing and dating certain pages of the copy. The test coordinator then checks that all necessary steps have been followed and signs off on the test copy that will be delivered.

Analyze Items for Statistical Merit, Content Issues, and Differential Item Functioning
In the week following an administration, test developers receive the measurement statistician's preliminary tasks and item analysis (PIA). In addition to item analysis PIA output contains a list of flagged tasks and items that test developers must examine to verify that each has a single best answer. Test developers are agreeing whether to score (or not to score) the tasks and items in a standard report prepared by the statisticians. Test developers must provide a rationale for the best answer to each flagged item as well as an explanation as to why certain flagged distracters are not keys.

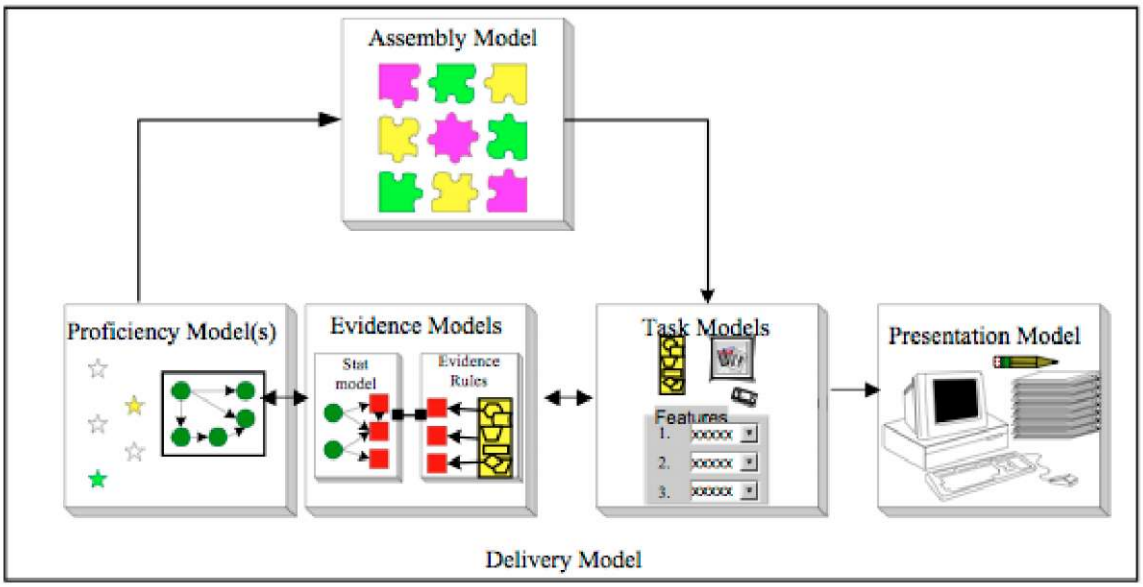
If it is decided not score a task or an item, official item keys are updated, tasks and items in databases are revised or deactivated, and the number of tasks and items used to compute and report scores are adjusted.

If there is sufficient test taker volume, Differential Item Functioning (DIF) analyses are run on a new ICT literacy test form to determine if subgroup differences in performance may be due to factors other than the abilities the test is intended to measure.

ICT Literacy Test Conceptual Assessment Framework (CAF) and Model Specifications

Since the ICT literacy assessment has a complex structure, the team used Evidence Centered Design (ECD: Mislevy, Almond, 2003) to organize the development. The Conceptual Assessment Framework (CAF) concerns technical specifications of ECD, that is, the materials and processes that embody ICT literacy assessments showed in Figure 3 below.

Figure 3: The Central Models of the Conceptual Assessment Framework (CAF)

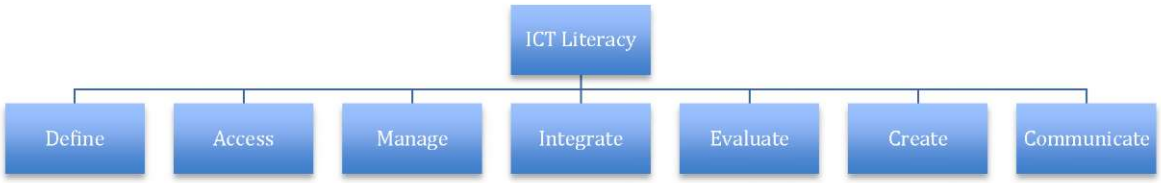


Below we will be describing the general **Assembly Model** with some elements of **Task Models**. Following these descriptions we will introduce the development of the **Proficiency Model**, and finally – development of **Evidence Model** for selected tasks and the base for prior distribution used.

Assembly model

Aligned with the nationally recognized *Association of Colleges and Research Libraries (ACRL) standards*, the ICT literacy assessment measures information literacy through seven proficiencies – Define, Access, Evaluate, Manage, Integrate, Create and Communicate – representing a range of ways that students handle information through digital technology (see Figure 4).

Figure 4: Proficiency Model



The organizational scheme, shown below in Figure 5, comprises of five components we view as essential for functioning in a knowledge society. It also includes the foundational set of skills and knowledge that underlie ICT literacy: cognitive and technical proficiency

Figure 5: Organizational Framework of ICT

Processes	Proficiencies Increasing complexity of knowledge and expertise →						
	Define	Access	Manage	Integrate	Evaluate	Create	Communicate
Cognitive							
Technical							
Ethical							

In Figure 6: Three processes are defined.

Figure 6: ICT Processes

Processes	Definition
Cognitive	The desired foundational skills of everyday life at school, at home, and at work. Literacy, numeracy, problem solving, and spatial/visual literacy demonstrate these proficiencies.
Technical	The basic components of digital literacy. It includes a foundational knowledge of hardware, software applications, networks, and elements of digital technology.
Ethical	It represents ethical and legal access and use of information, for example selecting a license agreement before downloading software.

Definitions of the seven proficiencies follow along with some example questions shown in Table 1.

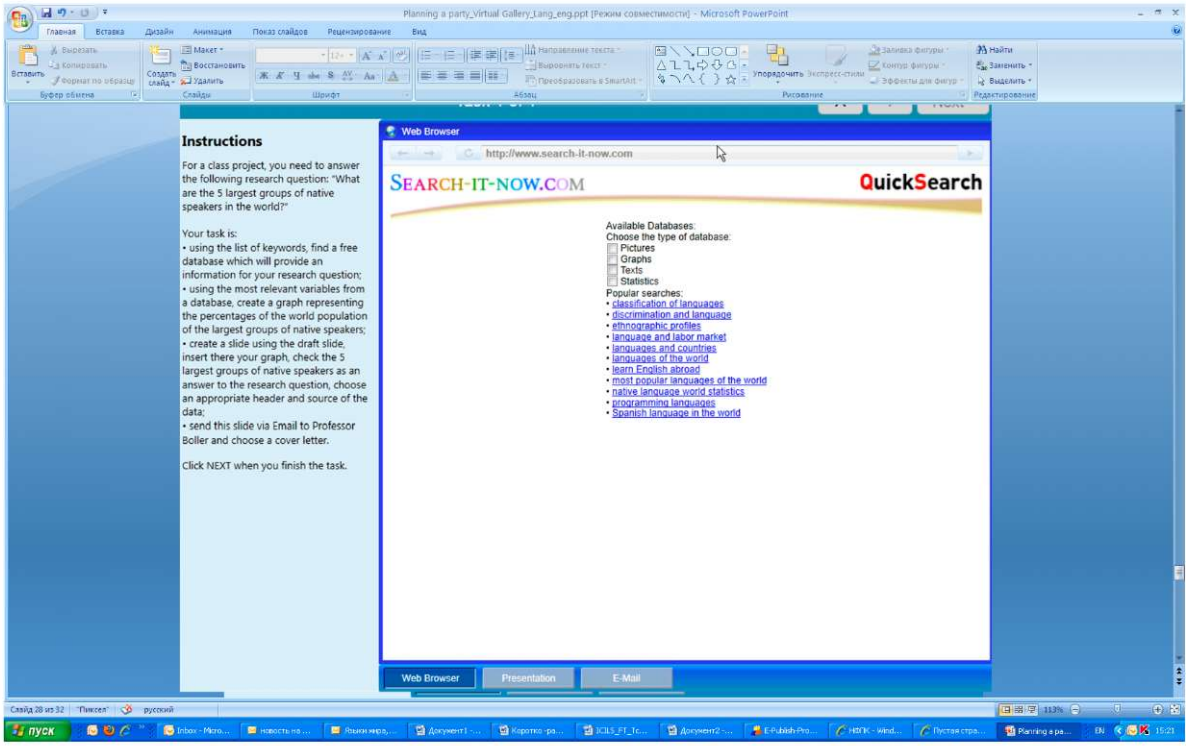
Table 1: ICT Proficiencies

Proficiencies	Description	Task Example
Define	Understand and articulate the scope of an information problem in order to facilitate the electronic search for information by: ✓ Distinguishing a clear, concise and topical research	

Proficiencies	Description	Task Example
	question from poorly framed questions, such as ones that are overly broad or do not otherwise fulfill the information need ✓ Asking questions of a "teacher" that help disambiguate a vague research assignment ✓ Conducting effective preliminary information searches to help frame a research statement	
Access	Collect and/or retrieve information in digital environments. Information sources might be web pages, databases, discussion groups, email or online descriptions of print media. Tasks include: ✓ Generating and combining search terms (keywords) to satisfy the requirements of a particular research task ✓ Efficiently browsing one or more resources to locate pertinent information ✓ Deciding what types of resources might yield the most useful information for a particular need	
Evaluate	Judge whether information satisfies an information problem by determining authority, bias, timeliness, relevance and other aspects of materials. Tasks include: ✓ Judging the relative usefulness of provided web pages and online journal articles ✓ Evaluating whether a database contains appropriately current and pertinent information ✓ Deciding the extent to which a collection of resources sufficiently covers a research area	
Manage	Organize information to help you or others find it later by: ✓ Categorizing emails into appropriate folders based on a critical view of the emails' contents ✓ Arranging personnel information into an organizational chart ✓ Sorting files, emails or database returns to clarify clusters of related information	
Integrate	Interpret and represent information using digital tools to synthesize, summarize, compare and contrast information from multiple sources. Tasks include: ✓ Comparing advertisements, emails or websites from competing sources by summarizing information into a table ✓ Incorporating information from different sources to conduct a scientific experiment and report the results ✓ Placing results from an academic or sports tournament into a spreadsheet to clarify standings and decide the need for playoffs	
Create	Adapt, apply, design or construct information in digital environments by: ✓ Editing and formatting a document according to a set of editorial specifications ✓ Creating a presentation slide to support a position on a controversial topic	

Proficiencies	Description	Task Example
	✓ Creating a data display to clarify the relationship between academic and economic variables	
Communicate	Disseminate information tailored to a particular audience in an effective digital format by: ✓ Formatting a document to make it more useful to a particular group ✓ Transforming an email into a succinct presentation to meet an audience's needs ✓ Selecting and organizing slides for distinct presentations to different audiences ✓ Designing a flyer to advertise to a distinct group of users	See Fig. 7

Figure 7: Example of Communicate Task Scenario Screen



The content areas and the technology topics measured are shown in Table 2.

Table 2: Content and Technology Specifications

Content areas (number of tasks)	Technology Topics (number of tasks)
Academic (6-8): ▪ Humanities ▪ Social Sciences ▪ Natural Sciences	Web Use: ▪ Email(1-2) ▪ Instant messaging(1-2) ▪ Bulletin board postings(1-2) ▪ Browser use (2-4) ▪ Search engines(3-5)
Business (4-6):	

▪ Practical Affairs Personal (4-6): ▪ Popular Culture	Database Management: ▪ Data searches(1-2), ▪ File management(1-2) Software ▪ Word processing (2-4) ▪ Spreadsheet (1-2) ▪ Presentations (1-2), ▪ Graphics (1-2)

A test form consists of tasks drown from 16 task model as shown in Table 3 below.

Table 3: Example of Test Blueprint for ICT Literacy Assessment for Task Model Complexity. Entries are labeled with difficulty targets and number of observables.

Task Model	Define	Access	Manage	Integrate	Evaluate	Create	Communicate	Observables
L1	EASY(3)							3
L2		EASY(1)						1
L 3			EASY(3)					3
L 4				EASY(1)				1
L5					EASY(1)			1
L6						EASY(3)		3
L7							EASY(1)	1
L8	HARD(3)							3
L9		HARD(1)						1
L10			HARD(3)					3
L11				HARD(3)				3
L12					HARD(1)			3
L13							EASY(2)	2
M1		HARD(2)			EASY(2)			4
M2				EASY(2)		HARD(2)		5
H1		EASY(3)			HARD(2)	HARD(1)	HARD(3)	9
Observables	6	7	6	6	6	6	6	Total: 43

Task Models L1-L13 are "low" complexity task, which would take each approximately 4 minutes to complete and target one proficiency each. Task Models M1 and M2 are "medium" complexity task, which would each take 15 minutes each to complete and target two proficiencies each. Task Model H1 is "high" complexity task, which each would take 30 minutes to complete and target four proficiencies. See Table 4 below for details.

Here HARD and EASY refer to the difficulty (amount of skills needed to solve tasks). The first 13 task models are for low complexity tasks, and only medium and high

complexity tasks target multiple proficiencies. Each task would a mini-simulation, which would yield multiple observed outcome variables (observables).

Table 4: Task Models Complexity and Target Proficiencies Specifications

Task Complexity	Time per task in minutes	Number of tasks in a test form	Number of proficiencies targeted per task	Description of targeted proficiencies (number of observables)
High	30	1	4	Access(1-3) Evaluate (1-3) Create (1-3) Communicate(1-3)
Medium	15	2	2	Access(1-3) Evaluate(1-3); Integrate(1-3) Create (1-3);
Low	4	13	1	Define(1-3) Access(1-3) Evaluate (1-3) Integrate (1-3) Manage (1-3) Create (1-3)

A test form that consists of technological topics is shown in Table 5 below.

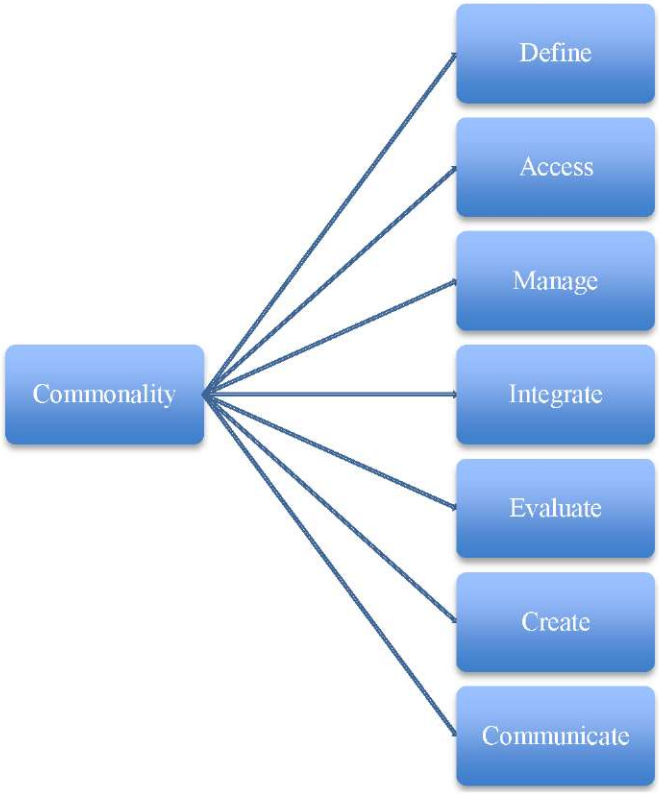
Table 5: Example of Test Blueprint for ICT Literacy Assessment for Technology Topics

Technology Topics	Define	Access	Manage	Integrate	Evaluate	Create	Communicate
E-mail							Task 1
Text editor				Task 2			
Spreadsheet							
Social network	Task 3						
Internet browser							
Other							

Proficiency Model

In the ICT literacy assessment, the 7 proficiencies were modeled as independent given overall ICT literacy. To increase the reliability of overall ICT proficiency estimation, the ICT literacy node, see Fig 4, was replaced with a Commonality node, see Figure 5. The Commonality was placed in the model to soak up the correlation between the seven proficiencies, but would not be used in reporting. Instead, the overall score would be based on a functional composite of scores for the seven proficiencies. This type of scoring would have the Law of Averages helping to decrease the variance.

Figure 5: Seven proficiency (P7) Bayesian Network



The seven proficiency variables can each take on three states in the model: Below Basic, Basic, and Above Basic. The Commonality variable also assigned three states.

The graphical structure shown in Fig. 5 is only part of the picture. To complete Bayesian Network conditional probability tables were introduced for each variable in the model given its parents in the graph, as network requires. Using these tables Commonality node was described having a multinomial distribution, and its conjugate prior is a Dirichlet distribution. The priors for the conditional probability tables consist of independent Dirichlet distribution for each row of the table. We used prior distribution that could be found in the report (Katz et, al., 2004). One of the distribution tables is shown in Table 6.

Table 6: Commonality Variable Distribution

Table	Create & Communicate		
Commonality	Above Basic	Basic	Below Basic
Above Basic	6.5	3.3	0.2
Basic	1.7	5.9	2.3
Below Basic	0.1	2.6	7.3

Standard-Setting Process

ICT literacy assessment consists of approximately 60 items (observables) derived from performance on 16 interactive, performance-based tasks. The scoring of the items follow rubrics that specify the nature of responses needed to gain High (1), Medium (0.5), or Low (0). The overall raw score on the assessment is the sum of all item scores.

Eighteen panelists, each with direct experience with the information literacy of high school, were divided into two subpanels. These subpanels independently made standard-setting judgments for each item in a test. On the final day of the study, the panel reconvened to reconcile any differences in their respective recommended benchmarks.

The standard-setting study followed a modified, extended Angoff procedure. Each panelist first independently judged the average score that would be received on each of the items by 100 qualified candidates. Members of each subpanel discussed their Round 1 judgments, and this process continued for a second round of judgments and discussion, followed by a third round of judgments. Task-level data were introduced during the Round 1 discussion and classification outcomes (percentage of examinees meeting or exceeding the preliminary cut scores) were presented during the Round 2 discussion. The benchmarks derived from the Round 3 judgments were reconciled in full panel discussions. Recommended by the panel a benchmark profile required for ICT literacy assessment shown in Table 7.

Table 7: Profiles Required for Proficiency Classifications

	Access	Manage	Communicate	Create	Define	Integrate	Evaluate
Advanced	P	P	P	P	P	P	P
Above Basic	P	P	P	P	P	A	A
Basic	A	A	A	A	A	A	A
Below Basic	A	A	A	A	A	U	U
Developing	U	U	U	U	U	U	U

P = Proficient, A = Acceptable; U = Unacceptable

Note: *The proficiencies are listed left to right from easiest to hardest.*

Evidence and Link Models

In the ECD framework the link has two parts: a) the evidence rules – which define how the work product are processed to produce the values of observable outcome variables and b) statistical model – which provides conditional probability distribution of the observed variable given proficiency state.

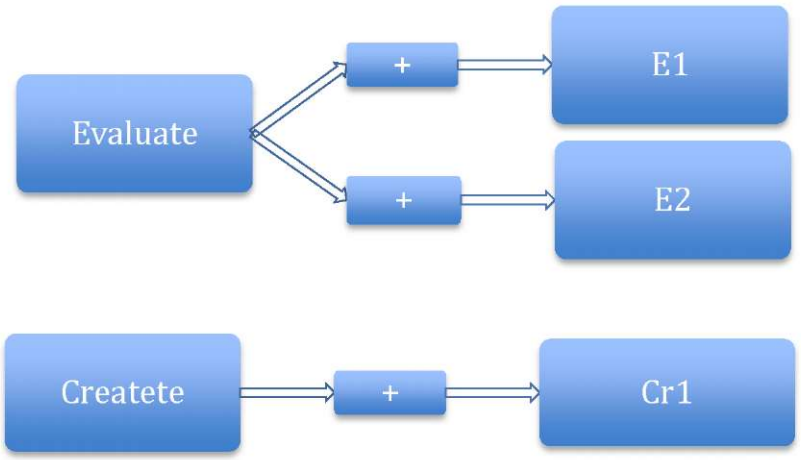
The rules of evidences for observable outcome variables in the link "degree of fulfilment of the information need" are shown in Table 8 below.

Table 8: Evidence Rules

Item (Observable Outcome)	Values
Degree of fulfilment of the information need	High: All fields in the slide contain at least something AND heading or source fields are empty, the graph is inserted Medium: The graph is inserted AND heading/source field is empty and 5 languages are selected OR heading and source fields are not empty but at least 1 language is not selected Low: all other cases

We will call the task specific version of the evidence models, link models. Evidence model defines a graphical structure related to observables. Figure 6 below shows an evidence model for a medium complexity task. It has three observable variables: two (E1 and E2) targeting the Evaluate proficiency and one (Cr1) targeting the Create proficiency. Bayesian network notations allow for modeling local dependence among observables using latent Context variable. This capability was not used this version of the test.

Figure 6: Evidence Model Evaluate and Create



The box with the plus sign in Figure 6 indicates the Compensatory distribution that is used to model the relationship between proficiency variable and observables. This is one of the Di Bello-Samejima distributions. It specifies the way for defining conditional probability network that drives Bayesian network with fewer parameters than the conditional multinomial distribution implies. Thus each box with the plus sign represents the conditional probability distribution for one of the observables given proficiencies Evaluate and Create. Since each both proficiency and observable variables each have

three levels, a conditional probability table is three by three table were each row represents the conditional probability of the observable given a certain proficiency (i.e., Evaluate). The Di Bello-Samejima distribution allows to assign each row of the table effectiveθ. Therefore we can use standards Item Response Theory (IRT) models, such as Samejima graded response, to fill out the probability table. In particular, the probability for each sell of the table could be represented by

$$\Pr[Y_{jk}|\theta] = log\ it^{-1}(a_j\theta - b_j - d_{jk}) \quad (1)$$

Where j is an index over observable, k is an index over states of observable, a_j is the discrimination parameter, b_j is the intercept parameter, d_{jk} is the level difficulty parameter.

The values in cells are found by taking differences of this equation for the possible states of the observable value y_k .

$$\Pr[Y_{ij}|\theta] = Pr(Y_{ij} \geqslant k|\Theta) - Pr(Y_{ij} \geqslant k+1|\Theta) \quad (2)$$

In order to make three rows of the table, the effective theta values for proficiencies were chosen based on equally spaced quintiles of a unit normal distribution. These values are Above Basic = 0.9674, Basic = 0, and Below Basic = -0.9674. For the set scale of θ metric value of -1 for d indicates an item that is somewhat easy for the examinees, 0 a typical item, and +1 a somewhat difficult item; further, a parameters typically range from about 0.3 to 3.

Scale Anchors

The sore report for ICT literacy assessment called for reporting the percentage of students at each of several proficiency levels for each proficiency variable. For example, the score reports supported inferences about which proficiency should be emphasized as needed improvements. To draw these types of inferences, the proficiency levels were anchored to ECD claims or to tasks that support those claims.

The set of anchors for a given proficiency all tap that proficiency and reasonably balanced with respect to difficulty so that the average difficulty is zero. The distribution of parameters for those observables was constrained so

$$\sum_j b_j = 0 \text{ and } \sum_j loga_j = 0; \quad (3)$$

Not all the observable parameters must be constrain, only a subset. Thus subset must be reasonably balanced with respect to Hard and Easy observables and should be representative of the domain.

The anchor set constraints are only needed the first time the ICT literacy assessment is calibrated. In the future administrations, the population distribution can be linked by including a set of common tasks in both administrations. This is a Bayesian version of Non –Equivalent groups Anchor test.

Initial Priors and Posteriors

To quantify the amount of information about the various parameters, the parameter-summary tables indicate a percentage increase in precision from priors to posteriors. It is calculated as follows:

$$\%\, increase\ in\ presision = 100 \times \frac{(posterio\ SD)^{-2} - (prioCD)^{-2}}{(prioCD)^{-2}}$$

A value of zero would indicate no new information, while a value of 100 would mean there was twice as much information about a parameter after seeing the data than before seeing it.

Psychometric Properties

Introduction

Statistical Analysis division has developed procedures designed to support the development of valid and reliable test scores for the ICT literacy. The item (observables) and test statistics are produced by both classical and Item Response Theory (IRT) analyses.

The psychometric procedures explained in this section follow well-established, relevant standards in *Standards for Educational and Psychological Testing (1999)* and the *ETS Standards for Quality and Fairness (2002)*. They are used extensively in the ICT literacy test and are accepted by the psychometric community at large.

As discussed in the CAF section, the ICT literacy test has a set of test specifications that is used to create versions of a test, called test forms. Each test form has a unique combination of individual tasks and items (observables). The data for the psychometric procedures described below are the test taker item (observable) responses collected when the test form is administered, most often by using the item responses from the first use of a test form.

Test-Scoring Process

It is described in details in the CAF section.

Item Analyses

Classical Item Analyses

Following the administration of a new test form, but before scores are reported, a PIA for all items (observables) is carried out to provide information to assist content experts and test developers in their review of the tasks and items. They inspect each item, using the item statistics to detect possible ambiguities in the way the items were written, keying errors, or other flaws. Tasks and items that do not meet quality standards can be excluded from scoring before the test scores are reported.

Preliminary and final analyses include both graphical and numerical information to provide a comprehensive visual impression of how an item is performing. These data are subsequently sent to test developers, who retain them for future reference.
Items are flagged for reasons including but not limited to:

- Low average item scores (very difficult items);
- Low correlations with the criterion;
- Possible double keys;
- Possible incorrect keys.

Speededness

Occasionally, a test taker may not attempt tasks and items near the end of a test because the time limit expires before she/he can reach the final items. The extent to which this occurs on a test is called – speededness. The ICT literacy test assesses speededness using four different indices:

- The percent of test takers who complete all items;
- The percent of test takers who complete 75 percent of the items;
- The number of items reached by 80 percent of test takers;
- The variance index of speededness (i.e., the ratio of not-reached variance to total score variance).

All four of these indices need not be met for a test to be considered speeded. If the statistics show that many test takers did not reach several of the items, this information can be interpreted as strong evidence that the test (or a section of a test) was speeded. However, even if all or nearly all of the test takers reached all or nearly all of the tasks and/or items, it would be wrong to conclude, without additional information, that the test (or section) was unspeeded. Some test takers might well have answered more of the items correctly if given more time. Item statistics, such as the percent correct and the item total correlation, may help to determine whether many test takers are guessing, but the statistics could indicate that the items at the end of the test are difficult.

Test-Form Equating and Scaling

Overview

ETS Standards for Quality and Fairness (2002) require the use of equating methodologies when —" results ... on different forms of an assessment are to be treated as though they were equivalent", as is the case for ICT literacy tests. The equating and scaling procedure is described in CAF section.

Test Statistics

Reliability

The reliability of a test refers to the extent to which test scores are consistent or stable. An index of reliability enables to generalize beyond the specific collection of items in a particular form of a test to a larger universe consisting of all possible items that could be posed to the test taker. Because tests consist of only a sample of all possible items, any estimate of a test taker's actual capabilities will contain some amount of error. Psychometrically, reliability may be defined as the proportion of the test score variance that is due to the "true" (i.e., stable or non-random) abilities of the test takers. A person's actual (or "observed") test score may thus be thought of as having a "true" component and an "error" component. Here, "error" is defined as the difference between the observed and true scores. Since true scores can never be known, the reliability of a set of test scores can not be assessed directly, but only estimated.

Reliability estimates for ICT literacy and equating scores are computed using the Kuder and Richardson (1937) formula 20 (KR 20). Reliability may be thought of as the proportion of item score variance that is due to true differences among the test takers with respect to the ability being measured:

$$reliability = 1 - \frac{error\ vriance}{total\ variance}$$

If the test is not highly speeded, the KR 20 reliability estimate will be an adequate estimate of alternate-form reliability.

Standard Error of Measurement

The standard error of measurement (SEM) is an estimate of the standard deviation of the distribution of observed scores around a theoretical true score. The SEM can be interpreted as an index of expected variation if the same test taker could be tested repeatedly on different forms of the same test without benefiting from practice or being hampered by fatigue. The SEM of a raw score is computed from the reliability estimate (r_x) and the standard deviation (CD_x) of the scores by the formula:

$$SEM_x = CD_x\sqrt{1 - r_x}$$

The standard error of measurement for the scaled score is:

$$SEM_S = A \times SEM_X \ ,$$

where A is the score conversion coefficient used in the scaled score conversion equation:

$$Scale\ Score = A \times (raw\ score) + B \ ,$$

When the raw-to-scaled score conversion for a test form is nonlinear, the A parameter is estimated using the ratio of the scaled score standard deviation to the raw score standard deviation.

Score Reporting

Score reporting is the process in which tests are graded and test results are reported to test takers and institution. ICT literacy assessment measures and reports student performance in two ways: norm referencing and criterion referencing.

Norm referencing: How students in one country or region perform when compare to another country or region.

Criterion referencing: How students perform against proficiency bands for each ICT competence.

In addition it also measure student performance over time.

The individual student reports include information about achieved ICT literacy levels, each ICT proficiency levels, and formative feedback.

The institutional reports include frequency distributions of test scores broken down by background variables (i.e., gender, ethnicity, educational level, etc.).