# Feasibility Study for the PISA ICT Literacy Assessment

## Report to Network A

Marylou Lennon, Irwin Kirsch, Matthias Von Davier, Michael Wagner, Kentaro Yamamoto

October 2003





## **TABLE OF CONTENTS**

ACKNOWLEDGEMENTS	5
SECTION 1: INTRODUCTION	6
Overview	6
Work of the ICT Expert Panel	7
Goals for the Feasibility Study	
Summary of General Conclusions	
SECTION 2: PROCEDURES	
Development of the Study Materials	13
General Technical Procedures	
General Translation/Adaptation Procedures	15
Developing the Assessment Tasks	
Section 1 – The Introductory Tasks	
Response Modes for the Background Questionnaire	
The Background Questionnaire	
Basic Technical Tasks	
Section 2 – The Short Scenarios E-mail Task	
Web Abstract Task	
Database Task	
Section 3 – The Web Search Task	
Part 1	
Part 2	
Section 4 – The Simulation Task	
Developing the Follow-up Questionnaire	27
Developing the Observation/Retrospective Interview Materials	28
Developing the Administrators' Manual	
Recruiting and Selecting Students	30
General Sampling Plan	30
Retrospective Interviews	31
Data Collection	
Operational Specifications	
Testing Procedures	32
SECTION 3: RESULTS	34
Introduction	34
Overview of the Analysis	
Cautionary Note	35
Characterising the Student Sample	35



Characterising Task Performance	38
Access and Manage Processes	39
Sections by Difficulty and Time	
Introduction	
Difficulty and Timing Data	
Section 1 – Basic Technical Tasks	
Section 2 – Short Scenarios	
The E-mail Task	
The Web Abstract Task	
The Database Task	
Section 3 – Web Search Task	
Section 4 – Simulation	
Characterising Student Performance	50
Introduction	
Background Characteristics and Performance	
Gender	
Computer Familiarity	
Observed ICT Proficiency	52
The Follow-up Questionnaire - Student Responses	55
E-Mail Task	
Database Task	56
Web Search Task	56
Simulation Task	57
Summary of Results	57
Curimary of Nocure	
SECTION 4: LESSONS LEARNED	59
Simulated Applications and Environments	50
Capturing Process Information	59 50
Translation and Adaptation	
Usability	
Evidence	
Conclusion	
APPENDIX A: THE ICT EXPERT PANEL	62
APPENDIX B: SAMPLE DATA FILE – WEB SEARCH TASKS	63
LIDE S. OAM LE SAIATIEL HEB OLAKOH IAOKOMMINIMI	
ADDENDIX C. FOLLOW UD QUESTIONNAIDE DESDONSES	G.A



## **TABLES**

Table 1: Construct Coding for the Feasibility Study Tasks	14
Table 2: PISA ICT Feasibility Study Sample	31
Table 3: Retrospective Interview Sample	32
Table 4: Section and Timing Information	33
Table 5: Self-Reported Academic Performance	36
Table 6: Self-Reported Location and Frequency of Computer Use	36
Table 7: Self-Reported Use – Operating System and Computer Type	37
Table 8: Self-Reported Computer Use	37
Table 9: Computer Use - Purpose and Frequency	38
Table 10: Process Coding for the Feasibility Study Tasks	39
Table 11: Access Tasks	40
Table 12: Manage Tasks	40
Table 13: Difficulty and Timing by Section	44
Table 14: Percentage of Students Correctly Completing Elements of E-mail Task	45
Table 15: Percentage of Students Correctly Completing Elements of Web Abstract Ta	ask
	46
Table 16: Percentage of Students Correctly Completing Elements of Web Search Tas	sk
	47
Table 17: Percentage of Students Correctly Completing Elements of Simulation Task	.49
Table 18: Correlations among Observed Proficiency, Computer Attitude, and Comput	er
Familiarity	54

### **ACKNOWLEDGEMENTS**

The ICT feasibility study represents the collaborative efforts of the Australian Council for Educational Research (ACER), the National Institute for Educational Policy Research (NIER) of Japan, the Tokyo Institute of Technology, and Educational Testing Service (ETS) in the United States. Many individuals within those organisations contributed to the ICT feasibility study. Those individuals, and their roles in the study, are listed below.

### Australian Council for Educational Research:

ACER Project Director	Ross Turner		
Technical assistance	Michael Castro and Sam Haldane		
Co-ordinator, test administration	Marten Koomen		
Test development	Juliette Mendelovits		
Test administration	Joy McQueen, Ross Turner and		
Test aurillistration	Margaret Wu		

## National Institute for Educational Policy Research of Japan:

NIER Project director	Ryo Watanabe		
Translation and adaptation, test	Hideki Maruyama, Taro Numano		
administration	and Masako Shinohara		

#### Tokyo Institute of Technology:

Co-ordinator	Kanji Akahori
Co-ordinator	Hiromitsu Muta

## **Educational Testing Service:**

Project director, ICT Panel Chair	Irwin Kirsch		
Systems design, development,	Kevin Bentley, Douglas Forer,		
integration, and data management	Lou Mang, and Michael Wagner		
Translation and adaptation	Yoko Futagi		
Test site co-ordination	Altamese Jackenthal		
Co-ordinator, test administration	Holly Knott		
Project co-ordinator and test	Marylou Lennon		
development	Warylou Ecillon		
Interface and screen design	Debbie Pisacreta		
Test administration	Peggy Redman		
Screen and graphic design	Janet Stumper		
Statistical analysis	Matthias Von Davier and		
Statistical analysis	Kentaro Yamamoto		



### **SECTION 1: INTRODUCTION**

#### Overview

- 1. This report presents the findings of a study conducted to explore the feasibility of developing and delivering an assessment of information and communication technology (ICT) literacy for the Programme for International Student Assessment (PISA). The study was conducted at the request of an expert panel convened by Network A and was funded in large part by the Organisation for Economic Cooperation and Development (OECD) with some support from Statistics Canada. The project reflects the collaboration of three organisations: the Australian Council for Educational Research (ACER) in Australia, the National Institute for Educational Policy Research (NIER) of Japan, and Educational Testing Service (ETS) in the United States.
- 2. Working together, ACER, NIER, and ETS developed a set of performance-based ICT assessment tasks that were tested on a small sample of students in Australia, Japan, and the United States. While the study itself represents a collaboration among these organisations, as well as the ICT Expert Panel, this report has been prepared by staff at ETS with limited input from others. This was not by design, but rather the result of a schedule that allowed only a few weeks for data analysis and reporting. Had there been additional time available, the report would most certainly have benefited from the input of everyone who worked so hard to ensure the study's completion.
- 3. This report is organised into four sections:
  - The Introduction section presents the background of the project, summarising the work of the expert panel in defining ICT literacy and developing a framework for assessing the skills and knowledge within that domain. In addition, it articulates the goals of the study and summarises the general conclusions drawn from the study results.
  - The Procedures section explains how the study instrument and all accompanying materials were developed. The sampling plan for the study is discussed and the procedures followed in setting up and running the data collection process are explained.
  - In the Results section, the procedures used for harvesting the data and defining variables for analysis and reporting are discussed. Student responses to the questions presented in the follow-up questionnaire are also summarised.
  - The report concludes by returning to the goals for the study, as set out in the introduction. We consider the lessons learned relative to those goals and their implications for moving forward with assessments of ICT literacy



#### **Work of the ICT Expert Panel**

- 4. Early in 2003, an expert panel established by Network A met to address the issue of including a measure of information and communication technology (ICT) literacy skills in the Programme for International Student Assessment (PISA). The panel consisted of international experts representing six countries and a range of perspectives on ICT<sup>1</sup>. These experts were charged with articulating a rationale for building an ICT assessment, developing a definition of ICT literacy, and defining the criteria to be used in creating ICT assessment tasks and reporting results to an international community.
- 5. The PISA assessments, developed and conducted under the auspices of the Organisation for Economic Cooperation and Development (OECD), measure the achievement of 15-year-old students as they approach the end of secondary school. These assessments are designed to provide policy makers, educators, researchers and the public with measures of how well students are acquiring the skills and knowledge necessary to participate fully in society, as well as to meet the challenges of the future. Recognising the growing importance of information and communication technologies in all aspects of people's lives, PISA decided to explore ways in which ICT literacy could be incorporated into its assessment design.
- 6. The ICT Expert Panel presented its framework for ICT literacy skills in April of 2003. Members of the Panel concurred that there was a strong rationale for including a measure of ICT literacy in the PISA assessments, citing the prevalence and increasing importance of information and communication technologies. The Panel articulated a set of underlying observations and beliefs that they felt provided the basis for this judgement. These are summarised as follows<sup>2</sup>:
  - ICT is changing the nature and value of knowledge and information.

    The growth of information and digital communication technologies, including capabilities for networks and shared environments, is changing the nature of social interactions and collaborative endeavours. Participating in this digital world is fast becoming a necessary condition for successful participation in society.
  - ICT literacy, in its highest form, has the potential to change the way we live, learn and work. Higher levels of ICT literacy have the potential to transform the lives of individuals who develop these requisite skills and knowledge. This will influence not only the kinds of activities those individuals perform at school, at home and in their communities but also how they engage in those activities. ICT has the potential to change how we think and learn, advantaging not just individuals but societies as a whole.
  - There is a lack of information about the current levels of ICT literacy both within and among countries. Data from large-scale global assessments, such as PISA, are critical for understanding the outcomes and effectiveness of current policies and educational programs, as well as for identifying potentially new and more effective strategies.

<sup>&</sup>lt;sup>2</sup> See "The PISA Framework for Assessing ICT Literacy: Report of the ICT Expert Panel", April 2003, for the complete discussion of these points.



7

<sup>&</sup>lt;sup>1</sup> See Appendix A for a list of ICT Expert Panel members

7. In order to clarify its views about what an assessment of ICT literacy would and would not include, the Panel developed a working definition of ICT literacy and articulated its underlying assumptions. The following definition was adopted by the Panel:

ICT literacy is the interest, attitude, and ability of individuals to appropriately use digital technology and communication tools to access, manage, integrate, and evaluate information, construct new knowledge, and communicate with others in order to participate effectively in society.

- 8. The Panel had specific views about what the term ICT means in the current technology environment. They distinguished between traditional information technology (IT), which commonly refers to the electronic display, processing, and storage of information, and ICT, which reflects the convergence of computer and communication technologies. In the Panel's view, ICT is characterised by unprecedented global flows in information, products, people, capital, and ideas. These flows are enabled by ICT; their sheer scale and pace would not be possible without the ability to connect vast networks of individuals across geographic boundaries at negligible cost.
- 9. In addition to developing a working definition of ICT literacy, the Panel identified a preliminary set of variables they believe influence performance on ICT literacy tasks. These include contexts, processes and technology environments as discussed below.

#### - Contexts

The Panel recognised that the intent of Network A was for the ICT literacy assessment to measure the skills and competencies associated with ICT, rather than measure a single academic domain such as science or mathematics using ICT. From their own experience, Panel members understood that students do not use information and communication technologies within a single context – for example, only in math class or only at home. As a result, assessment tasks must sample broadly across both "traditional" subject area domains and real-life contexts that represent the range of situations in which information and communication technologies are used.

The four context areas identified by the Panel included: ICT for personal use, public use, educational, and occupational purposes. The Panel believes that each of these areas represents real-life situations and is relevant to the overall goal of PISA to provide information about how well students are acquiring the ICT skills and knowledge they need today and will need in the future.

#### Processes

The Panel identified six processes that they see as critical components of ICT literacy: accessing, managing, integrating, and evaluating information, constructing new knowledge, and communicating with others. This list reflects the Panels' belief that ICT is not about specific technical skills, but rather it is about information gathering, knowledge construction and communication. As such, these processes reflect the integration of technical knowledge and skills with more traditional cognitive skills such as literacy and numeracy.



#### - <u>Technology environments</u>

Just as students do not use information and communication technologies in a single context, they typically do not work within a single ICT environment. The Panel identified three technology environments for the ICT assessment: Web, desktop, and e-learning environments. The Panel recognised that these environments do not represent all aspects of ICT, but selected these because of their prominence and because they are thought to be familiar to most students. While each technology environment may have overlapping functionality, the Panel believes that collecting evidence across a range of ICT environments will allow students to demonstrate a broader range of knowledge and skills and thus allow PISA to collect better evidence about what they know and can do.

10. In its framework document, the ICT Expert Panel recommended that a feasibility study be conducted to investigate a number of important questions for the ICT literacy assessment. It was proposed that the study be conducted in three countries with materials developed in two languages, English and Japanese. The study was scheduled to begin in May of 2003 and be completed by October 2003, when results would be reported to Network A. This report presents the findings of the ICT feasibility study that was implemented by the Australian Council for Educational Research (ACER) in Australia, the National Institute for Educational Policy Research of Japan (NIER), and the Educational Testing Service (ETS) in the United States.

## **Goals for the Feasibility Study**

11. A major purpose of the ICT feasibility study was to gather information about a number of key development and delivery issues, as defined by the ICT Expert Panel. This included exploring issues related to the following.

#### - Simulated applications and environments

The Panel recommended the use of simulated applications and environments in the assessment rather than using existing software. This decision was driven by a desire to ensure that all students had a consistent set of tools and that no one would be disadvantaged by a lack of prior knowledge about how a particular application functions. The Panel also felt that the use of simulated applications would facilitate the capture and scoring of performance, including data relating to the processes used by students to complete tasks. However, a number of questions related to the development of such applications and environments needed to be explored in the feasibility study. How much time would it take to develop simulated applications and environments with the capabilities necessary to present students with realistic tasks? What were the technical issues involved? What kinds of performance information could be collected within those applications and environments using an online data capture system?

#### - Translation and adaptation

Another critical issue for PISA was to understand the issues related to translating and adapting both the test interface and content. Could simulated applications and environments be developed that students across an international community could use? Would these function so that no particular group was disadvantaged? How would translation and adaptation affect the development process and



schedule? Exploring such issues was the motivation behind the Panel's recommendation that materials for the feasibility study be developed in two languages, English and Japanese, and be applicable across three cultures, Australian, Japanese, and American.

### - Usability

The study was also intended to provide information about how comfortable students are taking a computer-based test and how well they are able to perform in that environment. The study would allow initial data to be collected about usability as it related to the general interface, task interface, and directions used in the feasibility study software.

- 12. In addition to these development and delivery issues, the study was intended to provide preliminary information about task difficulty and timing. The assessment tasks proposed by the Expert Panel included innovative item types and limited prior experience was available to guide the test development process should ICT become part of the main PISA assessment. Which kinds of tasks would students find easiest and which would be more challenging? How long would it take students with varying levels of computer experience to complete the tasks? The study was intended to provide some initial information about the difficulty level of a variety of ICT tasks as well as information about how long students spent completing those tasks.
- 13. It is important to note that the feasibility study attempted to address a wide range of issues while operating under a number of constraints. First and foremost was the timeline, which allowed five months from initial task development to reporting results. This schedule required that developers focus on the purposes outlined above and did not allow time to explore the breadth or depth of the ICT literacy construct as outlined by the Expert Panel. Tasks had to be developed quickly, particularly given the requirement to translate and adapt them into a second language. The timeline and translation requirements also meant that there was insufficient time to experiment with the inclusion of more innovative multimedia elements (such as audio or video) in the assessment materials.

#### **Summary of General Conclusions**

- 14. The ICT feasibility study was designed to provide those responsible for PISA with information needed to proceed with a full assessment. To accomplish that purpose, the study ultimately had to address two fundamental questions. Is it feasible to develop and administer a computer-based assessment of ICT literacy? And, given the cost and time commitments required, is a computer-based ICT test worth doing? That is, can we learn things from such an assessment that are important and that cannot be learned from other types of assessments? While many questions remain to be studied, the considerable amount that was learned with respect to both of these questions would lead us to answer yes, such a test is both feasible and worth doing.
- 15. The feasibility question can be looked at from a number of different perspectives. From the test development perspective, the question was whether performance-based tasks could be designed that would assess ICT skills and knowledge as defined by the ICT Expert Panel. Working together, ACER, and ETS were able to demonstrate the feasibility of developing a set of performance-based tasks that covered many of the



contexts, processes, and environments specified by the Panel. In addition, NIER, ACER, and ETS successfully partnered to adapt and translate those tasks so that students in Australia, Japan and the United States were able to understand and perform them. Much was learned through this development process that can help inform future test development efforts.

- 16. The technical feasibility of programming and delivering the variety of tasks included in the ICT assessment was also demonstrated in the study. Software for assessing a wide variety of ICT skills was developed, and where possible, off the shelf components implementing specific ICT functionality were integrated. The test delivery system supported complex, multipart performance tasks and collected data on the process students used to produce a work product, along with their final response. In addition, the ability to develop and deliver complex tasks in two very different languages was demonstrated. Software technologies for supporting multibyte character sets were utilised so that a single codebase was implemented. Methodologies and processes for managing the process of developing computer-based assessments in multiple languages were created and potential improvements in this process were identified.
- Finally, from a psychometric perspective, the ICT assessment introduced opportunities beyond those offered by traditional paper and pencil assessments. New types of items delivered by a web-based platform allowed us to collect both performance or process indicators as well as variables that indicate knowledge in the ICT domain. Process indicators are variables such as response times and the sequence of user actions, both of which were collected and analysed. The range of defendable statistical analysis was restricted by the sample and domain coverage resulting from the time and design constraints of the study. Still, the analysis was able to extract a range of behavioural indicators from the raw data and show that they are systematically related. For example, within each section of the assessment, it was found that response time and response accuracy were positively related to self-reported computer proficiency skills as well as observed measures of ICT proficiency. The analyses show that response time alone may not be used as a performance measure, but it can serve as a valuable auxiliary variable in statistical models that combine both respondents' performance data and auxiliary information such as background data. The overall time used for the different sections, for example, was on average much longer for subjects rated with low ICT proficiency as compared to subjects rated as more proficient. Process variables specific to particular tasks began to emerge as performance indicators and will continue to be the focus of future investigation. The preliminary findings of the study support the belief that the automatic recording of user actions and timing data will allow more in-depth analysis in the future, given a larger sample and a broader range of items.
- 18. Given that an ICT assessment is feasible, is it worth doing? Can we learn things that are important and that cannot be learned from other types of assessments used in PISA? The belief of the ICT Panel was that a computer-based assessment of ICT literacy would have advantages over paper and pencil tests in terms of validity, data capture, and operational efficiency. One assumption was that delivering an assessment of ICT literacy in a computer environment would contribute to construct validity because the characteristics of the included tasks and response patterns more closely match the processes. A computer-delivered assessment based around authentic performance-based tasks, therefore, would provide more appropriate evidence about ICT knowledge and skills among students participating in PISA.



- 19. The computer-delivered instrument developed for the study was able to capture information that would be difficult or impossible to collect via a paper and pencil instrument. For example, the computer was able to record the time students spent on each task as well as where they clicked on the screen and the sequence of those clicks. The system went well beyond recording right/wrong responses to provide information about the path students took to complete tasks, giving us insight into the processes and strategies they employed.
- 20. Finally, one might expect that a computer-delivered assessment of ICT literacy would be worth doing because of operational efficiencies that would contribute to what could be learned. The study demonstrated several areas in which this was, indeed, the case. For example, administering the background questionnaire online worked very well. The students in the study sample had no difficulty with the limited technical skills required to complete the multiple-choice questions and, on average, were able to complete the background questionnaire in 3 minutes. We were able to collect a complete set of responses, as the software required students to answer each question, thus preventing skips and missed questions. In addition, the built- in branching capability meant that students did not need to spend time on unnecessary questions or make the type of errors that may be introduced on a paper and pencil questionnaire when students are directed to skip a particular question or set of questions. Managing the data files generated from the assessment tasks was one of the major challenges of the study. However, it was clear that as the parameters are better understood and defined, automatic coding of behaviours and generation of data files has the potential to save time and reduce errors.
- 21. In sum, based on our experience with the feasibility study we believe that developing and administering an ICT literacy assessment would be challenging, but successful. The challenges lie in the number and complexity of new (and sometimes unanticipated) hurdles that this type of endeavour will necessarily present. We think a planned phased approach would provide the best opportunity to identify and clear these hurdles. There is tremendous potential to learn more about what students know and can do in an ICT environment that would yield valuable information about ICT skills and would lay the groundwork for the future use of computers in the development and delivery of large-scale international assessments.



#### **SECTION 2: PROCEDURES**

- 22. This section explains the procedures followed to
  - Develop the study materials
  - · Recruit and select participants, and
  - Organise and conduct testing sessions.
- 23. First we will focus on the materials designed for the study, beginning with an explanation of the general development and item authoring procedures used to create the computer-based tasks. Samples of tasks within each test section will be presented and issues specific to those tasks will be discussed. Additional materials, including the follow-up questionnaire, retrospective interview packet, and administration manuals, will also be explained.
- 24. In order to successfully run the feasibility study, students in three countries had to be recruited and selected and testing sessions had to be organised in a consistent way. The sampling procedures followed for the study, both for the total participant sample and the retrospective interview sample, will be outlined. Finally, the guidelines developed for organising and running test sessions will be presented and discussed.

## **Development of the Study Materials**

- 25. Four sets of materials were developed for the ICT feasibility study: a background questionnaire, the assessment tasks, a follow-up questionnaire for participating students, and test administrators' materials including observation protocols and an administrator's manual.<sup>3</sup> Both the background questionnaire and assessment materials were computer-based materials. The follow-up questionnaire and administrators' manuals were developed as paper-and-pencil materials.
- 26. The assessment tasks were based on blueprints developed at the final meeting of the ICT Expert Panel in April of 2003. At that time, a set of draft tasks was presented and discussed and the Panel's input was obtained. Within a few weeks after that meeting, the test development team was organised and began refining and adapting those draft tasks. Early discussions were held with technical experts on the team to determine the range of functionality that could be developed and delivered in the available time. As tasks were designed, the interface and functionality were refined and revised in an iterative process in which test developers worked collaboratively with systems designers.
- 27. ETS and ACER staff members collaborated to draft initial prototypes of the tasks for the feasibility study. These tasks were designed to cover a range of the process components (access, manage, integrate, evaluate, construct and communicate), environments (Web, desktop, and simulation) and contexts defined under the Panel's ICT framework. As shown in Table 1, the team was able to develop tasks that covered a range of these construct areas.



<sup>&</sup>lt;sup>3</sup> A copy of all these materials is provided in the Materials Annex accompanying this report.

Table 1: Construct Coding for the Feasibility Study Tasks

	Access	Manage	Integrate	Evaluate	Construct	Communicate
E-mail task Environment: Desktop Context : Personal	Х	Х				Х
Database task		Х	Х			
Environment: Desktop Context : Public		^	^			
Web abstract task Environment: Web Context : Personal				Х		
Web search task Environment: Web Context : Educational	X	Х	x	X	x	X
Simulation task Environment: E-learning Context : Educational	Х	Х	x	Х	x	X

28. Environment and context coding was applied at the task level. In addition, each task was made up of a number of component elements and each of those was coded individually by process. For example, the database task was coded as a desktop environment task in the public context because it required students to use a simulated desktop application and was situated in the context of a community fundraising event. Within the task, students needed to search on or organise information based on the categories defined in the database (classified as a 'manage' component) and to compare and contrast information (an 'integrate' component) in order to identify the number of CDs meeting the specified criteria.

#### **General Technical Procedures**

- 29. The computer-delivered components of the study were implemented within the C3 simulation client of ETS's internet based testing system (iBT). C3 leverages Web browser technology to provide a rich, extensible platform for item display. C3 utilises dynamic HTML, as well as Java and Flash, to display a wide range of item types, ranging from simple multiple choice to complicated E-Learning and web search simulations. The item display system is backed up by a robust and flexible data logging facility, which provides for the collection and scoring of complex user responses.
- 30. A toolbox of ICT user interface components was assembled for the study. Some of the components, such as the simulated Web browser and search engine, had already been developed by ETS for other projects and were adapted for use in this study. Other components, such as sortable tables and bar/line chart displays, were obtained from third party sources, either as open source or commercially licensed software. Finally, some components, such as the simulated e-mail application, were developed from scratch. Test developers utilised this toolbox in the design of the tasks for the study. Software developers worked closely with the test developers, advising on the feasibility of implementing proposed tasks.



31. The multilingual requirements of the ICT feasibility study necessitated some minor adaptations to the C3 system. Unicode, and in particular the UTF-8 format, was selected as the base format for the electronic study materials because it supported all the character sets that might be of interest in a larger PISA assessment. The technologies on which C3 is built, Java, HTML, XML, JavaScript, Flash, and Microsoft SQL Server 2000, all support Unicode. There were a limited number of C3 system components with embedded assumptions about 8-bit wide characters, which Unicode violates, but these were easily identified and corrected.

## General Translation/Adaptation Procedures

- 32. Two versions of materials were developed for the feasibility study: one in English and one in Japanese. Ideally, a third Australian version would have been created, but time constraints made that impractical. All materials were developed first in English. Static screen mock-ups were created and staff members at ACER and ETS reviewed all materials. In addition, elements of the interface (labels for icons, user prompts, general direction screens, etc.) were also reviewed. During this process discrepancies between British and American English were resolved wherever possible. For example, one question on the background questionnaire asked students "What was your average grade in each of the following subjects over the past academic year?" The first response option was originally written as "Math." Reviewers pointed out that students in Australia would refer to that subject as "Maths." A compromise was reached by using the term "Mathematics" which would be familiar usage for students in both countries.
- 33. Once we had agreement on the English version of the text, materials were prepared for translation. The procedure followed was to create an Excel file with three columns: one contained the English text, one provided space for the Japanese translation, and one contained adaptation notes. The notes column contained adaptation guidelines provided by ETS to ensure consistency across versions (e.g., the name of the e-mail recipient can be replaced with any Japanese name but that change must also be reflected in lines 33, 45, and 60). Other acceptable modifications to accommodate differences, such as variation in measurement and monetary systems, were also listed. Translators from NIER also used this column to provide information about adaptations that needed to be made to the English version of the materials (for example, references to reduced book prices in some Web pages were eliminated since, in Japan, books are sold only at list price).
- 34. The English text in these translation files was presented line-by-line, and sometimes word-by-word. This level of detail was necessary because ETS development staff, who could not read Japanese, needed to integrate the translated text into the HTML and .XML files created for the tasks, system messages and directions, and include it in the graphics files (. JPG and .GIF) for the icons and screen design. By matching each line of English text in the translation file with the appropriate line in the HTML code or GIF image, the corresponding Japanese text could be identified and put into the Japanese version of the software. Although it was somewhat cumbersome, we found that this process functioned until we needed to make final revisions to the English text. At that point, it became impossible for someone without any knowledge of the language to locate the corresponding Japanese characters in the code. One ETS staff member, a linguist who was a native Japanese speaker, assisted the development team by checking the Japanese version against the English, working with the programmers to



locate and make necessary revisions within the code, and finally checking the revised Japanese version for errors.

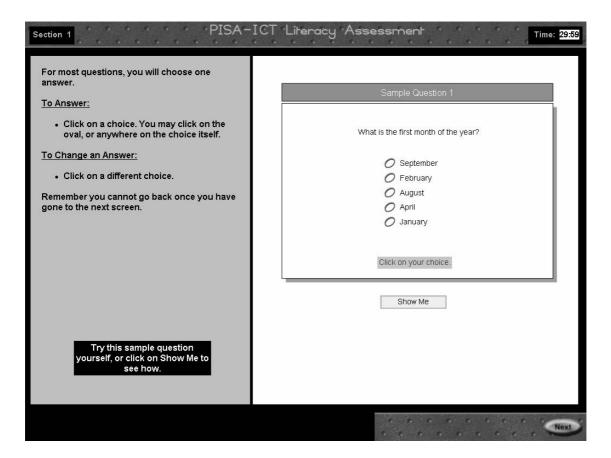
## **Developing the Assessment Tasks**

### Section 1 – The Introductory Tasks

- 35. The test included a short set of tasks designed to be administered prior to the assessment tasks. Initial plans called for the survey to begin with a tutorial that would present students with simple functional tasks including using a mouse to move, point and click, scrolling, typing, and basic editing such as inserting, deleting, and changing text. Those students who had difficulty with such tasks would receive some brief instruction and practice exercises before going on to the assessment tasks. Upon review, this plan was judged to be overly instructive, providing students with too much ICT-related information.
- 36. The revised introductory section had three parts: how to answer questions on the background questionnaire, the background questionnaire itself, and a brief set of tasks designed to obtain baseline data about students' basic technical skills. Samples from each section are included in the following descriptions and complete sets of screen print outs, both in English and in Japanese, are available in the Materials Annex.



#### RESPONSE MODES FOR THE BACKGROUND QUESTIONNAIRE

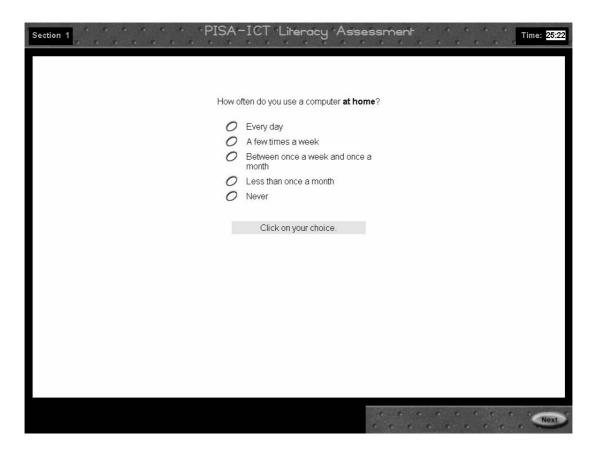


- 37. Students were first introduced to the three response modes required to complete the online background questionnaire: single-selection multiple choice (as shown above<sup>4</sup>), multiple-selection multiple choice, and type in. Students were presented with explanations on how to answer each question type and then asked to complete a practice item and/or click on Show Me to see an animated demonstration. The design of these screens was based on tutorials ETS developed for programs such as the Test of English as a Foreign Language (TOEFL). Leveraging this past experience allowed us to take advantage of the usability testing and field experience associated with these earlier development efforts.
- 38. The explanations and practice in this section were intended to ensure that students would be able to correctly complete the background questionnaire. If any student had insufficient technical skills (a situation that did not arise in the feasibility study), test administrators were instructed to provide assistance by clicking or typing responses supplied by the student and clicking on 'Next' to move through the questions.

<sup>&</sup>lt;sup>4</sup> Upon the recommendation of our colleagues at NIER, this particular item was replaced with a parallel one for the Japanese version. The question would be a nonsensical one for Japanese students as months are designated numerically ("1" month, "2" month, etc.) in Japanese.



#### THE BACKGROUND QUESTIONNAIRE



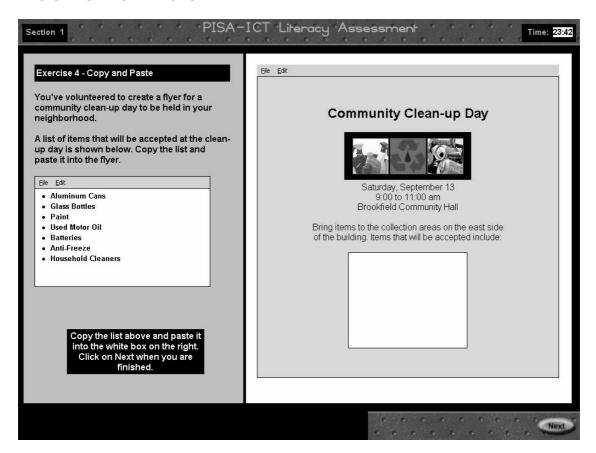
- 39. The background questionnaire formed the second part of the introductory section. This online questionnaire contained 22 questions organised into 5 general areas:
  - General participant information (gender, age, performance in school)
  - Access to computers
  - Frequency and type of computer use
  - Frequency and type of applications used
  - · Attitudes about computers

The background questions were drafted at the last meeting of the ICT Expert Panel in April of 2003, with some revisions in wording and format made as the final questions were refined and authored into the delivery system.

40. The questionnaire was designed with two branching points. If students responded that they had never used a computer prior to the assessment, they skipped questions about patterns of computer use and branched to the attitude questions. Students who indicated that they never use the Internet to communicate or collaborate with others skipped the question about the type of Internet connection they used.



#### BASIC TECHNICAL TASKS



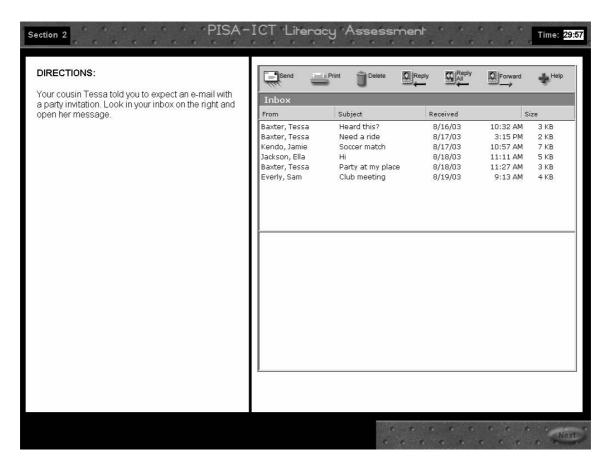
41. A brief assessment of technical skills followed the background questionnaire. The purpose of this section was to obtain evidence about very basic abilities such as vertical and horizontal scrolling, typing, highlighting, and then being able to cut and paste a block of text (as shown above), dragging an object to a desired location on the screen, and using both pull-down and drop-down menus. These skills were selected because they represented the set of basic technical skills required to complete the subsequent assessment tasks. In a larger PISA assessment, a similar set of tasks might act as a screener, allowing students a way out of the survey if they lacked the most basic technical skills required to complete the assessment tasks.

### Section 2 – The Short Scenarios

42. The intent behind the short scenarios was to provide easy, discrete tasks that would broaden the range of task difficulty sampled in the study. These tasks were included in the feasibility study to provide information about their potential usefulness in the main assessment to collect information about students at the low end of the ICT proficiency scale. Three short scenarios were included in the feasibility study: an e-mail task, a Web abstract evaluation task, and a short database task. The development of each is briefly described below.



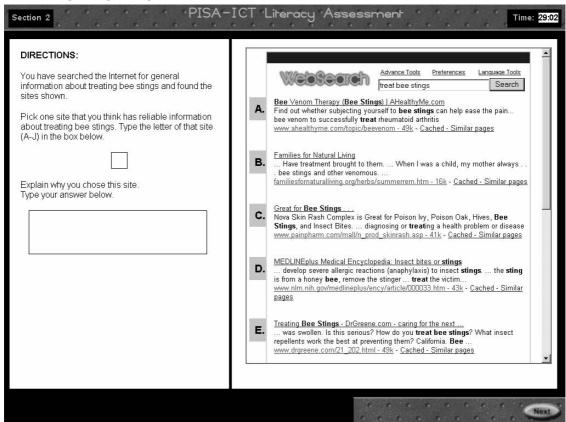
#### E-MAIL TASK



- 43. In this task, students were asked to use a simulated e-mail application. Students were directed to open a message in an in-box and forward it to a friend. Upon clicking 'Forward,' a new set of directions displayed. Students were instructed to add a short message and then send the e-mail to the address provided, along with a copy to another friend. This task required students to access, manage, and communicate information, at a basic level. The task was very directed, with each step explained in turn.
- 44. The interface for this e-mail task was designed to be easily usable for students familiar with a range of e-mail applications. Icons were designed to be unique to this assessment but they incorporated features of existing PC and Mac-based applications. Given the somewhat novel interface elements, a Help screen was developed to provide information about the icons and their functions. Information on the Help screen was listed in a simplified way that was intended to reduce the reading load.
- 45. The software was able to capture a clear representation of student actions in this task. Because the task was divided into three distinct parts (opening the e-mail, forwarding it, including a message and sending it) the sequence of clicks and other actions could be easily mapped to the path students took through the task.



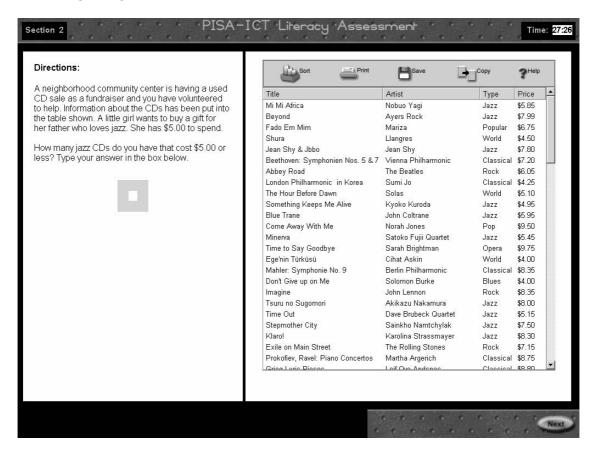
#### WEB ABSTRACT TASK



- 46. The Web abstract task presented students with a static page of hits resulting from a Web search. Students were asked to first identify a site that had reliable information about treating bee stings. The page contained ten sites, all of which had some mention of bees or bee stings. None of the links on the page were active so students had to rely on the title, description, and URL information presented. Students had to scroll down the page in order to view all ten sites. Only a few of the sites presented information about the treatment of bee stings and, of those, some were commercial sites advertising products, some were sites sharing personal anecdotal information and others were government or education-based sites. A second task asked students to identify another site from the list that contained information about treating bee stings but was not as reliable.
- 47. For both tasks, students were asked to type an explanation for their selection. This was followed by a multiple-choice item that listed possible reasons for selecting sites and asked students to indicate which ones applied to their choice. In the main assessment, a multiple-choice item would be preferable for scoring efficiency, particularly across multiple languages. Ideally, one would want the multiple-choice options to cover the range of responses students might provide in an open-ended item. Since including an open-ended item first in a pilot or pre-test might be one way to inform the development of those options this double-questioning process was tried in the feasibility study.



#### DATABASE TASK

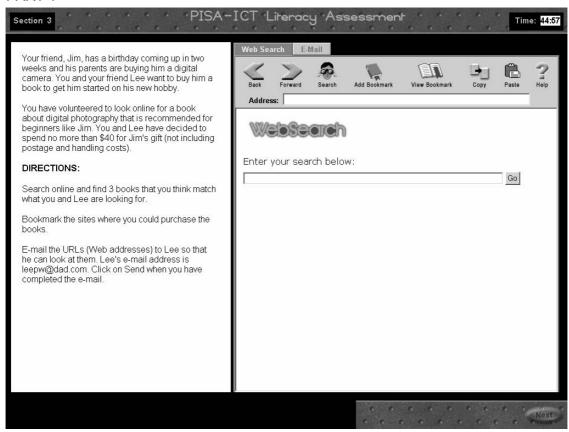


- 48. This task presented students with a database of 100 CDs listed by title, artist, type, and price. Students were asked to determine the number of jazz CDs that cost \$5.00 or less. There were two ways to complete the task either by scrolling down the list and counting the CDs matching the specified criteria, or using the 'Sort' function. When students clicked on they were asked to identify the columns to sort by, and the order, similar to the sorting function in programs such as Excel.
- 49. To create a version of this task in Japanese, the translators replaced the titles with other Japanese CDs. The placement of the correct responses was maintained, so that the correct CDs were located on lines 11, 34, and 80 in both the English and Japanese versions of the task.
- 50. The computer was able to capture the response students entered as well as any screen clicks so it was possible to record if a student clicked on 'Sort' and, if so, how he or she sorted (number of variables selected, ascending or descending order, etc.)



## Section 3 – The Web Search Task

#### PART 1



- 51. One of the charges presented to the development team by the ICT Panel was to explore the development of a simulated Web universe to support a performance-based Web search task. While the team was aware of other efforts to create such universes for research purposes, we did not know of anyone who had tried to create parallel universes across languages. As we worked to design this task, we had to consider the time and expense involved in developing a simulated Web universe in 10 or more languages, the ultimate target for the main PISA assessment. We also had to keep in mind that this universe would need to function comparably across multiple versions.
- 52. One of the challenges inherent in working in the real Web universe is that it is both broad and deep. That is, it contains billions of pages and on each page the user is both able to drill down many levels and link to other pages. However, as we began working on our Web task it quickly became obvious that creating a universe that was both broad and deep simply wasn't feasible for PISA. The number of pages involved and the resulting translation and adaptation requirements would be prohibitive, both in terms of time and expense. Therefore we decided to experiment.
- 53. We created a universe that was relatively broad, containing 65-70 pages for the study with a goal of about 100 pages in subsequent versions. What was different about this universe was that it was not at all deep. Students could enter a search, read

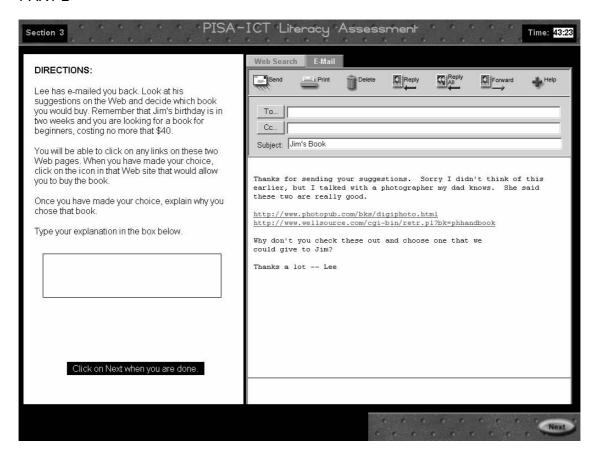


abstracts, open any of the hits, go back and revise searches, bookmark the hits, etc. But each hit was only one level deep – that is, none of the links on the Web page were active (as was emphasised to students in the section directions for the task).

- 54. To assess skills and knowledge associated with a deep universe, future efforts could include an additional task where students have to navigate around one very deep Web site. The task could place students at a particular point in that site and require that students use hyperlinks to navigate around, as well as up and down the levels of one particular site.
- The Web universe was created first in English. Within this universe, three pages 55. were designed to be correct responses based on the criteria set out in the task. These pages were modelled after real book sites and created as HTML pages by ETS development staff. The three pages were then translated by NIER to ensure that they were identical in both English and Japanese. The other English sites in the universe were located by searching on related, but incomplete, queries such as 'digital' (including a page about digital televisions) and 'photography' (including pages about careers in photography and online photography courses). Specifications based on the queries, as well as the resulting hits (e.g., one book about underwater digital photography and one collection of famous photographs) were developed to provide the translators with guidelines for the selection of parallel sites. Given the iterative nature of this selection process and the limited timeframe, the sites for the Japanese Web universe were selected by an ETS staff member (a native Japanese speaker) working with the test developer who selected the English sites. The Japanese sites were then verified by a second ETS staff member who was also a native speaker.
- 56. In Part 1 of the task, students were asked to conduct a search to look for three books that would make a suitable gift for a friend who was just learning about digital photography. Students were asked to locate books meeting the specified criteria and email the URLs to another friend who was sharing in the purchase. To successfully complete the task, students had to enter one or more queries, locate appropriate books, bookmark the URLs and/or copy them into the e-mail, and send the e-mail to the address provided. Students were able to toggle back and forth between the search and e-mail windows by using one of two tabs located on the screen. This tabbed-based interface was not an authentic environment, but was provided as a substitute for the typical manipulation of multiple windows across different applications. This trade-off was made to simplify the implementation as well as to avoid interactions when students tried to manipulate simulated windows in the test delivery application by using Windows functionality (e.g., ALT-TAB). Complete handling of this situation would have required patching the operating system or implementing low-level device drivers.



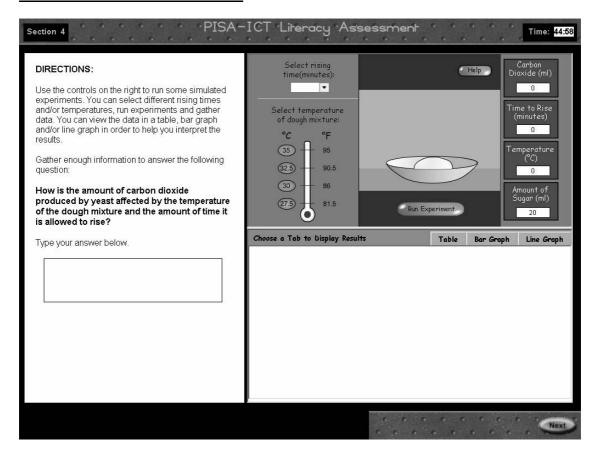
#### PART 2



- 57. In Part 2 of the task, students "received" an e-mail from their friend that contained two URLs for books he suggested for purchase. Students needed to use the links to check both sites and select one of the books based on the criteria provided. This second part was included to test the feasibility of creating a multi-step scenario in which the final response was not dependent on earlier actions. To make the final book selection all students were given the same two choices. This made it possible to evaluate whether students could correctly use the criteria to evaluate and select a book irrespective of the choices they had made in Part 1.
- 58. The two sites included in this part of the task were also created based on real sites located on the Web. Like the three correct sites in Part 1, these were directly translated to ensure that they would be parallel in the English and Japanese versions of the task.



#### Section 4 – The Simulation Task



- 59. The simulation task was situated in the context of understanding variables that influence the production of carbon dioxide in bread dough. Students were asked to run experiments in order to gather data about the relationships among temperature, rising time, and production of carbon dioxide. While no prior knowledge about these relationships was necessary, students did need to run a number of experiments and display the data in a way that made those relationships clear.
- 60. The interface for the task was not explicitly defined (although once again, a 'Help' function was available that provided a basic explanation of how the tools functioned). One goal of the task was to see if students could explore an unfamiliar interface and figure out what they needed to do in order to collect the necessary data and do so with some fluency. The task was open-ended, leaving students to decide when they had collected a sufficient amount of information and were ready to answer the question.
- 61. In the second part of the task, students were asked which of the representations (table, bar graph, or line graph) best illustrates the relationship between time, temperature and the production of carbon dioxide. Students could select a response or choose to run more experiments and/or make new tables or graphs. Once they responded, they were asked a final multiple-choice question about the relationship they had been exploring via the simulation.



62. The simulation task was implemented with a combination of Flash, HTML, and Java applet technology. The simulated experiments were implemented with Flash, which supports rapid development of animations and engaging interfaces. The bar and line charts were implemented with a third-party Java applet, the JetChart library from Jinsight Informatica Ltda. (http://www.jinsight.com/jetchart/). HTML was used to communicate between the simulations and the charts, as well as to implement the tables.

## **Developing the Follow-up Questionnaire**

63. A paper and pencil follow-up questionnaire was developed collaboratively by ACER and ETS to collect information about student's perceptions of the test. Students filled out the questionnaire after they had completed their testing session. Students were asked to share their opinions about a range of issues including how well the survey allowed them to demonstrate their computer skills and knowledge, how interesting they found the tasks and whether they felt that any features of the simulated applications interfered with their ability to complete the tasks. A sample page is shown below and a complete copy of the questionnaire can be found in the Materials Annex.

Next, please answer a few questions about each of these four tasks. Circle one answer (1, 2, 3 or 4) for each task.









E-mail task

Database task

Web search task (photography book)

Yeast simulation

12. Did you have enough time to carefully complete each task?

	More than enough time	Enough time	Almost enough time	Not enough time
E-mail task	1	2	3	4
Database task	1	2	3	4
Web search task	1	2	3	4
Yeast simulation	1	2	3	4

13. How easy did you find each task?

	Very easy	Easy	Difficult	Very difficult
E-mail task	1	2	3	4
Database task	1	2	3	4
Web search task	1	2	3	4
Yeast simulation	1	2	3	4

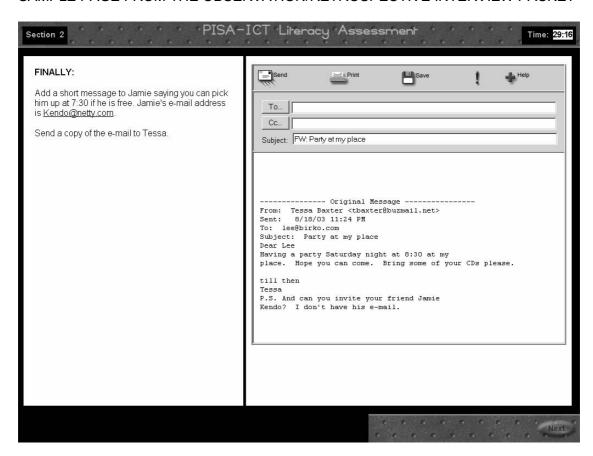


## **Developing the Observation/Retrospective Interview Materials**

- 64. We anticipated that the retrospective interviews would vary across test administrators and students, and did not want to require a lock-step set of procedures. Test administrators needed to have some flexibility so that they could use their own judgement about the best way to gather useful information from each student participating in a retrospective interview. However, some degree of consistency was necessary if the results were to be aggregated and summarised.
- 65. In an effort to provide a minimum level of consistency, an observation/retrospective interview packet was prepared which contained clips of most of the computer screens in the test. Test administrators used the packet both while they observed students and while they conducted the retrospective interviews. Each page was annotated with observation guidelines and spaces for recording students' actions and responses as they worked through the assessment. In addition, each page contained one or two suggested questions for the follow-up interview which test administrators conducted as they reviewed the screens with students once they had completed the assessment. Test administrators were asked to complete a packet for each student participating in a retrospective interview. A sample is included on the following page and the complete packet can be found in the Materials Annex.



#### SAMPLE PAGE FROM THE OBSERVATION/RETROSPECTIVE INTERVIEW PACKET



### Observation

L	Types e-mail message – includes appropriate information (pick up at 7:30)?	Yes	No
L	☐ Includes Jamie's e-mail address – circle: cuts/pastes types		

☐ Includes Tessa's e-mail address in CC – circle: cuts/pastes types

☐ Clicks 'Send' button

☐ Clicks on other icons – circle all that apply: 'Print' 'Save' '!'

☐ Clicks on 'Help'

Is the student able to easily complete the e-mail task? Yes No

#### Follow up

Tell me how you completed this part of the task.

Probe if they did not include Tessa's address (not notice the directions? Unable to locate address?)



## **Developing the Administrators' Manual**

- 66. While this small-scale data collection did not require major administrator training efforts, we did want to ensure that the test administrations were as similar as possible across the testing sites in Australia, Japan, and the United States. A short manual was developed with the following contents:
  - Important Information at a Glance

Contact information and a list of short-cut keys that would allow the test administrators to refresh the screen, pause the software, etc.

How to Run the Software

More detailed procedural information such as how to log on to the system, how to pause, etc.

- Administration Record

This form was used to record student information, testing date, and type of computer used (laptop or desktop).

Session Materials and Procedures

Lists of materials needed for testing sessions as well as guidelines for intervening if a student asked for help or needed assistance during the testing session.

Event Log

This provided a place for test administrators to record any irregularities that occurred during a test sessions (hardware problems, interruptions that might impact task timing, etc.)

- Troubleshooting Guide

Guidelines for understanding problems that might arise with the software and what to do about them were provided in this section.

### **Recruiting and Selecting Students**

#### General Sampling Plan

- 67. The sampling plan for the ICT feasibility study specified 35 participants per country. The students, like those in the PISA population, were to be enrolled in secondary school and around age 15 (between 14 years/8 months and 16 years/4 months). Within each group of 35, there was to be a roughly even split between females and males. The plan also called for an ability distribution, with about 15 high-ability students and 20 lower-ability students. For the purposes of the study, a self-reported rating of academic performance was used to estimate ability. Students were asked in the background questionnaire to rate their overall performance in school during the past academic year. The five rating options included: excellent (mostly As), above average (mostly Bs), average (mostly Cs), below average (mostly Ds), and poor (mostly Fs). For the study's purposes, students reporting excellent and above average academic achievement were considered the high-ability group and other students were counted in the low-ability group.
- 68. The achieved sample, by country, is shown in Table 2.



Table 2: PISA ICT Feasibility Study Sample

	Desired	Achieved		
		Australia Japan U.S		U.S.
Number of participants	35	34	55	29
Age	About 15 years	34	55	29
Gender distribution	Roughly even	19 females	26 females	15 females
		15 males	29 males	14 males
Ability Distribution	15 high/20 low	28 high	33 high⁵	16 high
	-	6 low	17 low	13 low

- 69. The number of participating students varied across countries, ranging from a high of 55 in Japan to a low of 29 in the United States. Thirty-five students were originally scheduled for testing in Australia. However, one student was unable to attend a testing session and because this occurred on the last day of data collection it was not possible to reschedule another student.
- 70. Data collection had to occur over in the month of August when students in Japan and the United States were not in school, which made it challenging to for those countries to recruit. Despite this challenge, Japan was able to test more than the requested thirty-five students by recruiting students from two schools who participated in summer club activities.
- 71. Twenty-nine students were tested in the U.S. While additional high-ability students could have been included in the U.S. sample, the decision was made not to oversample the high-ability students and hold the total at twenty-nine students.

#### Retrospective Interviews

- 72. Plans for the study called for a subset of the participants in each country to participate in what was called the retrospective interview. These students were to be observed one-on-one by test administrators during the testing session. Once a student completed the session, an administrator reviewed the assessment screens with that student, asking about the processes he or she used to complete the tasks. Each country was asked to have 12 randomly selected students participate in these retrospective interviews, with the group roughly split between male/female and including about 5 high-ability and 7 low -ability students. In actuality, the process worked somewhat differently in each of the participating countries.
- 73. In Australia, random selection was achieved by having each student select a ball from a bag, with one (sometimes two) students being selected for the retrospective interview in each testing session. Selection for the retrospective interview was conducted as soon as possible after all students had arrived for the test. In Japan most participating students were tested in larger groups, ranging from 5 to 10 students at a

<sup>&</sup>lt;sup>5</sup> Ability data is available for only 50 of the 55 participating Japanese students. No responses for any items in Section 1 (including the background questionnaire) were collected for five students. Records indicate the ALT+R key combination was used to skip this section and we believe this may have been due to Internet problems.



,

time. Students were selected to participate in the retrospective interview according to their seating location in the school computer lab, with one student per row selected. For a smaller set of 8 students who took the survey at NIER, half were randomly selected to participate in the retrospective interview. Because of scheduling flexibility in the U.S. and the availability of staff from the ETS usability lab, each of the twenty-nine students participating in the study was observed and participated in the retrospective interview.

74. The achieved sample for the retrospective interview, by country, is shown in Table 3.

Desired Achieved Australia Japan U.S. Number of participants 12 12 15 29 5 females 7 females 15 females Roughly even Gender distribution 8 males 7 males 14 males 5 high 9 high 8 high 16 high Ability Distribution 7 low 3 low 3 low 13 low

Table 3: Retrospective Interview Sample

#### **Data Collection**

### **Operational Specifications**

- 75. All sites participating the in data collection for the feasibility study were provided with the following specifications for the hardware to be used in the study.
  - PC running Windows 98 or later
  - Internet Explorer version 5.5 or later
  - 300MHZ Pentium II processor or faster
  - Minimum of 64 MB of RAM
  - Display set to 1024 x 728 screen resolution
  - 16 bit screen colour depth or better
  - 128kbit available Internet bandwidth
- 76. Sites were requested to run a PC certification test to ensure that all their computers were set to run the software. An additional requirement was to check the Windows font size settings to ensure that they were set to Small Fonts/Normal Size. If set to Large Fonts, the screen layouts were affected and content would be too large to display properly on the screen.

#### **Testing Procedures**

77. The assessment tasks were divided into the following four sections, with timing allowances as noted in Table 4 below.



Table 4: Section and Timing Information

	Section	Time
Section 1	Introductory Material, BQ, Basic Tasks	30 minutes
Section 2	Short Scenarios	30 minutes
Section 3	Web search task	45 minutes
	Break	15 minutes
Section 4	Simulation	45 minutes

The section times listed above were intended to be maximums and students were allowed to continue on to the next section when ready. These times were selected because we wanted to allow enough time to gather the maximum amount of information from students while staying within the section times proposed in the design for the main assessment.

- 78. Students worked individually on a desktop or laptop computer. They began their testing sessions by logging into the system using an ID number assigned by the test administrator. Each participating country was assigned a block of ID numbers as follows: Australia 1001-1035, Japan 2001-20035, and the U.S. 3000-30035. Any country testing more than the requested 35 students simply appended additional numbers to the suggested range.
- 79. After logging in, students proceeded to the assessment tasks. Those students participating in the retrospective interviews were observed by a test administrator throughout the testing session. Test administrators used a retrospective interview packet that contained screen printouts for recording observations, suggested points to observe, and questions for the follow-up interview (with the expectation that test administrators would adjust their questioning to get the best information from each particular student).
- 80. If any student had questions during the testing session, test administrators were asked to follow the procedures outlined in the TA manual to ensure consistent responses. Test administrators were also asked to note any session irregularities in an event log.
- 81. Once students finished the online materials, they were asked to complete the paper and pencil follow-up questionnaire. Finally, students participating in the retrospective interview went through the software a second time to discuss how they completed each task with the test administrator who had observed them.



## **SECTION 3: RESULTS**

#### Introduction

- 82. The feasibility study demonstrated a range of capabilities that, taken together, present a promising first step in the development of an innovative assessment of ICT skills and knowledge. These include capacities to:
  - develop assessment tasks based on the framework established by the ICT Expert Panel,
  - translate and adapt those tasks across two very different languages so that students in three countries were able to use them.
  - deliver the assessment over the Internet with no major technical problems,
  - capture performance-based data including complex records of student actions,
  - report that data back over the Internet and create clean data files, and
  - begin developing a scheme for interpreting and understanding the results.
- 83. This section will focus on the data that was collected and the ways in which it might be aggregated and organised to understand the ICT skills of 15-year-old students. Results are discussed in terms of what they suggest about factors influencing the performance of different task types as well as what they suggest about the performance of students. The conclusions based on these analyses are tentative, given the nature of the feasibility study, but it is hoped that they will spur discussion and ultimately lead to more definitive models that can guide future data collection efforts.

#### Overview of the Analysis

- 84. As anticipated, the ICT feasibility study yielded an extensive amount of raw data. For each participating student, the computer captured time spent on each task, clicking responses (including opening a Web site, sending an e-mail, or performing some other action such as running a simulated experiment), and typed responses. To provide a sense of the sheer size of the resulting data files, an example of what was generated for one task is included in Appendix B.
- 85. One way to summarise this data is at the level of individual behaviours 92% of the students clicked on the 'Back' button during a Web search, 60% used bookmarks, etc. However, focusing at this level of detail does not provide much insight into general ICT proficiencies. The challenge for the study was to identify a set of variables that could guide the process of aggregating and analysing the data. We chose to approach this task from two perspectives, looking at both task and student performance. At both the task and student level, we attempted to identify variables that might impact performance. To do this, we looked to the ICT construct as defined by the Panel, variables already validated in previous PISA assessments (for example, task specificity and complexity of stimulus materials), and the observed performance of students in the feasibility study. The data were then analysed to see if those variables were supported or if others were suggested. The result of this process is a set of initial hypotheses about factors that might characterise ICT tasks and student performance on those tasks.
- 86. To identify task variables, we focused on basic questions related to task difficulty and timing as posed in the goals for the study. What kinds of tasks would students find



easiest and which would be more challenging? Which tasks would they be able to do quickly and which would take longer? And, could we identify task characteristics that might account for those differences?

- 87. It was hoped that the tasks could be analysed to investigate the extent to which performance could be mapped back to the six processes outlined in the ICT framework. Exploring whether or not these processes would yield distinct scales was clearly beyond the scope of this study. However, we felt that an initial analysis would provide some suggestion about the extent to which these processes help explain task difficulty and serve as indicators of ICT proficiency. As will be discussed, we were able to investigate how two ICT processes accessing and managing information were related to task difficulty and timing data.
- 88. To identify variables related to student performance, we began by looking at student background characteristics including gender, computer familiarity and attitudes about computers, and then investigated their relationships to task performance. In addition, we analysed the notes made by test administrators who observed and debriefed students in the retrospective interviews in order to define behaviours that seemed to distinguish proficient from not-so-proficient performance. These behaviours were aggregated into an observed ICT proficiency measure and the relationships among that measure, background characteristics, and task performance were investigated.

## **Cautionary Note**

Throughout the analysis process we were careful not to over-interpret the data that had been collected – a caution that readers should keep in mind as well. This feasibility study involved a limited number of participants: 34 in Australia. 55 in Japan and 29 in the United States, for a total of 118 students. Each group represented a convenience sample, constrained by factors such as the time of year (summer in Japan and the United States, when many students were on vacation or not easily accessible), available testing locations, and the limited amount of time available for recruiting. The achieved sample, as detailed below, over-represented high-ability students who were also highly familiar computer users. In addition, it must be remembered that the tasks included in this assessment were selected to serve particular purposes such as testing the feasibility of creating a simulated Web environment that could operate in more than one language. The tasks were not designed to assess the full range of skills and knowledge represented in the ICT construct as defined by the Expert Panel. Taken together, these factors limit the generalisability of the data yielded by the study. However, the study was able to accomplish the goal of providing information to inform future endeavours and exploring the types of questions a future ICT assessment might address.

### **Characterising the Student Sample**

90. The information supplied by students in the background questionnaire provides information about their academic performance as well as their experiences with computers. While 118 students participated in the study, 5 did not have complete response records for the background questionnaire and so are excluded from this part of



the discussion<sup>6</sup>. Of the remaining 113 students, 56% were females and 44% were males. As shown in Table 5, almost 68% of the students in the sample characterised their academic performance as excellent or above average. Only about 6% rated their performance as less than average.<sup>7</sup>

Table 5: Self-Reported Academic Performance

Rating	Frequency	Percent
Excellent (Mostly As)	22	19%
Above average (Mostly Bs)	55	49%
Average (Mostly Cs)	29	26%
Below average (Mostly Ds)	7	6%
Poor (Mostly Fs)	0	0%

91. The students in the study sample tended to be regular computer users with experience across a range of applications and uses. One hundred percent of the students said they had used a computer prior to taking the survey. Most students reported that they use a computer more frequently at home than at school or in some other location. Almost 70% use a computer at home a few times a week or more, with 42% reporting daily usage. Forty-three percent of the students reported using computers at school a few times a week or more while only 12% use a computer at school on a daily basis.

Table 6: Self-Reported Location and Frequency of Computer Use

How often do you use a computer	At home	At school	In other locations
Every day	42%	12%	2%
A few times a week	27%	29%	11%
Between once a week and once a month	12%	50%	31%
Less than once a month	11%	6%	37%
Never	7%	3%	20%

92. The types of computers and operating systems used by students in the study sample are shown in Table 7. Many more students reported using a Windows operating system (95%) than a Mac (20%) or other operating system (9% for Unix and 2% for Linux). In the follow-up questionnaire, 55% of the students said they use desktop computers, 15% use laptops and 28% use both.



36

<sup>&</sup>lt;sup>6</sup> The reason for this missing data is not entirely clear but notes in the administrator records file indicate Internet problems during these testing sessions.

See the Materials Annex for a full copy of the questions as presented in the background questionnaire and responses.

Table 7: Self-Reported Use – Operating System and Computer Type

Which of the following operating systems are on the computers you use?	Frequency	Percent
Mac	23	20%
Windows	107	95%
Unix	10	9%
Linux	2	2%
Don't know	3	3%

What types of computers do you typically use?	Frequency	Percent
Laptop	18	15%
Desktop	65	55%
Both	33	28%
Neither	2	2%

93. Not only do students in this sample regularly use computers, but they have been doing so for quite some time. As can be seen from Table 8, 68% of the students report that they have been using computers for three years or more. Only about 7% have been using computers for less than one year.

Table 8: Self-Reported Computer Use

How long have you been using computers?	Frequency	Percent
Less than one year	8	7%
One to three years	28	25%
Three to five years	32	28%
More than five years	45	40%

94. Given their high level of computer use, an interesting question is what are these students using computers for? As shown in Table 9, students reported that they use computers for a range of purposes including communicating over the Internet, going on the World Wide Web<sup>8</sup> and using applications such as word processors, spreadsheets and drawing tools. They do not use educational software with any regularity, with 48% of the students reporting that they never use this type of software. While not shown in this table, few students reported making Web pages or writing programs (60% said they never do those things and 25% engage in those activities less than once a month).

<sup>&</sup>lt;sup>8</sup> While not specifically addressed in the BQ, according to test administrators all of the Japanese students participating in the study regularly use a cell phone for network access and Internet mail (exchanging as many as 100 messages a day). Reportedly, most of these students can input information more rapidly using cell phone buttons than the computer keyboard.



Table 9: Computer Use - Purpose and Frequency

How often do you use	Internet	World Wide Web	Tools such as word processors	Educational software	Games
Every day	31%	35%	12%	0%	11%
A few times a week	25%	31%	27%	4%	25%
Between once a week and once a month	17%	16%	32%	12%	22%
Less than once a month	12%	15%	25%	35%	26%
Never	16%	3%	4%	48%	12%

95. This overview of the student sample quantifies the point made earlier in the cautionary note. According to the information they provided in the background questionnaire, the majority of students participating in the feasibility study were highly familiar computer users with excellent or above average academic performance. These characteristics of the study sample must be kept in mind as we investigate variables related to both task and student performance.

## **Characterising Task Performance**

96. One way to look at the feasibility study results is in terms of what they tell us about the various tasks included in the study instrument. The data were analysed to investigate task timing and difficulty, and to see if variables related to task performance could be identified. One question the Expert Panel was interested in investigating was the extent to which the six process variables were related to task performance. Therefore, the development team tried to ensure that the tasks developed for the assessment sampled across the range of processes. The process variables included across assessment tasks are listed in Table 10.



Table 10: Process Coding for the Feasibility Study Tasks

	Access	Manage	Integrate	Evaluate	Construct	Communicate
E-mail task	Χ	Χ				X
Database task		X	X			
Web abstract task				X		
Web search task	Χ	X	X	X	X	X
Simulation task	Х	Х	Х	Х	X	X

97. The question we were interested in investigating was to what extent are the underlying processes related to task difficulty? To explore this question we needed a sufficient number of task elements to obtain reliable information. While each 'X' on the table above indicates the presence of at least one task element for a particular process, the actual number of elements per process varies. Of the six processes, Access and Manage were the two with the largest number of scorable task elements in the study. Therefore, these two were analysed to provide a first look at how the processes might be related to task performance.

## Access and Manage Processes

98. As shown in Table 10 above, three of the tasks contained Access components: the e-mail, Web search and simulation tasks. Table 11 details the specific task elements coded as Access components, as well as the performance required in order for each element to be scored as correct. The performance requirements were defined during the test development process and then refined based on observations of student performance during the study.



<sup>&</sup>lt;sup>9</sup> Note that in some cases desired task elements or performance requirements could not be included in the analysis because the elements were not captured or scored by the delivery software.

Table 11: Access Tasks

Task	Task Elements and Performance Requirements		
E-mail Task	Selects email, clicks on 'Forward' and 'Send' icons		
Web Search Task, Part 1	Use of search – enters one or more search terms Clicks on 'Go' Clicks on one or more sites to open them Clicks on 'Back' icon at least 2 times (necessary in order to view any 3 opened sites)		
Web Search Task, Part 2	Clicks on links provided in e-mail to open Web sites Clicks on e-mail tab at least one time (necessary in order to open both sites)		
Simulation Task, Part 1	Makes one or more tables Makes one or more bar graphs Makes one or more line graphs		

99. The Manage tasks are similarly identified in Table 12. While the table has eleven tasks listed, only eight are included in the analysis. The three Manage components in the Web search task were not included because of technical difficulties that limited the amount of data captured.

Table 12: Manage Tasks

Task	Task Elements and Performance Requirements		
E-mail Task	Enters correct e-mail address in 'To' field Enters correct e-mail address in 'CC' field Enters e-mail text, mentioning 7:30		
Database Task	Uses 'Sort' functionality to identify CDs		
Web Search Task, Part 1	Bookmarks URLs and/or includes URLs in the email Enters correct e-mail address 'To' field Enter URLs into email		
Simulation Task, Part 1	Specifies variables to make one or more tables Specifies axes to make one or more bar graphs Specifies axes to make one or more line graphs Varies display type		

- 100. In order to analyse these components, students were given a single Access and a single Manage score, each based on the total number of correct task elements per process. For example, because there were ten Access elements, a perfect Access score equalled a ten. A perfect Manage score was an eight. Difficulty was then determined by looking at the mean percent correct obtained by the study sample.
- 101. On the whole, the Manage tasks were more difficult than the Access tasks. Students were able to get 76% of the Access tasks correct but only 65% of the Manage tasks. This may have been due to the fact that the Access skills included in the study



tasks were generally easy and posed no difficulty for most of the students in this sample. The Manage components tended to be more deeply embedded within the tasks. That is, students had to have completed a number of preliminary steps before reaching the Manage component of a task. An additional explanation for the difference in difficulty may lie in the performance requirements that were specified for scoring. For the Access tasks, the requirements generally included the ability to collect or retrieve information with no judgement about the correctness of that information. In order to demonstrate the ability to organise and manage information, students often had to go beyond 'doing something' and complete a specified action correctly.

- 102. Another difference found between Access and Manage scores was that, although both were reliable, only the Manage score correlated with student variables such as computer familiarity<sup>10</sup>. That is, students with limited computer experience (relative to the skills of the sample group) were much less successful in completing the required eight Manage tasks than students in the higher groups (42% correct for the low group versus 69% for the medium and high groups). The Access score did not reveal a similar relationship to these student variables, perhaps because the Access tasks were easier overall and had a limited response variability in this sample.
- 103. Based on these initial findings for Access and Manage, we would expect that the process components might well function as variables that are related to task difficulty. This expectation would require additional testing using an instrument with a greater number of task elements per process in addition to a larger and more varied student sample.

## Sections by Difficulty and Time

### **INTRODUCTION**

- 104. To learn more about task performance, we also looked at the data in terms of more general task type. Four types of tasks were developed for the study: basic technical tasks, short application-based tasks, and longer web- and simulation-based tasks. Each section of the test included tasks of a particular type. Therefore, analyses were run to look at how the test sections varied in terms of difficulty and time. Task variables were explored to see if they might be related to variations in how the tasks functioned. These variables included the number of processes required to complete a task, how directed or open-ended the task was, and characteristics of the interface or task presentation. The sections are introduced below in terms of some of these variables and then discussed in greater detail with the accompanying data.
  - Section 1 contained the background questionnaire and a set of basic technical tasks that required students to scroll, copy and paste, etc. These tasks were not application specific, but intended to represent foundational mouse/touchpad and keyboard skills. It was expected that most students would find these tasks very easy and be able to complete them quickly.

<sup>&</sup>lt;sup>10</sup> See 'Characterising Student Performance' for an explanation of how the computer familiarity index was created.



\_

- Three short scenarios made up Section 2: the e-mail, Web abstract and data base tasks. These tasks required more knowledge of application-related functionality; however, they were also very structured, as each presented a directed sequence of steps for students to follow. In addition, each task required the application of only a few of the ICT processes: three for the e-mail task, two for the database task and one for Web abstract task (see Table 10).
- Section 3 presented the Web search task. Part 1 was somewhat open-ended, presenting a task to be completed by accessing information in a simulated Web environment. It required a range of processes, from access to evaluate, and those processes varied in terms of difficulty. It was expected that this would be the more challenging part of this task. In Part 2 of the Web search task, students were directed to use two provided Web sites to make a purchasing decision. This was more structured than Part 1, although students had to pay close attention to the purchase criteria in order to evaluate the two books and make the correct choice.
- Finally, Section 4 presented the simulation task. This was expected to be the most difficult task for a number of reasons. First, the simulation was situated in a novel, open-ended environment in which students needed to navigate without explicit user instructions. In addition, this task required students to apply all of the six ICT process components. To successfully complete Part 1, students needed to access data by running experiments, manage the results by displaying them in a table and/or graph, integrate information across a number of experiments and representations in order to evaluate the results, and then communicate their conclusions in response to an open-ended question. Part 2 focused students' investigations on the selection of the best representation for the experimental data, requiring them to construct additional displays as needed, and then evaluate how effectively each illustrated the relationship among the variables in the experiment. Part 3 was the most structured of all, consisting of a multiple-choice question to summarize what had been learned through the simulation.
- Before discussing the specific results for each of the test sections, a general point should be made regarding the difficulty data. We have reported difficulty in terms of the percent of students getting each task correct. Defining what constitutes a correct response is a challenge in any performance-based assessment. This challenge is compounded for an assessment in a relatively new domain, such as ICT literacy, where there are few agreed-upon standards for what constitutes evidence of ICT skills and knowledge. We took the following approach for the feasibility tasks. For each task, a list of steps or elements was developed and a preliminary set of completion criteria for each element was defined. These criteria were sometimes adjusted based on observation reports and/or performance data. So, for example, the initial list of required actions for a correct response on part 2 of the simulation task (the multiple-choice question about data representation) included identifying the line graph as the best representation for the data. However, the data revealed that even for this particularly able student sample, correctly selecting the line graph had the very low frequency of 0.23, or less than chance for this 3-option multiple-choice question. The observation reports also indicated that students found the interface for selecting variables difficult to use. For many students the line graph that they were able to create was not, in fact, the best representation of the data. Taken together, these findings caused us to revise our "correct" criteria for this



preliminary analysis of the difficulty results. Such adjustments to scoring criteria are to be expected and as items are piloted with larger and more varied student samples those judgements can be made with greater confidence.

#### **DIFFICULTY AND TIMING DATA**

106. The difficulty and timing data based on the study results are shown in Table 13 below. One caveat must be shared about the data for Part 1 of the Web search task and the resulting Section 3 totals. As a result of a technical problem, portions of the performance records for this task were not collected for the first 54 students tested. We had no record from the computer of the search term entered or the URLs that students included in the e-mail they sent indicating their book choices. Fortunately, we were able to use information from the observation reports to reconstruct these results for 26 of the students. However, due to these difficulties, the data for this task is heavily patched and should be interpreted with particular caution.



Table 13: Difficulty and Timing by Section

	Difficulty	Time (in	minutes)
	Percent correct	Maximum	Mean
Section 1: Technical tasks			
Section Total:	90% <sup>11</sup>	20	5
Section 2: Short scenarios			
E-mail task	25%	17	3
Web abstract task	24%	18	6
Database task	67%	4	2
Section Total:		39	11
Section 3: Web search task			
Part 1 (book search)	63%	32	10
Part 2 (view both book choices)	78%	15	4
Section Total:	52%	47	14
Section 4: Simulation task			
Part 1 (experiment/draw conclusions)	20%	33	8
Part 2 (create additional displays)	66%	32	2
Part 3 (select correct summary )	51%	6	1
Section Total:	23%	71	11

N = 113 for section 1 and 118 for the sections 2-4 with the exception of parts 1 and 3 in the simulation task which had an N of 60.

#### SECTION 1 - BASIC TECHNICAL TASKS

107. Ninety percent of the students were able to successfully complete the basic technical tasks presented in Section 1, which is not surprising given the level of computer expertise possessed by the students in the sample. In addition, students were able to complete these tasks very quickly with a mean time of 5 minutes. We would expect that implementing a number of interface revisions based on observation reports and student comments (for example, clarifying performance expectations for the scrolling tasks and providing expected keyboard short-cuts for familiar actions such as saving a file) would improve task performance even more.



<sup>&</sup>lt;sup>11</sup> This score excludes performance on the typing and vertical scrolling tasks. Difficulties with data capture and interface on the scrolling task and discrepant scores for the Japanese students on the typing task due to character set and keyboard issues negatively impacted the performance scores in ways judged to be irrelevant to the skills being measured.

<sup>&</sup>lt;sup>12</sup> Because Section 2 contained the most heterogeneous set of tasks, it was expected that student performance would differ across tasks. Therefore time and difficulty results are only reported separately for each individual task and no total is reported for this section.

108. While most students in this sample found these basic technical tasks extremely easy, even in this population there were students who had some difficulty. It would be helpful to administer these tasks again with students who had little or no computer experience to determine if they provide useful information about the basic technical skills of a less experienced population. Such a test would also be useful for determining if these tasks, or ones like them, might function as a screener to identify students who would most likely be unable to complete a computer-based assessment because of limited technical skills.

#### SECTION 2 – SHORT SCENARIOS

#### The E-mail Task

109. To receive a correct score for the e-mail task, students were initially required to complete the actions listed in Table 14 below.

Table 14: Percentage of Students Correctly Completing Elements of E-mail Task

Element	Percent Correct
Select the correct e-mail	77% on first click
	87% within two clicks
Correctly click on the 'Forward' icon	90% of the students clicked on
	'Forward' before clicking on any other
	buttons (including 'Next')
Enter the correct 'TO' address in the e-mail	75%
response	
Enter the correct 'CC' address in the e-mail	28%
response	
Type a message that, at the minimum, included	46%
the time (7:30) that Jamie would be picked up	
Correctly click on the 'Send' icon	68%
Task Total	25%

Several adjustments to the scoring were made based on the observation reports. Initially it was expected that students should easily locate the correct e-mail to open in the in-box. Students were looking for an e-mail from Tessa Baxter with information about a party she was having. However, based on student actions and comments it became clear that the first message in the in-box, which was from Tessa Baxter, contained a subject line that was general enough ("Heard this?") to be a plausible response. Therefore, many capable students were clicking on it before noticing the correct e-mail later in the list ("Party at my place"). As a result, students were given credit for a correct response on this part of the task if they clicked on the correct e-mail on the first or second try.

110. A second adjustment concerned the 'CC' address. Students were clearly not familiar with the idea of copying another person by including an address in the 'CC' field.



Only 28% completed that action as expected. The data records revealed that some students treated this field as a subject line, some entered their message on the 'CC' line, and many left it blank. This observation was confirmed in the retrospective interviews where many students inquired what the 'CC' meant. Because of this confusion, the requirement that the 'CC' information be correctly entered was dropped from inclusion in the total task score.

Even with these adjustments, only 25% of the students were able to correctly complete the e-mail task. Why such a low percentage? One explanation may be related to frequency of use. While 85% of the students stated on the follow-up questionnaire that they had used e-mail, test administrators indicated that during discussions students reported using instant messaging, chat rooms, or text messaging on their cell phones more frequently than they use e-mail. A second explanation, revealed through the observations, is that many students did not read the directions displayed on the left side of the screen and/or did not notice as those directions changed during the task. This resulted in entering incorrect information, not entering information at all (for example, not typing a response of any kind before sending the e-mail), or clicking on incorrect icons. Students did not spend much time on this task (on average, about 3 minutes), further suggesting that they were able to hurry through it. It is therefore possible that the results for this task underestimated students' abilities. The challenge for future iterations of tasks such as this one is to refine the display and interface to ensure that students notice and read directions or to revise the task and scoring criteria, allowing students to explore their way through the task rather than follow an explicit set of instructions.

#### The Web Abstract Task

112. Two correct responses were required for a correct score on the Web abstract task – selecting a reliable and an unreliable site with relevant information based on the abstracts provided. About 50% of the students successfully completed either of those tasks, as shown below. Only 24% completed both parts correctly.

Table 15: Percentage of Students Correctly Completing Elements of Web Abstract Task

Element	Percent Correct
Correctly choose one of the reliable sites	52%
Correctly choose one of the unreliable sites	47%
Task Total	24%

113. On average, students took six minutes to complete both parts of this task, a reasonable time given that they had to scan the abstracts for ten Web sites and type explanations for their responses. We believe students had difficulty with this task for several reasons. First, based on students' responses to the open-ended 'Explain why you chose this site' question, it was clear that many students interpreted 'reliable' and 'unreliable' to mean 'relevant' and 'irrelevant'. This interpretation was also noticed and reported by test administrators. As a result of this, some students focused only on the content of the sites and did not consider the source. In addition, very few students reported that they use URLs to help them judge whether a source is reliable or not and even fewer know the meaning of extensions such as '.gov' or 'edu.'



#### The Database Task

- 114. Of the short scenarios, the database task was the easiest. In order to receive a correct score on this task, students had to enter the correct number of jazz CDs listed in the database. Sixty-seven percent of the students were able to do this and most were able to complete the task quickly, with a mean time of 2 minutes. Of those who answered the question correctly, only 12% used the 'Sort' function. According to the observation reports, many students simply did not notice the 'Sort' button on the screen. In fact, when the 29 U.S. students reviewed the screens during the retrospective interview and were asked, "Do you see another way you might have completed this task?" almost half noticed the 'Sort' icon and were able to use it correctly.
- 115. The students who did not sort were most likely able to find the correct answer because the list of 100 CDs was short enough to scroll through and because the two columns with the critical information (type and price) were adjacent in the table, making it easier to scan and count. Revising the task so that the database was longer and more complex might encourage students to find and use the sorting functionality.
- 116. Of those students who got the question wrong, 12% answered either '14' or '15.' Counting the total number of CDs costing \$5.00 or less but disregarding the requirement to locate jazz CDs in that price range results in an answer of '15.' Making the additional mistake of not counting the one CD that costs exactly \$5.00 results in an answer of '14.' This analysis suggests that for some students who scrolled and counted, the difficulty in this task lay in matching all the specified conditions in the question.

#### SECTION 3 – WEB SEARCH TASK

117. To receive a correct score for the Web search task, students were required to complete the task elements listed below.

Table 16: Percentage of Students Correctly Completing Elements of Web Search Task

Element	Percent Correct
Part 1	
Enter a search query and click on 'Go'	96%
Have last query contain 2 or more relevant key	
words or forms thereof (e.g., digital,	76%
photography, book, beginner)	
Open 3 or more sites	88%
Include one or more correct URLs in e-mail	51%
Total for Part 1 elements	63%
Part 2	
View both books listed in friend's e-mail	78%

Task Total	52%

118. For almost all of the students in our sample, searching on the Web is a familiar activity. There were only a few (3%) who reported never using the Web. Despite their familiarity, only about 52% of the students were able to correctly complete both parts of



this task. In addition, students spent, on average, more time on this task than any other in the assessment.

- 119. In Part 1, students were able to complete each element with varying degrees of success. Ninety-six percent of the students entered a search query and students opened an average of 4 pages per search, suggesting that they were engaged in the search task. However, only 51% of the students were able to locate and list one or more correct URLs. This task element required the integration of multiple processes and fluency in both ICT and reading skills. Some students did not persist in accessing and evaluating sites, as they failed to scroll through the list of hits, go beyond the first page of hits, or open enough sites to make a judgement about which included suitable information. Based on observation reports, it was clear that some students used prior experience with other search engines as a basis for limiting their search. For example, some students said they did not go beyond the first page of hits because the search engine they typically use does a good job of giving them the most relevant hits early in the list.
- 120. In the second part of the task, students were supplied with links to two books and asked to determine which best met the specified criteria. Seventy-eight percent of the students viewed both the recommended books, demonstrating that they could use the provided links to access the sites and navigate back and forth between the simulated email and Web applications. In order to complete the task successfully, students needed to click on a 'Check Availability' button on one Web page, and then click on a 'Buy Now' or 'Add to Shopping Cart' button to indicate their book choice. Unfortunately neither action was included in the performance record because the software could not detect clicks within an HTML page such as the ones students opened via the links provided, so we do not have data about the extent to which students were able to use the specified criteria to evaluate the choices provided.

#### **SECTION 4 – SIMULATION**

121. The simulation task was, as expected, more difficult than any other task in the study. Only 17% of the students were able to successfully complete all three parts of the task. The mean time for task completion was 11 minutes. The definition of successful completion was particularly challenging in this task, given its open-ended nature. Because there were two multiple-choice questions associated with this task, the simplest option would have been to assume that correct responses to those questions equalled successful completion. However, the software captured a great deal of additional information about how students actually went about completing the task. Therefore, we included additional actions in our definition of "successful" although we chose to score those actions leniently (for example, scoring only whether or not students created a graph or table, not whether that display was correctly represented). Given a larger sample of students, definitions such as this one could be validated with additional performance data. However, for our purposes, successful completion was defined as a composite of the following task elements.



Table 17: Percentage of Students Correctly Completing Elements of Simulation Task

Element	Percent Correct
Part 1	
Conducts one or more experiments	99%
Creates one or more displays	96%
Enters a response to the question	96%
Enters a correct response to the question	68% correctly related time or temperature to carbon dioxide production 42% correctly related both time and temperature to carbon dioxide production
Total for Part 1 elements	20%
Part 2	
Creates additional displays	66%
Part 3	
Selects correct explanation from multiple- choice list	51%
Task Total	23%

- 122. In Part 1, students had to conduct one or more experiments, create one or more displays, and enter a response that mentioned the three correct variables and the fact that levels of carbon dioxide rose. In Part 2, students had to create at least one additional display and in Part 3 they had to select the correct multiple-choice option. Clearly, with different scoring criteria, performance profiles on this task would change. Given more time, the impact of different scoring criteria could be evaluated and both tasks and scoring programs could be refined to optimise measurement.
- 123. We assumed that one reason the simulation task would be difficult for students was its open-ended nature. Students were presented with a range of actions, which they could complete in a sequence of their choosing. They also had to decide when they had accumulated a sufficient amount of information to answer the presented question. This task also required the application of the full range of ICT processes, from Access through Communicate. Despite these factors, students in this sample were able to explore the interface and figure out how to run experiments. Ninety-nine percent of the students ran one or more experiments and, on average, students ran 17 experiments in Part 1. They were also able to create representations of the data. Fifty-eight percent of the students made one or more tables, 50% made at least one bar graph and 45% made one or more line graphs. However, many students had difficulty using the simulation to explore the relationships among time, temperature, and volume in some systematic way. In the future, a similar simulation task, perhaps in a different context, with fewer variables to manipulate and an improved graphing interface might be explored to see if it allowed more students to demonstrate and apply ICT skills to this type of problem-solving task.



## **Characterising Student Performance**

## Introduction

124. A second way to look at the data from the feasibility study is to view the results at the student level, that is, to look at student characteristics and behaviours and see how they are related to variations in task performance. We had three types of evidence available to inform this investigation. First, performance data collected by the software provided a record of student actions and responses to questions. Then, data from the background questionnaire provided information about student variables such as gender and computer familiarity. Finally, we collected observational data for a significant number of the students participating in the study. As shown in the figure below, we found that these three types of evidence interacted, with each informing the interpretation of the others.

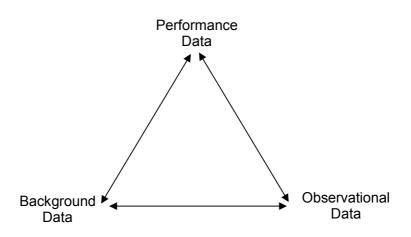


Figure 1: Sources of Evidence for Student Performance

125. For the purposes of this analysis, we have focused on the relationship between background characteristics and performance as well as the relationship between what we have called "observed ICT proficiency" (based on observational data) and performance.

## **Background Characteristics and Performance**

126. We looked at three variables associated with background characteristics: gender, computer familiarity, and attitudes about computers. However, because the feasibility study included three different convenience samples, comparisons across these groups



50

<sup>&</sup>lt;sup>13</sup> We have observation records for 70% of the English-speaking students and 27% of the Japanese students. Analysis in this section is based on the English-speaking students only because, unfortunately, there was neither sufficient time nor resources to translate the Japanese reports.

do not have much meaning. A few highlights are included here as they suggest patterns that might be explored in subsequent studies conducted on a larger scale.

#### **GENDER**

127. There were few differences in performance based on gender found in this study. The greatest differences were found in the Web search task. Females spent, on average, 3 more minutes on the task and 60% of them completed it correctly compared with 50% of the males.

#### **COMPUTER FAMILIARITY**

- The computer questions in the background questionnaire (BQ) were collapsed to create two indices – a computer familiarity index and a computer attitude index. The computer familiarity index was generated based on the BQ questions that asked how often and in which environments (home/school) students use computers, and how often they used computers for specific purposes (homework/chat/games/surfing the Internet). The attitude index was based on the final four questions in the BQ that asked how important students think it is to learn about and use computers as well as how comfortable and confident they feel using computers. These BQ items were analysed using classical test theory (CTT) reliability analysis 14 that yielded a scale with Cronbach's alpha of about 0.8 for familiarity and slightly less than 0.7 for attitude towards computers. This means that each group of items could be meaningfully summarised with a single number, so that both a computer usage/familiarity index and a computer attitude index could be formed. Because the computer attitude index was found to be less reliable than self-reported computer familiarity, only the latter was used to analyse results. Future efforts might focus more on the impact of interest and attitudes, given their inclusion in the definition of ICT literacy.
- 129. The computer familiarity index was used to define three groups of students those with low, medium, or high computer familiarity. The data were then analysed to see if any there was any relationship between this index and performance on the tasks in the study instrument Computer familiarity showed some relationship to both difficulty and time, particularly on the Web search task. Highly familiar computer users were much more likely to complete the Web search task successfully. Eighty-five percent of them completed that task correctly as opposed to 63% of the moderately familiar and 45% of the low familiarity group. As might be expected, highly familiar computer users tended to complete most tasks more quickly than less familiar users. On average, the highly familiar students completed the Web search task in 9 minutes as opposed to 11 minutes for the low familiarity group. Even greater variation in timing results was found for the short scenarios. The highly familiar students spent 9 minute completing the three tasks in this section, whereas the low familiarity group spent 12 minutes.



<sup>&</sup>lt;sup>14</sup> Classical test theory methods as well as exploratory data analysis rather than IRT based methods have been used in this first analysis of the feasibility data. Reports of reliability coefficients such as the Cronbach alpha values for computer familiarity and attitudes towards computers should be viewed as very preliminary numbers due to the limited sample size and population coverage.

# Observed ICT Proficiency

- 130. One important source of information in the feasibility study was the thorough set of observation records collected for the retrospective interview students. We were interested to see if we could use these observations to identify patterns of actions or variables that could help us identify and characterise groups within our sample. If so, these groups could then be compared in terms of variables such as task performance and computer familiarity.
- 131. We began by reviewing the observation records and test administrator reports for the purpose of identifying actions and behaviours that seemed to differentiate student performance. The behaviours listed below made up the observed ICT proficiency and were coded using information in the observation reports. These included some observer judgements (such as the ease with which students completed tasks), some actions that were not captured by the software, and some that we did not have time to extract from the student records.

## Observed ICT Proficiency

Did the student easily complete each of these tasks?<sup>15</sup>

Section 1 tasks

E-mail task

Web abstract task

Database task

Web search task

Simulation task

Use of 'advanced' techniques

Keyboard short cuts

Right clicking

Web search queries (use of Boolean operators, etc.)

#### Section 1

Tries right click to copy – adapts copy method when right click doesn't work

Web abstract task

Uses URLs and/or extensions to evaluate sites

Database task

Uses 'Sort' functionality and/or clicks on column headings to sort

<sup>&</sup>lt;sup>15</sup> Observers were asked to indicate this judgement for each of the tasks in the assessment. Test administrators from ACER summarised their approach to this judgement as follows. "The Test Administrators coded students who were hesitant, distressed or anxious in completing a task as 'No' for these items. Test Administrators coded a 'Yes' if the student was confident in completing the task and demonstrated facility in the task. Whether the response was correct or incorrect was not taken into account." For consistency, the coding for U.S. students was reviewed to ensure these same criteria were applied for all students.



#### Web search task

Gathers information about book availability in Part 2

#### Simulation

Figures out the experiment interface (selects time, temperature and clicks on 'Run Experiment')

Figures out table interface/able to create a table

Figures out graph interface/able to create a graph

Experiments in systematic way (for example, holds temperature constant and runs experiments across multiple rising times or vice versa)

- 132. Each behaviour was coded as either a 1 (present) or 0 (absent) for the 44 students and those numbers were totalled to yield the observed ICT proficiency score. The purpose for experimenting with this observed ICT proficiency measure was to see if we could identify groups of students at the extreme ends of the sample those who moved through the tasks with great ease and success versus those who struggled and were unable to complete many tasks. Once total "scores" were calculated, three groups were identified those with a high, medium, or low observed proficiency.
- 133. We then analysed the data to see if these groups, in fact, performed as one would predict on the ICT tasks and if their reported levels of computer familiarity also matched what one would expect. In many cases, task performance matched observed ICT proficiency. That is, students classified in the high proficiency group, based on observations, tended to do better on the assessment tasks than those in the low group. For example, on the simulation task the high proficiency group performed almost 3 times as many experiments as the low proficiency group (a mean of 20 experiments versus 7). All the high proficiency students made more than one representation of the data (that is, they used both a table and graph or used both a bar graph and line graph). Only 42% of the low proficiency group used more than one form of representation for the data. Finally, 70% of the high proficiency group correctly answered the final question about the relationship among the variables as compared with 30% of the low proficiency group.
- 134. Table 18 shows the correlation between the observed groups and both the computer attitude and computer familiarity scale scores. It is evident that all three measures are related positively, indicating that the variance between students' observed ICT proficiency and students' computer familiarity and attitudes can be captured by these different measures with slightly varying foci.

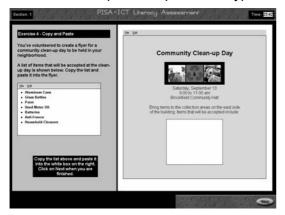


Table 18: Correlations among Observed Proficiency, Computer Attitude, and Computer Familiarity

		Observed ICT Proficiency	Computer Attitude	Computer Familiarity
Observed ICT	Pearson Correlation	1.000	.394 (*)	.514 (**)
Proficiency	Sig. (2 tailed)		.011	.001
Proficiency	N	41	41	41
Computer	Pearson Correlation		1.000	.543 <sup>(*)</sup>
Computer Attitude	Sig. (2 tailed)			.000
Attitude	N		113	113
Computer	Pearson Correlation			1.000
Computer Familiarity	Sig. (2 tailed)			
ranimanty	N			113

135. Finally, the observation reports suggest additional behaviours that might be further explored with larger samples in the future. One such set of behaviours was mentioned by a number of test administrators. These test administrators noted that some students were more willing than others to explore the simulated environments presented in the assessment. If these students got stuck, they were able to apply prior knowledge or use information available in the tasks to figure out how to continue. They tried a number of different approaches or clicked on the 'Help' function. Some students even demonstrated the ability to learn new approaches based on information presented in the tasks (for example, several students noticed keyboard short cuts identified on the 'Edit' menu and used them for the first time to complete the copy/paste task). Other students seemed unable to adapt their approach if familiar interface elements or functionality were not available. When they got stuck, these students often gave up and asked the test administrator what they should do next.

136. A simple example of this type of adaptable behaviour can be found in the



copy/paste task in Section 1, which contained an inadvertent discrepancy in functionality. The only way to copy the text was to either select 'Copy' from the pull-down 'Edit' menu or highlight the text and then use the CTRL+C keyboard shortcut. However, it was possible to paste the text into the flyer using a right mouse click. One third of the students observed in Australia and the U.S. demonstrated an interesting and consistent response to this situation. They tried to right click to copy the text, found that the right click functionality would not work and then easily switched to a



<sup>(\*)</sup> Correlation is significant at the 0.05 level (2-tailed)

<sup>(\*\*)</sup> Correlation is significant at the 0.01 level (2-tailed)

second method to copy (either CTRL+C or using the pull-down menu). Once the text was copied, they used the right click functionality to paste it into the new window.

- 137. One interpretation of this adaptability is that these students had a number of different strategies at hand to complete the task. They knew more than one way to accomplish what they needed to do and had no problem switching approaches. Many of these adaptable students displayed similar behaviours across a variety of the presented tasks. Identifying clusters of behaviours such as this one is another way to develop profiles of the skills and knowledge possessed by students at various points along a continuum of ICT proficiency.
- 138. It is possible that an ICT literacy assessment should include tasks that focus on measures of adaptability or fluency. We know from the literature, for example, that fluency is an important component of reading. As students develop fluency in decoding, they are able to focus on understanding and interpreting text. It may be that a similar fluency in ICT allows students to focus on the problem they are trying to solve using technology rather than on aspects of the technology itself. Regardless of whether or not this particular hypothesis is substantiated in future investigations of ICT literacy, we believe the study has illustrated the importance of using multiple sources of data including background characteristics, performance data, and observations to investigate the construct and its component skills.

# The Follow-up Questionnaire - Student Responses

- 139. Students participating in the feasibility study were asked to complete a 34-item paper and pencil questionnaire once they had finished testing. Those students selected for the retrospective interviews completed this questionnaire prior to discussing the test and their performance with test administrators. The questionnaire is included in the Materials Annex and a summary of student responses is listed in Appendix C at the end of this report.
- 140. Students were first asked for general impressions of the test they had just completed. Ninety-five percent of the students felt that it was easy or very easy to navigate around the test and 94% found the overall look of the screens to be appealing or very appealing. Students' assessment of the difficulty level of the test was more varied. Forty-five percent of the students felt that, overall, the test was easy or very easy; 36% felt it was of average difficulty and 22% found it difficult or very difficult.
- 141. Eighty-one percent of the students felt that survey allowed them to demonstrate their computer skills and knowledge either well or very well. Fifty-three percent reported that many or most of the things they typically do using computers were included in the test, but 42% said that only a few of the things they typically do were included. When asked for examples of things they normally do on the computer, either in or outside of school, that they felt should be included in a test like this, many students listed word processing. Other popular suggestions were specific applications (Word, Excel, PowerPoint), instant messaging, games, and downloading music.
- 142. Students were asked a number of questions about the major tasks included in the test. Their responses are presented as follows, organised by task:



## E-Mail Task

143. Sixty-two percent of the students felt that the e-mail task was interesting or very interesting. It was, in fact, the most popular task in the test with the Web search task coming in second, the database task third and the simulation last. Ninety-one percent of the students thought that the directions for the task were clear and 95% thought they had enough or more than enough time to complete the task. Although only 25% of the students correctly completed the task, 93% felt it was easy or very easy. Eighty-five percent of the students reported that they had used e-mail programs before taking this test. When asked if the functionality of the simulated e-mail program in the test interfered with their ability to complete the task, 82% said not at all or very little. Eighty-nine percent said that the way the program looked (different icons, etc.) interfered not at all or very little with their task performance. And 83% felt that using this e-mail program to communicate with friends and arrange a social event somewhat or very well reflected the way they typically use e-mail.

## Database Task

144. Fifty-one percent of the students felt that the database task was interesting or very interesting. Ninety-four percent thought the directions were clear and 92% felt they had enough or more than enough time to complete the task. Ninety-three percent felt that the task was easy or very easy. Only 47% of the students reported that they have used database programs previously. Of the students who had used database programs before, 87% felt that any differences in functionality between this simulated database program and what they have used in the past interfered not at all or very little with their ability to complete this task. Ninety-five percent said that they way this program looked interfered not at all or very little with their ability to complete this task. And 93% felt that using this database to sort information reflected the way they typically use databases somewhat or very well.

#### Web Search Task

145. Fifty-five percent of the students responding found the Web search task interesting or very interesting. Seventy-nine percent thought the directions were clear and 89% thought they had sufficient time to complete the task carefully. Students felt this task was more difficult than the short scenarios (e-mail and database tasks). Sixty-seven percent found it easy or very easy while 33% thought it was difficult or very difficult. Ninety percent of the students report that they have used search engines to search the Web. When asked if the functionality of the simulated Web environment used in the test interfered with their ability to complete the task, 92% said not at all or very little. Ninety percent said that the way the program looked (different icons, etc.) interfered not at all or very little with their task performance. And 90% felt that the way they were asked to use the simulated search engine reflected somewhat or very well they ways in which they typically use search engines.



## Simulation Task

146. The simulation task was the most negatively rated of the four tasks included in the follow-up questionnaire. Forty-nine percent of the students found it interesting or very interesting and only 55% thought the instructions were very or somewhat clear. Eighty-four percent of the students thought they had enough or more than enough time to complete the task. Fifty-seven percent found it difficulty or very difficult, much higher than the 7% for the e-mail and database tasks and the 33% for the Web search task. Students were also less familiar with this type of task. Sixty-four percent said they had never used simulations or similar learning tools before. Of those who had, 71% said that any differences in functionality interfered not at all or very little with their ability to complete the task presented in the test and 80% felt that the way the program looked interfered not at all or very little. Eighty-three percent of those who had used simulations before thought that the simulation task in the test reflected the way they typically use such tools somewhat or very well.

## **Summary of Results**

- 147. At the beginning of this section, we said that the analysis of results from the feasibility study would hopefully yield a set of initial hypotheses about factors that might characterise ICT tasks and student performance on those tasks. A number of task and student performance characteristics have emerged from the data and areas for further exploration have been identified.
- 148. The study data suggest that variations in task performance may be related to variables such as the following.
  - The processes in which students must engage in order to complete the tasks.
     We found initial evidence that task difficulty may vary by process (with Manage tasks being more difficult than Access tasks) and suggest that the number of different processes required to complete a task may also impact difficulty.
  - The degree to which the task and/or the simulated environment is directive versus open-ended.
  - The amount of information presented, its organisation, and how it could be accessed (for example, the limited amount of information in the database task made that task easy to complete without the benefit of more advanced functionality).
  - Requisite background knowledge of ICT terminology and functions (for example, knowing what a URL is or understanding the function of the 'CC' field in the email task).
  - Interface and functionality issues.
- 149. Student performance is clearly related to computer familiarity as defined by amount and type of computer use. A set of behaviours including ease of task



completion, use of more advanced functionality, and the ability to navigate an unfamiliar interface and application also appear to be related to performance. There is a suggestion, based on the observational data, that issues related to fluency or adaptability might warrant additional exploration, particularly with a sample of students that represents a wider range of ICT proficiency than the one included in this study.

150. Finally, the study results illustrate the importance of including multiple types of evidence to inform interpretations of behaviour and guide future test development efforts. Extensive observational data collection is necessary to set the stage in the usability testing phase and additional observations from the larger more varied pilot would allow developers to extract and aggregate relevant variables from the extensive data records collected by the delivery system.



# **SECTION 4: LESSONS LEARNED**

151. In this final section of the feasibility study report, a number of lessons learned are summarised. We discuss the lessons relative to the original goals for the study and in terms of their implications for moving forward with assessments of ICT literacy

# Simulated Applications and Environments

- 152. One goal of the study was to learn more about the use of simulated applications and environments in an assessment of this type. Lessons learned include the following:
  - In general, the simulated applications developed for the study worked quite well and students were able to use them. Creating applications that were similar to existing applications across a variety of platforms appears to have been a reasonable compromise so that students were not unfairly advantaged or disadvantaged based on familiarity with existing applications.
  - Novel environments (such as the one created for the simulation task) provide useful information about the extent to which students can explore and ultimately navigate their way through an unfamiliar application.
  - The limited functionality necessary in simulated applications seems to allow more proficient students to demonstrate abilities to adapt and use a variety of approaches to complete a task. Whether the negative impact on the performance of less proficient students is too great should be explored with a more diverse sample of students.
  - For the simulated Web task, the broad universe concept (creating a universe that was relatively broad but only a single level deep) allowed the instrument to be constructed in two languages within the time and budget constraints of the project. The preliminary data indicates that the Web search task worked equally well in both English and Japanese. Based on information from the follow-up questionnaires, students did not seem to mind using web pages with non-functioning links and they, in general, did not seem to think that the design interfered with their ability to complete the task.

#### CAPTURING PROCESS INFORMATION

The use of simulated applications successfully allowed us to capture a range of student actions. Capturing process information, rather than simple right/wrong responses, informed our understanding of performance and proficiencies, as illustrated in the following example from the simulation task. Two students failed to correctly answer the question posed in the task – "How is the amount of carbon dioxide produced by yeast affected by the temperature of the dough mixture and the amount of time it is allowed to rise?" If they received a simple right/wrong score, their performance would be rated similarly – neither was able to successfully complete the task. However, process information revealed a very different pattern of actions. One student selected a single rising time, tried to run an experiment, and then typed a response. The other student ran a number of



- experiments, systematically varying temperature or time, and displayed the resulting data in both a table and a bar graph. While both final responses would be considered wrong, the two students demonstrated very different sets of skills as supported by evidence from the process information collected.
- Despite positive initial results, issues related to the capture and scoring of process-related behaviours still require additional work. In order to meet the development schedule for the study instrumentation, ETS leveraged existing software that it had developed for previous computer-based assessments. This software came with some limitations in terms of what could be captured and scored. For example, while the test developers were interested in capturing detailed information about methods used to copy and paste text in various tasks (such as menu commands and keyboard shortcut), the delivery system could only determine the text that was copied or pasted, not the method by which it was done. Such low-level actions are usually handled automatically by the operating system and will be difficult and expensive to monitor and record in most test delivery systems. Observational information should provide useful guidance about the importance of capturing process behaviours that pose similar technical challenges.

## Translation and Adaptation

- 153. Because of the time constraints under which the feasibility study was conducted, we needed to simultaneously develop, adapt, and translate tasks. This resulted in inefficiencies such as having to translate multiple versions of some tasks as they were revised during programming. Based on our experience, we can vouch for the importance of a design model such as the following:
  - 1. Develop the initial interface/functionality for the test and the simulated applications.
  - 2. Send the proposed interface for country review and make revisions and adaptations based on those reviews.
  - 3. Develop a functional prototype for a task or set of tasks in English.
  - 4. Usability test the prototype and then send it for country review.
  - 5. Make revisions based on usability testing and reviews this becomes the finalised English version.
  - 6. Author the task(s) into the test delivery system. (This system will also be used to manage the translated versions of the tasks.)
  - 7. Develop translation guidelines and send the authored task(s) for translation. Translated output will be XML or some other data that can be fed directly to the delivery system.
  - 8. While translation is occurring, develop and finalise next set of tasks.
- 154. A plan that allows for a completed English version prior to translations minimises the need to go back for translations of pieces of items that might be overlooked (e.g., popup messages, help screens, etc.). While this is probably not the fastest way to get things done (ideally the English version of the software could be developed while the translations were being completed), this method is most likely to be successful.



## **Usability**

- The study illustrated the importance of iterative usability testing. By conducting a series of small-scale tests with 5-8 students as tasks are being developed, problems with the interface and directions can be identified and fixed prior to larger scale testing.
- The development schedule needs to allow for software testing in participating countries. We were able to do this in Australia and had no technical problems. However, scheduling constraints prevented a thorough test in Japan and issues did arise related to Internet connectivity and the screen display of buttons controlled by the Japanese operating system (such as the IME button that controls the Japanese character input function).

# **Evidence**

- The study confirmed the importance of defining evidence early in the development process. Specifying what will constitute evidence of proficiency allows test designers to design tasks to elicit that evidence and systems designers to design a delivery system to capture it. When working in a relatively new domain, such as ICT, this typically involves an iterative process of defining, testing, and refining evidence models.
- The importance of multiple sources of evidence was also illustrated in the study. In order to make inferences about ICT proficiency, performance data, background data, and observational data were collected and analysed. The observational data was particularly important in helping us to refine and focus our analysis of all the performance data captured by the delivery system. Therefore, another important step in the test development process is to observe students using prototyped tasks. Such a process provides information about measurement issues (does this task appear to be measuring what we thought it would?) as well as usability issues (can students use the tools and interface presented in the tasks?).

#### Conclusion

155. The ICT feasibility study should be considered a success in terms of the scope and range of lessons learned, as well as the results that provide a preliminary understanding of variables related to performance on ICT tasks. The study has demonstrated that a computer-based assessment provides tremendous potential to learn more about what students know and can do across a variety of ICT environments. Continued investigation would yield valuable information about students' ICT skills and also lay the groundwork for the future use of computers in the development and delivery of large-scale international assessments in any domain.



## **APPENDIX A: THE ICT EXPERT PANEL**

#### PANEL MEMBERS

**Irwin Kirsch** (Chair) Educational Testing Service Princeton, New Jersey

**Robert Kozma**, SRI International Menlo Park, California

**Chul-Hwan Lee**, Inchon National University of Education Inchon, Korea

**Jörgen Nissen**, Department of Information Technology Uppsala, Sweden

**Jean-Paul Reeff**, LIFE Research and Consult GmbH Esch-sur-Alzette, Luxembourg

**Catherine Régnier**, Ministère de l'Education Nationale Paris, France

**Arnold Spee**, Ministry of Education Zoetermeer, The Netherlands

**Wouter van Joolingen**, University of Amsterdam Amsterdam, The Netherlands

### PROJECT STAFF AND COUNSEL

**Tracy Gray**, American Institutes for Research Washington, D.C.

**Marylou Lennon**, Educational Testing Service Princeton, New Jersey

**Jay Moskowitz**, American Institutes for Research Washington, D.C.

**Maria Stephens**, American Institutes for Research Washington, D.C.

**Claudia Tamassia**, OECD, Education Directorate Paris, France



## APPENDIX B: SAMPLE DATA FILE - WEB SEARCH TASKS



In this file, user actions taken in the Web search task are encoded. An action identifier and any associated data are stored in a fixed width format. Those records with many trailing zeros represent students who performed only a few actions.



# **APPENDIX C: FOLLOW-UP QUESTIONNAIRE RESPONSES**

Question	Tally	%	% of students answering
How easy was it for you to get around the survey, that is, to move from screen to screen and section to section?			98%
Very easy	58	50%	
Easy	52	45%	
Difficult	4	3%	
Very difficult	2	2%	†
What did you think about the overall look of the survey (for	_	_ /0	
example, the graphics, colors, and fonts)?			99%
Very appealing	22	19%	
Appealing	88	75%	
Unappealing	7	6%	
Very unappealing	0	0%	1
What types of computers do you typically use?			100%
Laptop	18	15%	
Desktop	65	55%	1
Both	33	28%	
Neither	2	2%	
Which types of input devices do you typically use?	_		120%
Mouse	107	75%	
Touch pad	20	14%	
Track ball/joy stick	11	8%	
Other	4	3%	
Did the type of hardware used for the survey (for example, the kind of computer, input device, or keyboard) interfere with your ability to complete the tasks?			99%
No	97	83%	
Yes	20	17%	
If yes, how much do you think it interfered with your ability to complete the tasks?			17%
Very little	13	65%	
Somewhat	5	25%	
Very much	2	10%	
How would you rate the overall difficulty of this survey?  Overall, the tasks were			100%
Very easy	11	9%	
Easy	42	36%	1
Average	42	36%	1
Difficult	22	19%	1
Very difficult	1	1%	1



Appendix C: Follow-up Questionnaire Responses, continued

Question	Tally	%	% of students answering
How well did the survey allow you to demonstrate your computer skills and knowledge?			97%
Very well	13	11%	
Well	80	70%	1
Not very well	17	15%	1
Not at all	1	1%	1
Not applicable/I don't use computers	4	3%	1
How well do you think the survey covered the ways in which	•	070	
you typically use computers?			100%
Most of the things I typically do were included	17	14%	
Many of the things I typically do were included	46	39%	_
Only a few of the things I typically do were included	50	42%	
Nothing I typically do was included	5	4%	+
Not applicable/I don't use computers	0	0%	
How interesting was the email task?	<del>  0</del>	0 /0	99%
Very interesting	16	14%	33 /6
Interesting	57	49%	+
<u>_</u>	34	29%	+
Somewhat interesting  Not at all interesting	10	9%	_
	10	9%	000/
How interesting was the database task?	-	70/	99%
Very interesting	8	7%	_
Interesting	52	44%	
Somewhat interesting	39	33%	-
Not at all interesting	18	15%	200/
How interesting was the web search task?		400/	99%
Very interesting	14	12%	_
Interesting	50	43%	
Somewhat interesting	42	36%	
Not at all interesting	11	9%	
How interesting was the yeast simulation?			100%
Very interesting	29	25%	
Interesting	29	25%	
Somewhat interesting	37	31%	
Not at all interesting	23	19%	
How clear were the instructions for the email task?			99%
Very clear	58	50%	
Somewhat clear	49	42%	_
Confusing	10	9%	]
Very confusing	0	0%	
How clear were the instructions for the database task?			99%
Very clear	67	57%	
Somewhat clear	43	37%	
Confusing	5	4%	
Very confusing	2	2%	



Appendix C: Follow-up Questionnaire Responses, continued

Question	Tally	%	% of students answering
How clear were the instructions for the web search task?			99%
Very clear	40	34%	
Somewhat clear	53	45%	_
Confusing	22	19%	
Very confusing	2	2%	
How clear were the instructions for the yeast simulation?			99%
Very clear	23	20%	
Somewhat clear	41	35%	_
Confusing	40	34%	
Very confusing	13	11%	
Did you have enough time to carefully complete the email task?			97%
More than enough time	84	73%	
Enough time	25	22%	
Almost enough time	5	4%	
Not enough time	1	1%	
Did you have enough time to carefully complete the database task?			99%
More than enough time	84	72%	
Enough time	24	21%	
Almost enough time	7	6%	
Not enough time	2	2%	
Did you have enough time to carefully complete the web search task?			99%
More than enough time	79	68%	
Enough time	25	21%	
Almost enough time	8	7%	
Not enough time	5	4%	
Did you have enough time to carefully complete the yeast simulation?			96%
More than enough time	68	60%	
Enough time	31	27%	
Almost enough time	8	7%	
Not enough time	6	5%	
How easy did you find the email task?			99%
Very easy	52	44%	
Easy	57	49%	
Difficult	7	6%	1
Very difficult	1	1%	
How easy did you find the database task?			99%
Very easy	56	48%	
Easy	53	45%	
Difficult	7	6%	
Very difficult	1	1%	



Appendix C: Follow-up Questionnaire Responses, continued

Question	Tally	%	% of students answering
How easy did you find the web search task?			98%
Very easy	27	23%	
Easy	51	44%	
Difficult	28	24%	
Very difficult	10	8%	
How easy did you find the yeast simulation?			99%
Very easy	7	6%	
Easy	43	37%	
Difficult	51	44%	
Very difficult	16	14%	
Have you ever used e-mail programs before taking this survey?			99%
Yes	99	85%	
No	18	15%	
The e-mail program in this survey may have worked differently from others you have used. How much did any differences interfere with your ability to complete the task?			75%
Not at all	29	33%	
Very little	44	49%	
Somewhat	11	12%	
Very much	0	0%	
There were no differences	5	4%	
The e-mail program in this survey may have <u>looked</u> different from others you have used (for example icons/buttons may have been different, the inbox may have been presented differently). How much did any differences interfere with your ability to complete the task?			84%
Not at all	56	57%	
Very little	32	32%	
Somewhat	8	8%	
Very much	0	0%	
There were no differences	3	3%	
In this task you were asked to use e-mail to communicate with friends and arrange a social event. How well does this reflect the way you typically use e-mail?			83%
Very well	33	34%	
Somewhat	48	49%	
Very little	14	14%	
Not at all	3	3%	
Have you ever used database programs before taking this survey?			99%
Yes	55	47%	
No	62	53%	



Appendix C: Follow-up Questionnaire Responses, continued

Question	Tally	%	% of students answering
The database program in this survey may have worked differently from others you have used. How much did any differences interfere with your ability to complete the task?			47%
Not at all	29	53%	
Very little	19	35%	
Somewhat	3	5%	
Very much	0	0%	
There were no differences	4	7%	
The database program in this survey may have <u>looked</u> different from others you have used. How much did any differences interfere with your ability to complete the task?			47%
Not at all	30	55%	
Very little	22	40%	1
Somewhat	0	0%	
Very much	0	0%	
There were no differences	3	5%	
In this task, you were asked to use a database to sort information. How well does this reflect the way you typically use databases?			46%
Very well	21	39%	
Somewhat	29	54%	
Very little	4	7%	
Not at all	0	0%	
Have you ever used search engines (such as Google or Alta Vista) to search the Web before taking this survey?			98%
Yes	104	90%	
No	12	10%	
The search engine used in this survey may have worked differently from others you have used. How much did any differences interfere with your ability to complete the task?			89%
Not at all	58	55%	
Very little	39	37%	
Somewhat	2	2%	
Very much	0	0%	
There were no differences	6	6%	
The search engine used in this survey may have <u>looked</u> different from others you have used. How much did any differences interfere with your ability to complete the task?			89%
Not at all	57	54%	
Very little	38	36%	
Somewhat	4	4%	]
Very much	0	0%	]
There were no differences	6	6%	



Appendix C: Follow-up Questionnaire Responses, continued

Question	Tally	%	% of students answering
In this task, you were asked to use a search engine to look for information on the Web. How well does this reflect the way you typically use search engines?			87%
Very well	62	60%	
Somewhat	31	30%	
Very little	6	6%	
Not at all	4	4%	
Have you ever used simulations or other similar learning tools before taking this survey?			99%
Yes	42	36%	
No	75	64%	
The simulation tool in this survey may have <u>worked</u> differently from others you have used. How much did any differences interfere with your ability to complete the task?			36%
Not at all	16	38%	
Very little	14	33%	
Somewhat	11	26%	
Very much	1	2%	
There were no differences	0	0%	
The simulation tool in this survey may have <u>looked</u> different from others you have used. How much did any differences interfere with your ability to complete the task?			35%
Not at all	17	41%	
Very little	16	39%	
Somewhat	4	10%	
Very much	4	10%	
There were no differences	0	0%	
In this task, you were asked to use a simulation to collect and			
display information. How well does this reflect the way you	1		36%
typically use simulations or other learning tools?			
Very well	15	36%	
Somewhat	20	48%	
Very little	6	14%	
Not at all	1	2%	



ICT Feasibility Study

**Materials Annex** 

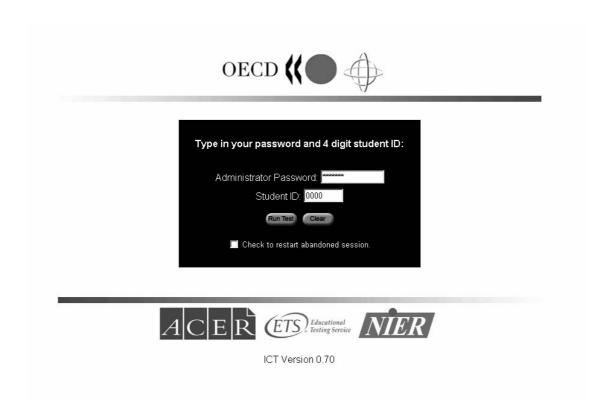


# Contents

Section 1 – ICT Screens	3
Section 2 – Follow-Up Questionnaire	69
Section 3 – Observation/Retrospective Interview materials	78
Section 4 – Test Adminstrator's Manual	126



# **SECTION 1 – ICT SCREENS**

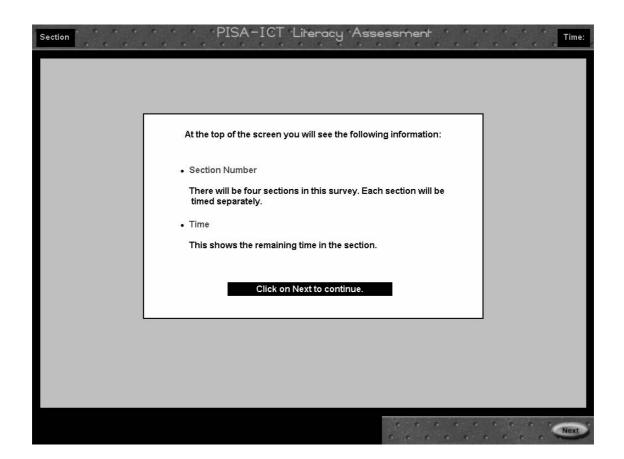








#### ICT Screens - Introduction

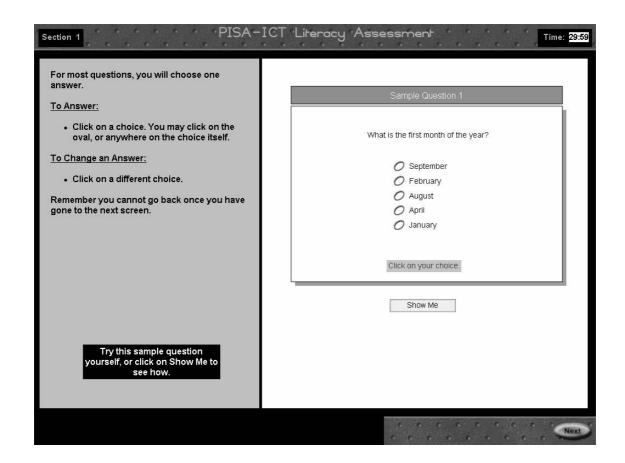






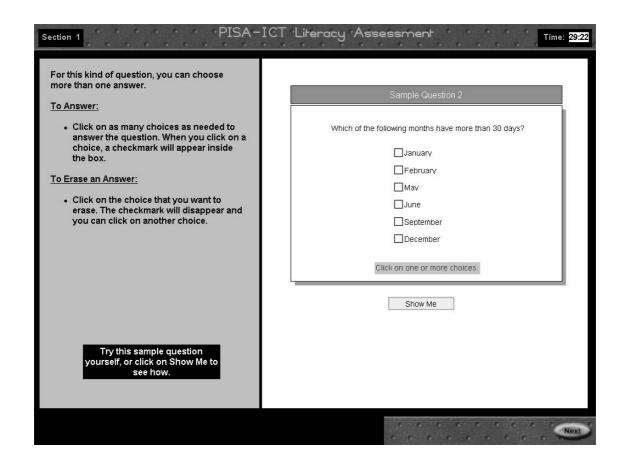


### ICT Screens – Response Modes for Background Questionnaire



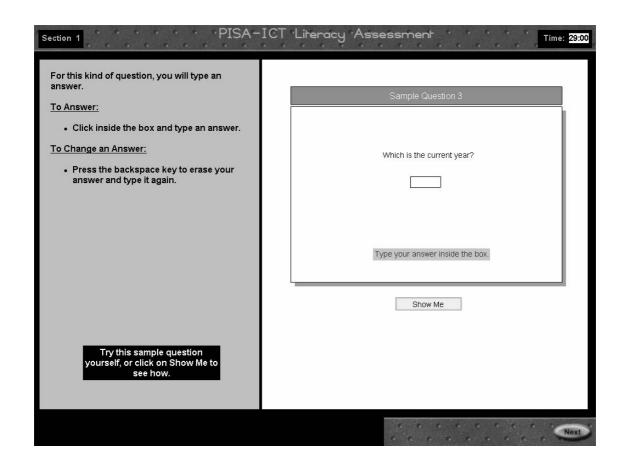


### ICT Screens – Response Modes for Background Questionnaire





### ICT Screens – Response Modes for Background Questionnaire







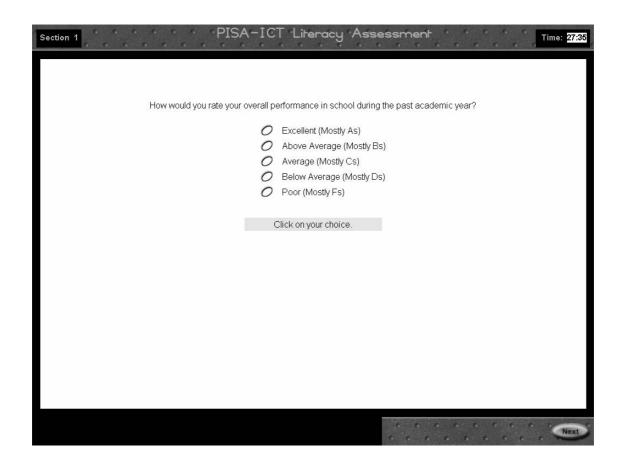




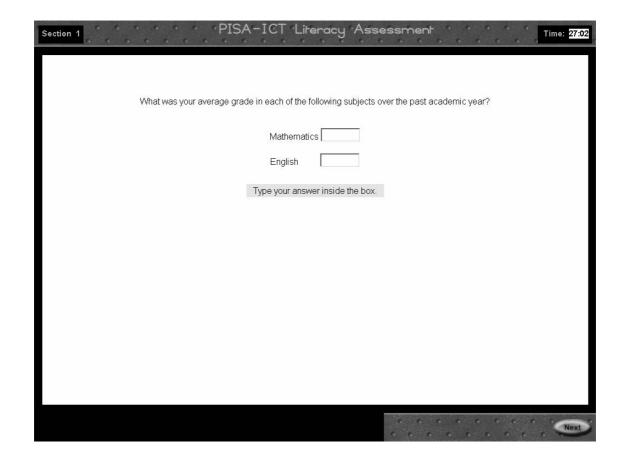




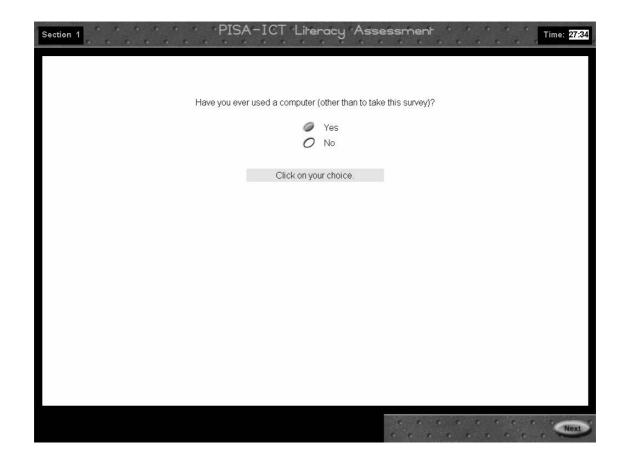








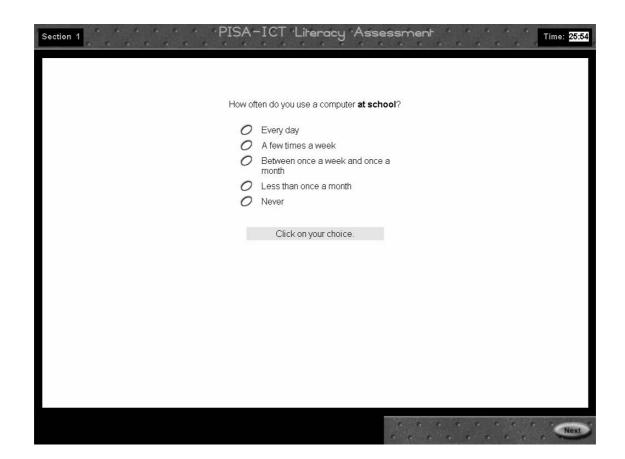




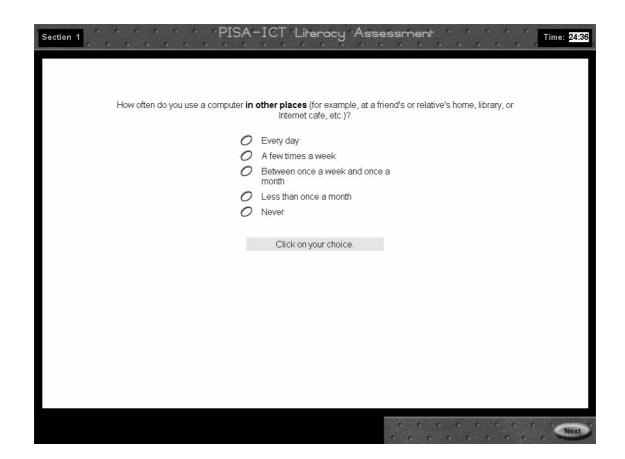




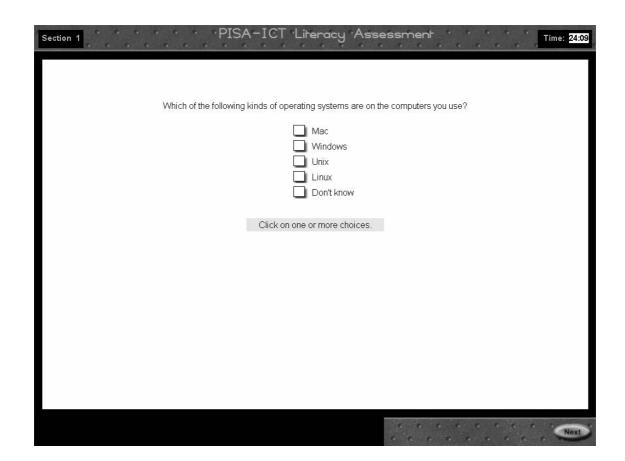




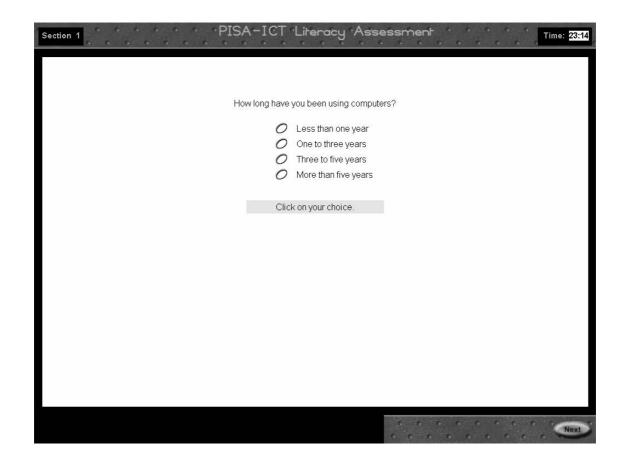




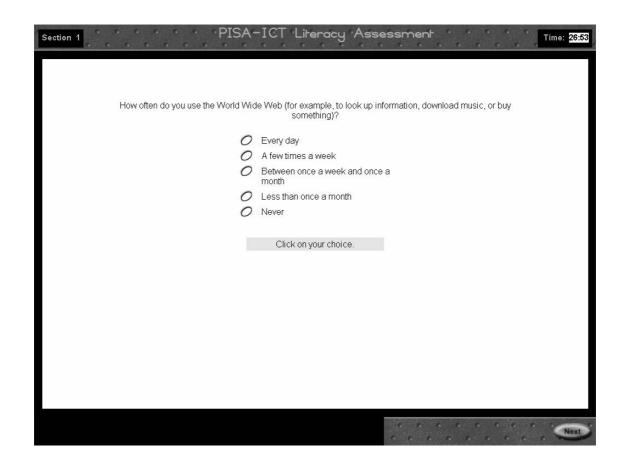




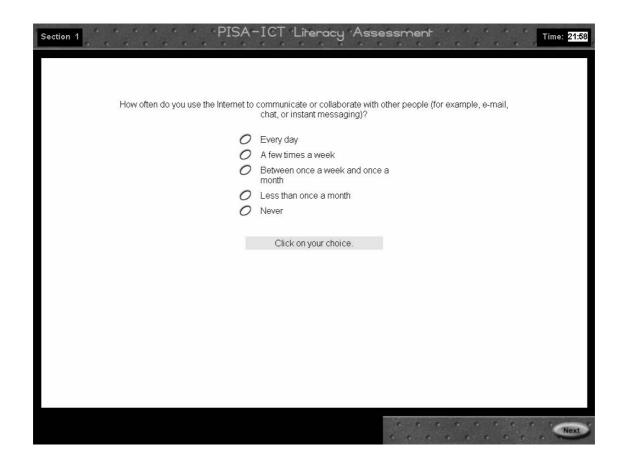




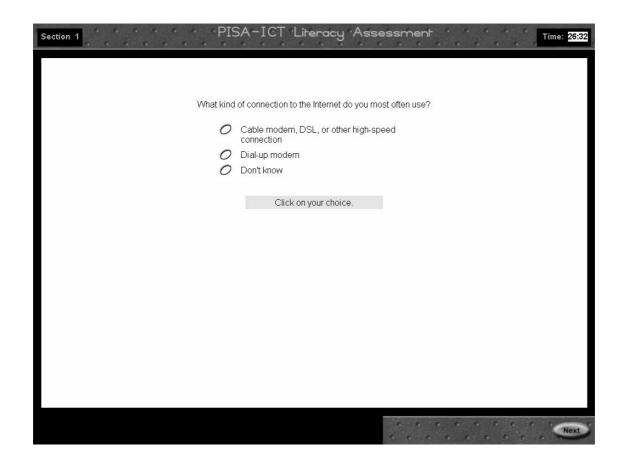




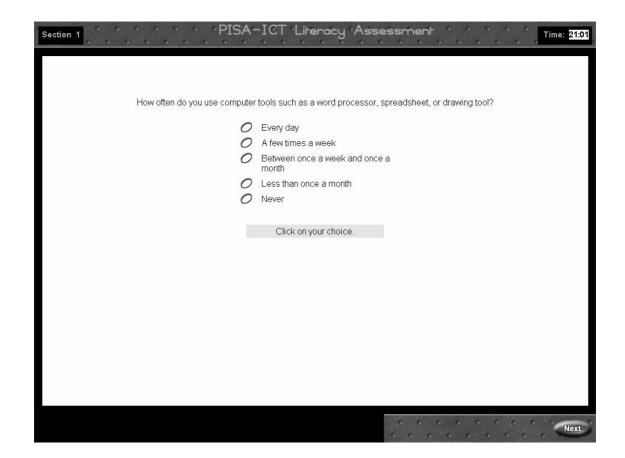




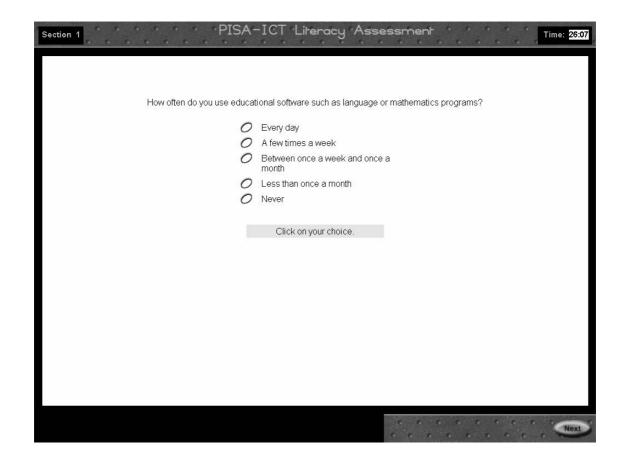




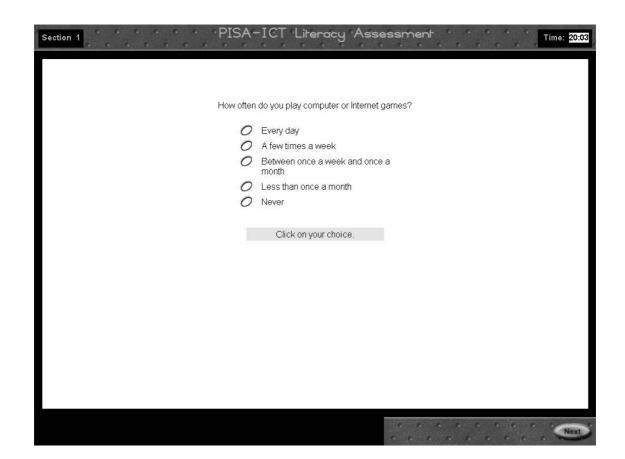




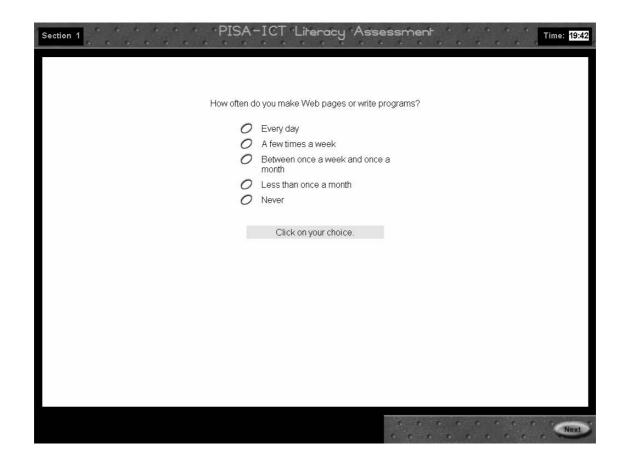




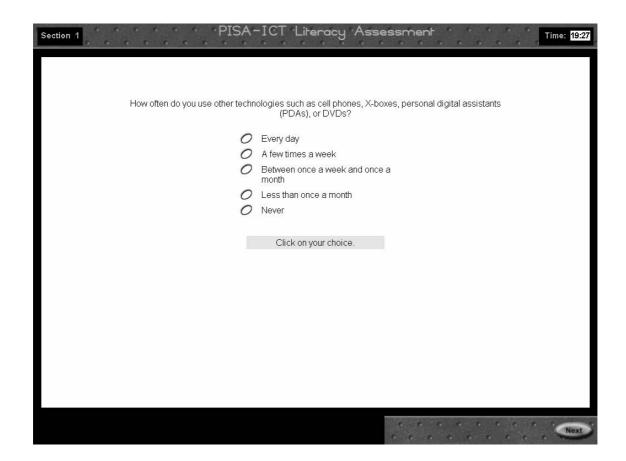




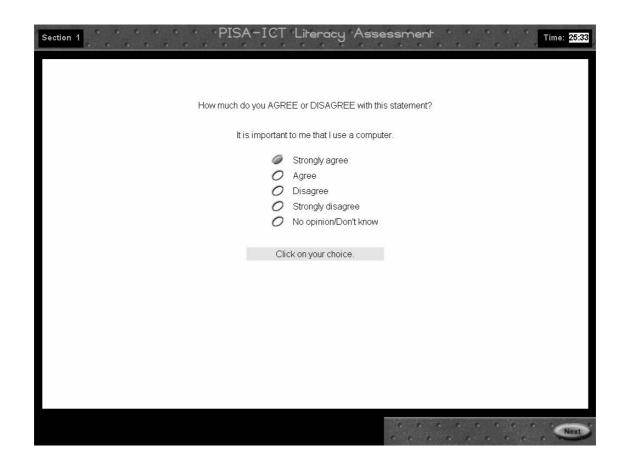




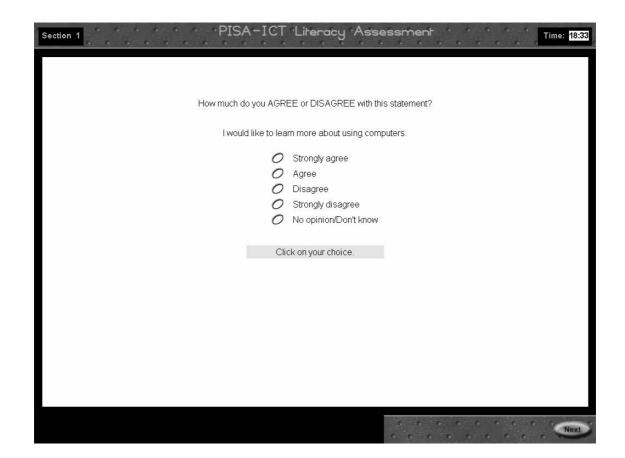




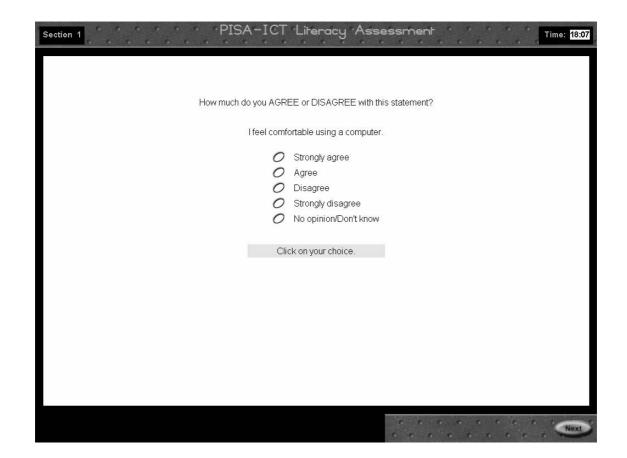




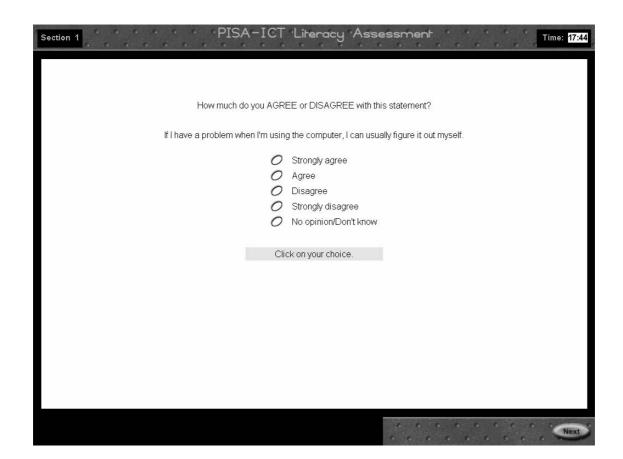




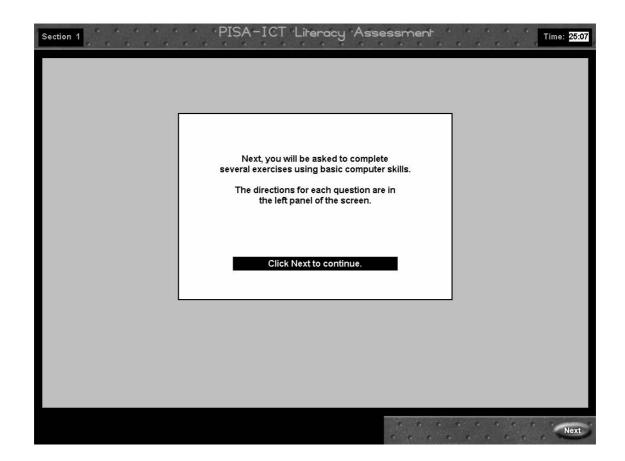




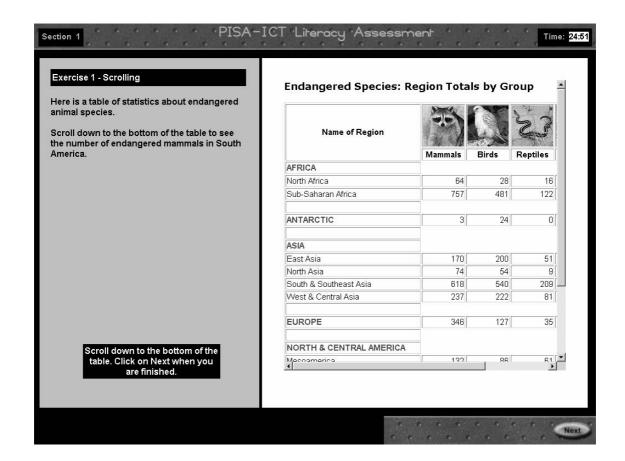




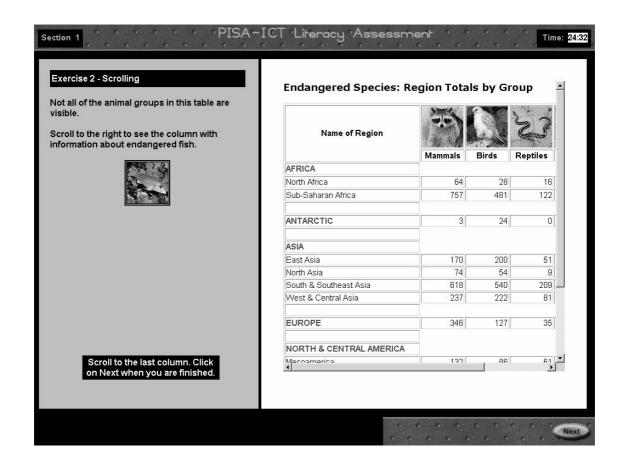


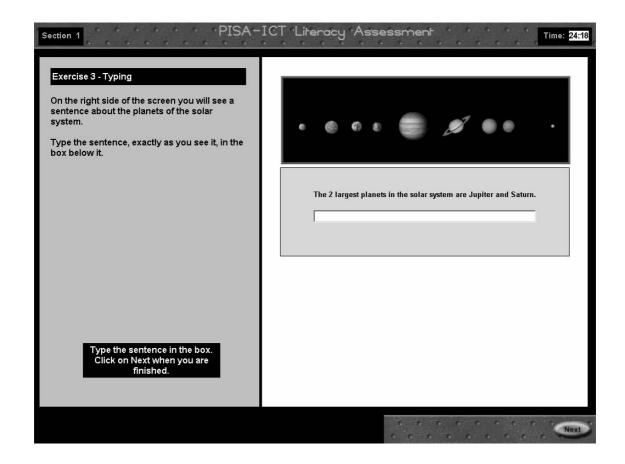




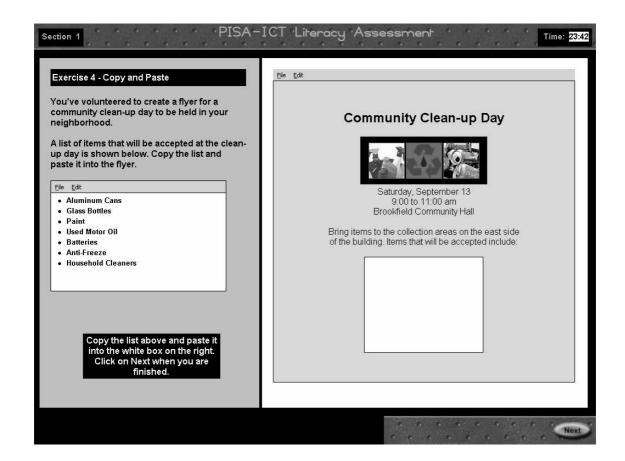




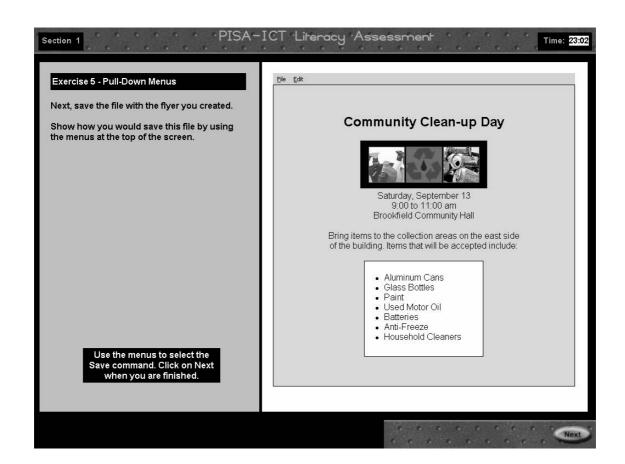




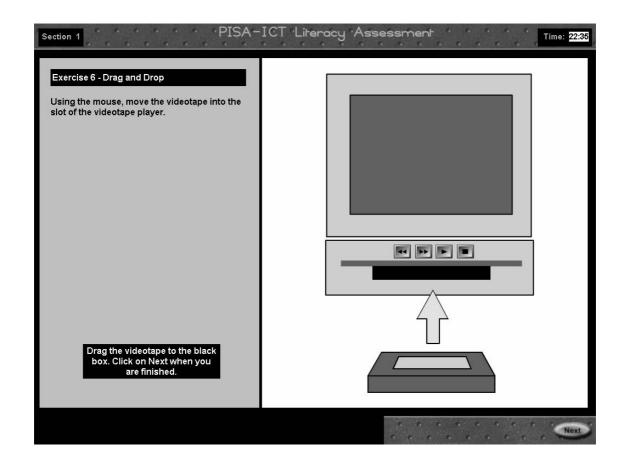






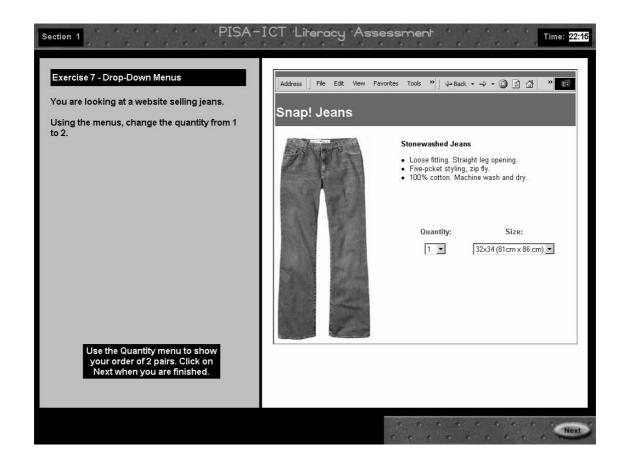






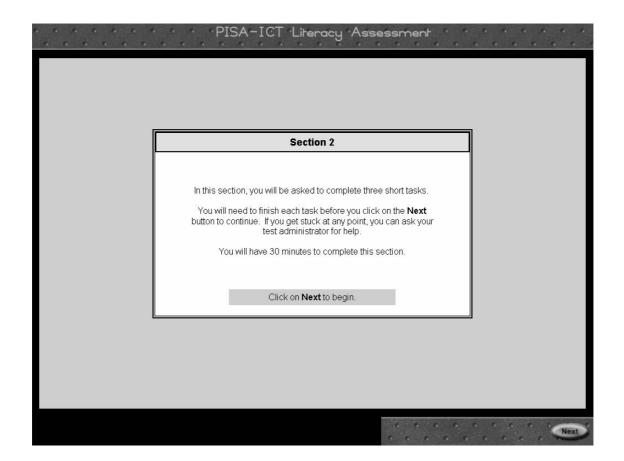


#### ICT Screens - Assessment of Basic Technical Skills



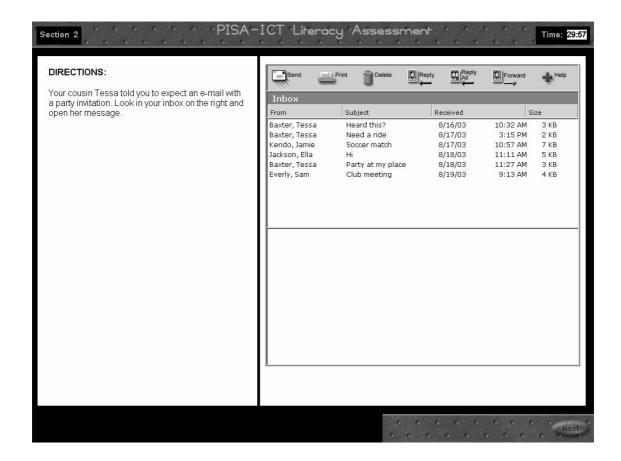


## ICT Screens - Section 2: Short Scenarios

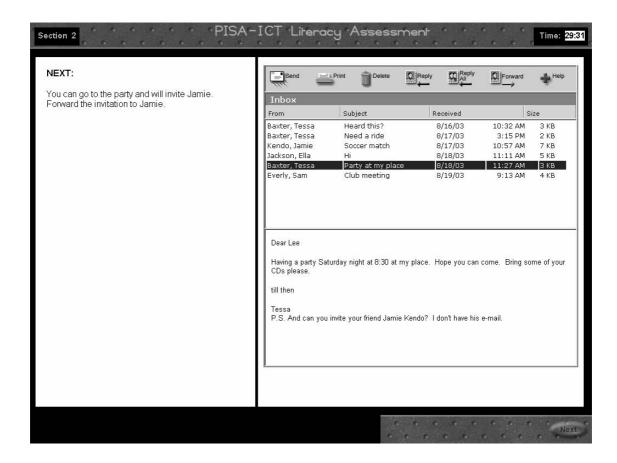




## ICT Screens - Section 2: Short Scenarios, E-mail Task



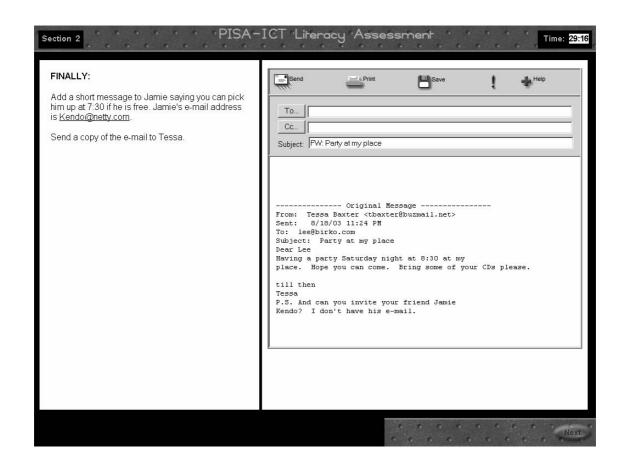




Correct e-mail selected – next set of directions displays.

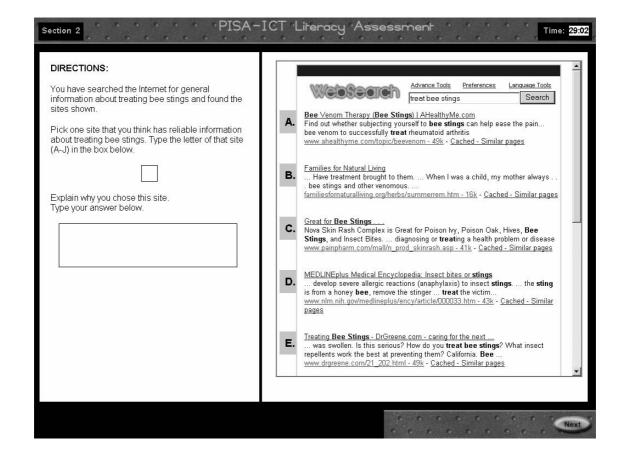


## ICT Screens - Section 2: Short Scenarios, E-mail Task



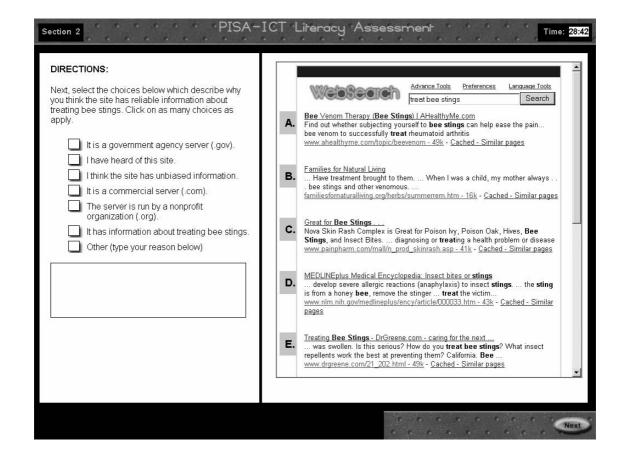


#### ICT Screens - Section 2: Short Scenarios, Web Abstract Task



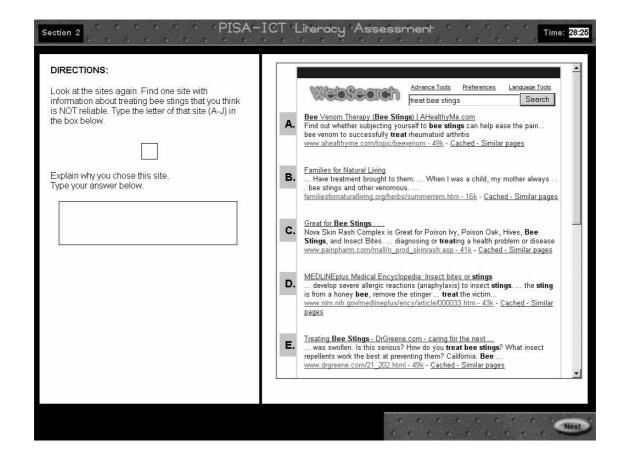


#### ICT Screens - Section 2: Short Scenarios, Web Abstract Task



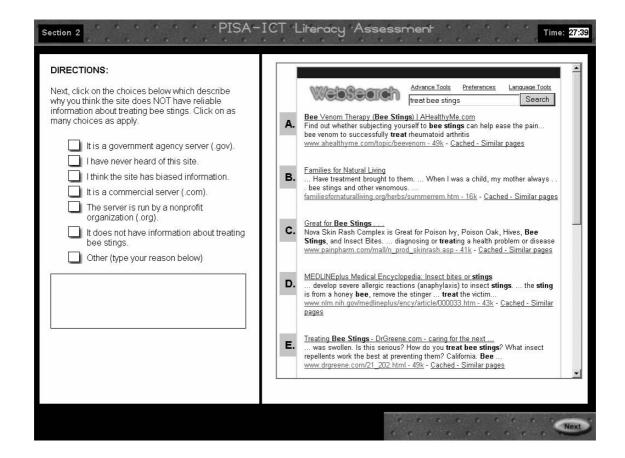


#### ICT Screens - Section 2: Short Scenarios, Web Abstract Task



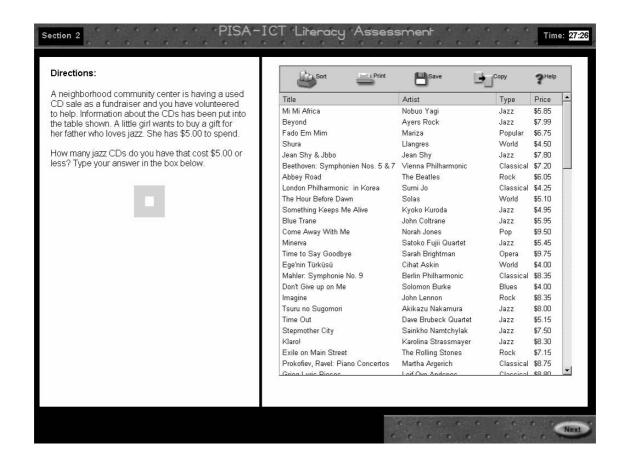


#### ICT Screens – Section 2: Short Scenarios, Web Abstract Task

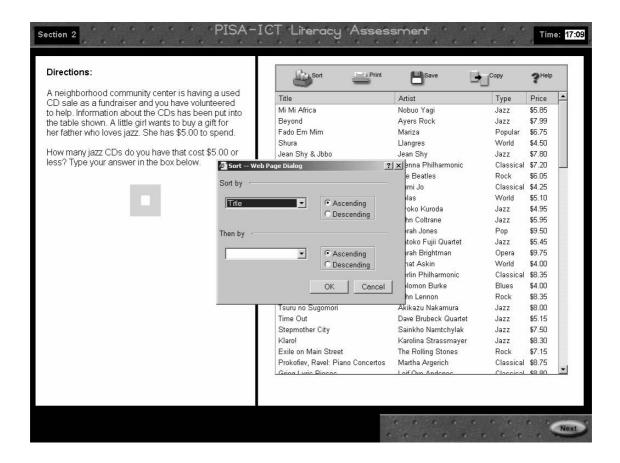




#### ICT Screens - Section 2: Short Scenarios, Database Task

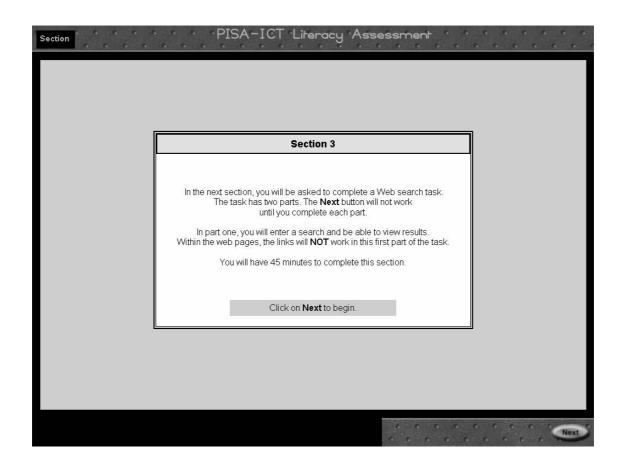




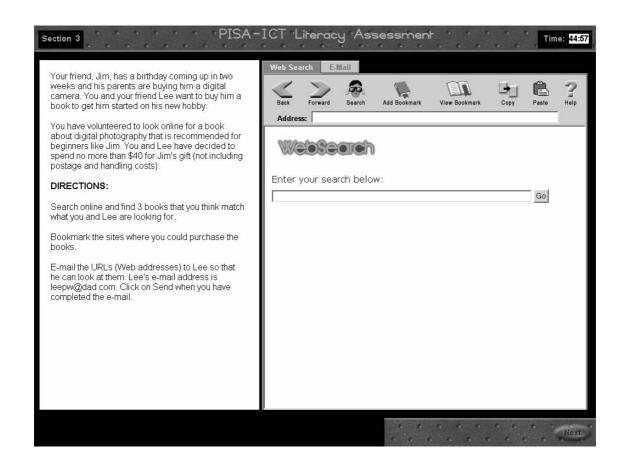


Dialog that displays when 'Sort' icon is clicked

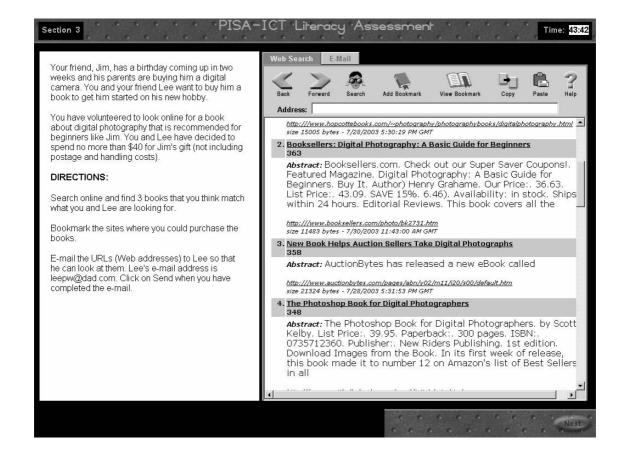






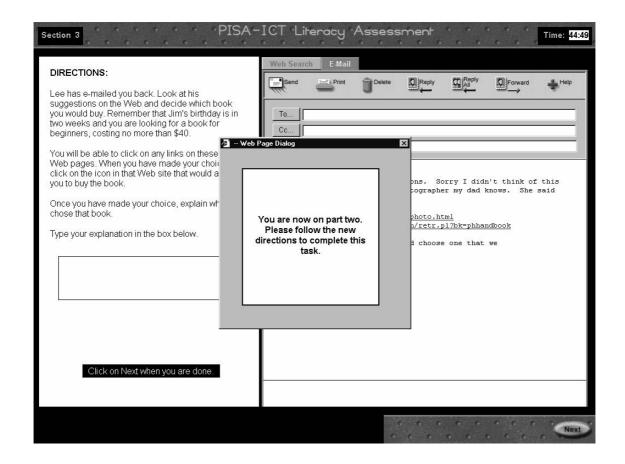




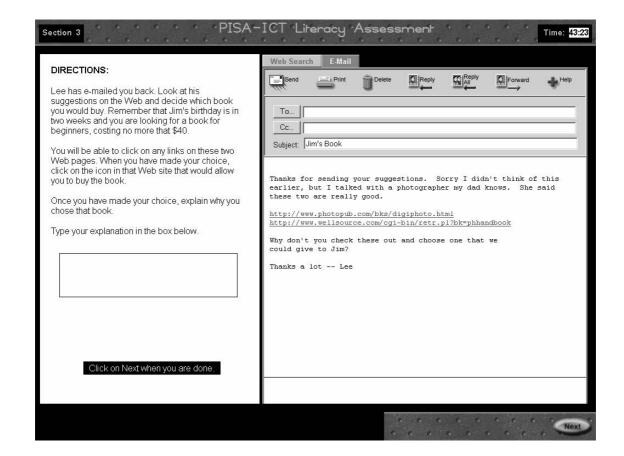


Sample page of hits.



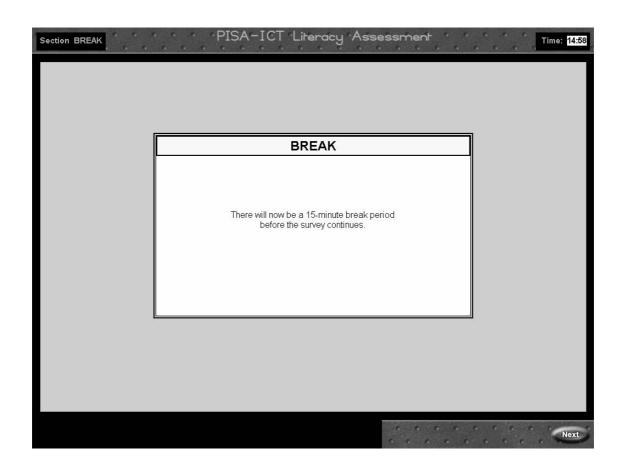




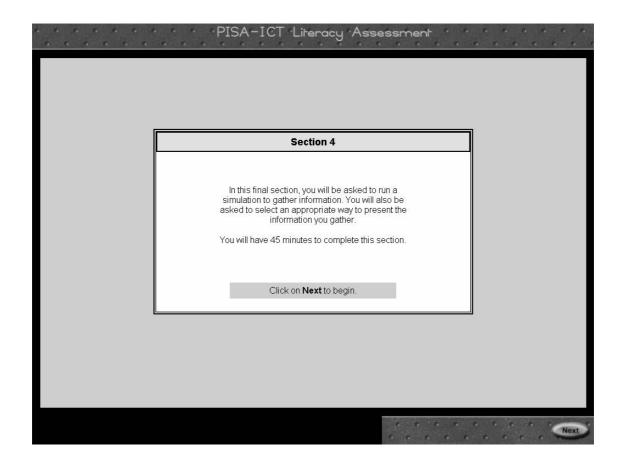




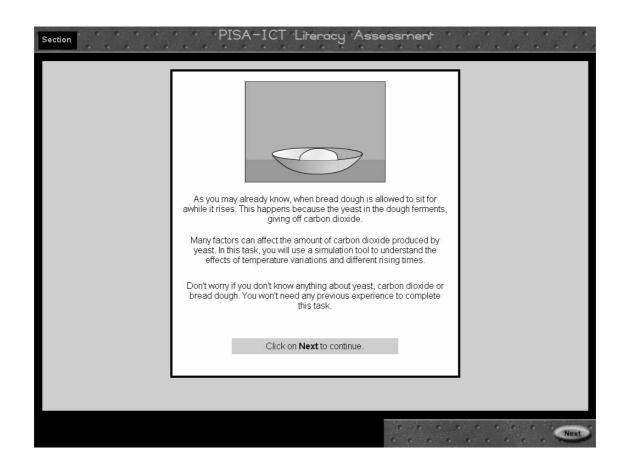
## **ICT Screens**



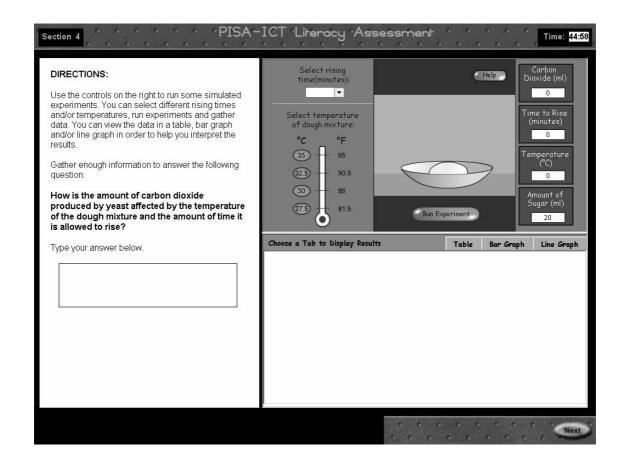




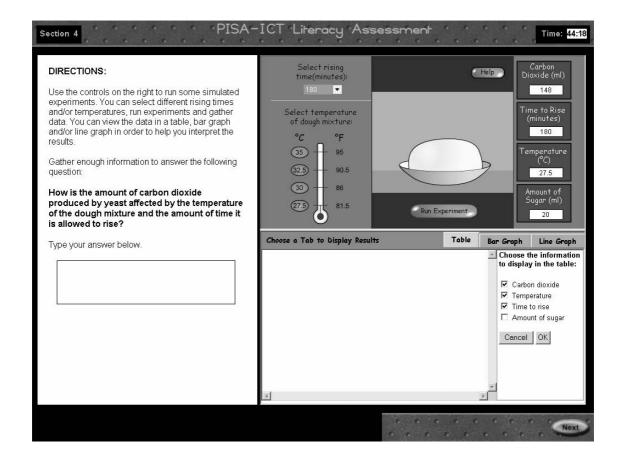






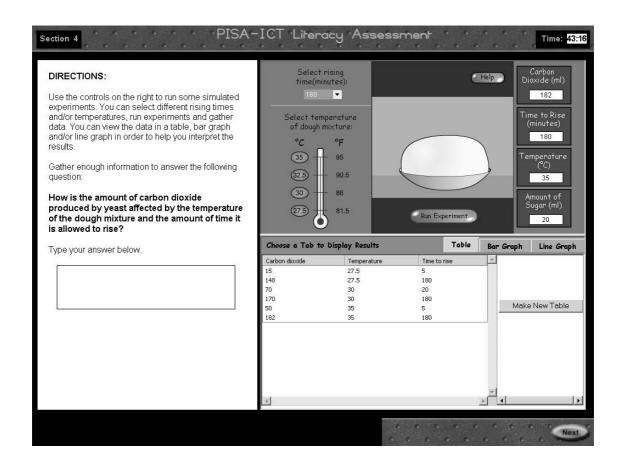






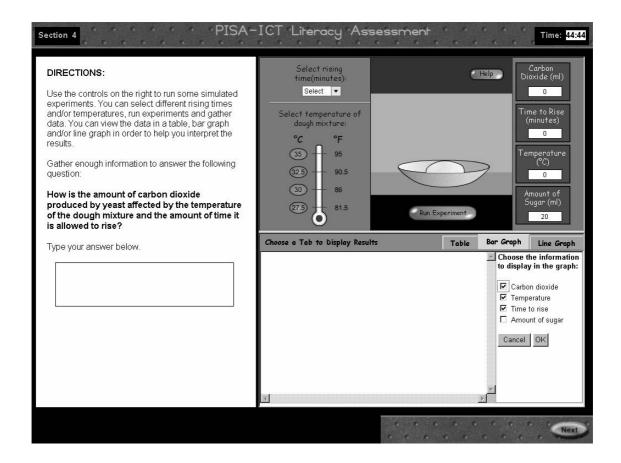
Selecting information for table display





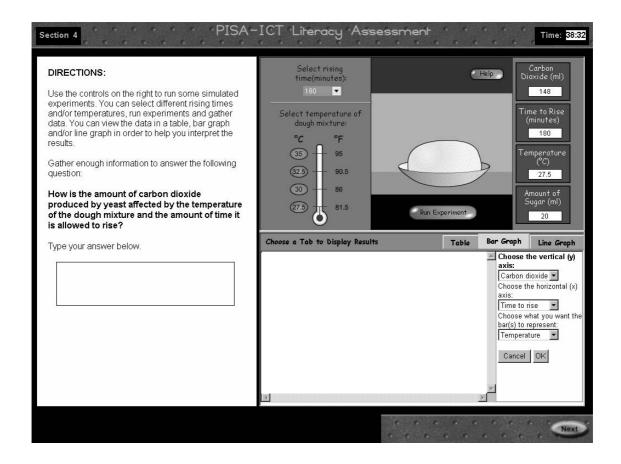
Sample table after six experiments have been run





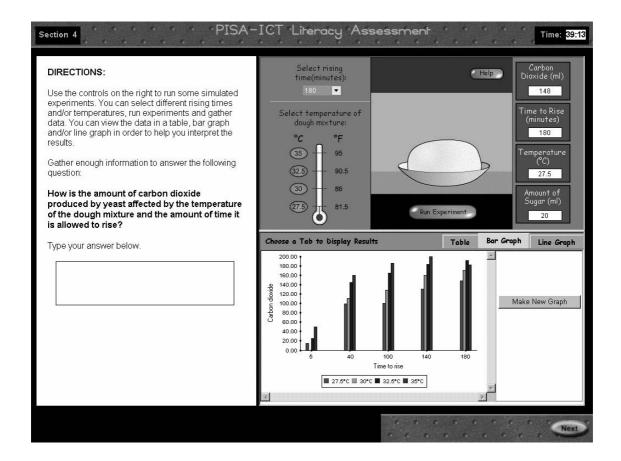
Selecting information for bar graph display





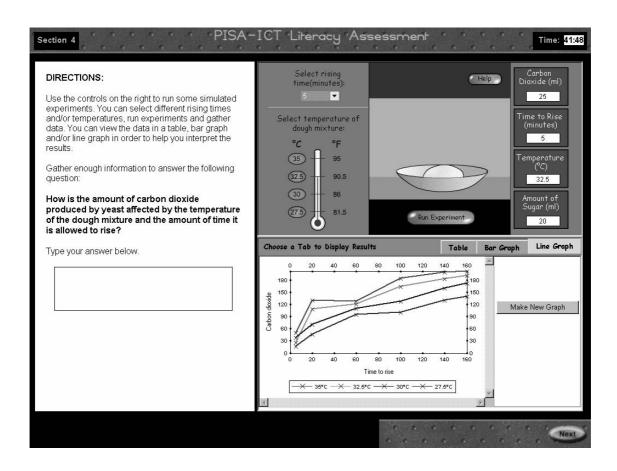
Selecting display for bar graph (pull-down menus allow users to choose any selected variable for axes and bars)





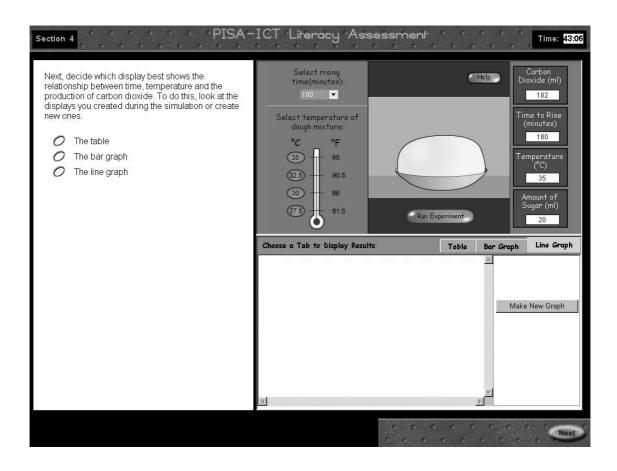
Sample bar graph display





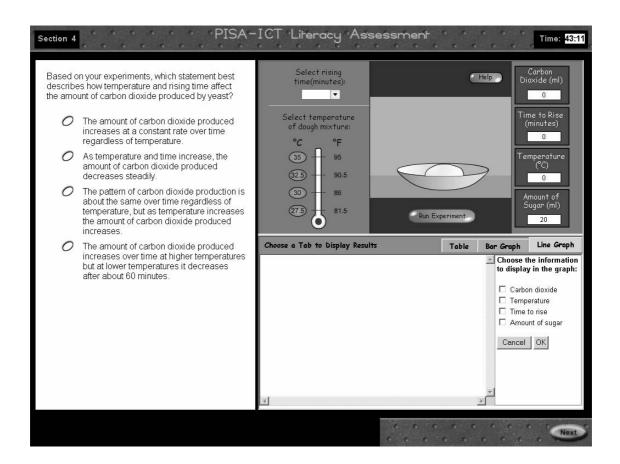
Sample line graph display





Simulation, Part 2. Users can choose to view displays created in Part 1 and/or create new displays to answer the question presented.

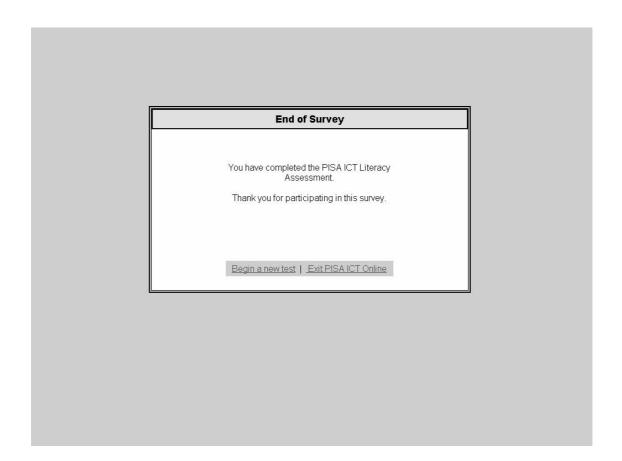




Simulation, Part 3



# **ICT Screens**





## **SECTION 2 - FOLLOW-UP QUESTIONNAIRE**

Student ID	Number	
Ctaaciit ib	1 10111001	

PISA - ICT Literacy Assessment

Follow-up Questionnaire

The ICT literacy assessment is an experimental version of one that will be developed for students your age. In order to improve this survey, your reactions and comments are very important. We appreciate your time and your help.



1.	How easy was it for you to get around the survey, that is, to move from screen to screen and section to section?  Uery easy Easy Difficult Very difficult
	Additional Comments:
2.	What did you think about the overall look of the survey (for example, the graphics, colors, and fonts)?  Very appealing  Appealing  Unappealing  Very unappealing
	Additional Comments:
3.	What types of computers do you typically use?  Laptop Desktop Both Neither
4.	Which types of input devices do you typically use?  ☐ Mouse ☐ Touch pad ☐ Track ball/joy stick
5.	Other  Did the type of hardware used for the survey (for example, the kind of computer, input device, or keyboard) interfere with your ability to complete the tasks?  □ No □ Yes
	If yes, how much do you think it interfered with your ability to complete the tasks?  ☐ Very little ☐ Somewhat ☐ Very much
	Additional Comments:



6.	How would you rate the overall difficulty of this survey? Overall, the tasks were  ☐ Very easy ☐ Easy ☐ Average ☐ Difficult ☐ Very difficult
	Additional Comments:
7.	How well did the survey allow you to demonstrate your computer skills and knowledge?  Very well  Well  Not very well  Not at all  Not applicable/I don't use computers
	Additional Comments:
8.	How well do you think the survey covered the ways in which you typically use computers?  Most of the things I typically do were included Many of the things I typically do were included Only a few of the things I typically do were included Nothing I typically do was included Not applicable/I don't use computers
	Additional Comments:
9.	Are there things you normally do on the computer, either in or outside of school, that you think should be included in a survey like this so students can demonstrate what they know?



Next, please answer a few questions about each of these four tasks. Circle one answer (1, 2, 3 or 4) for each task.









E-mail task

Database task

Web search task (photography book)

Yeast simulation

10. How interesting was each task?

	Very interesting	Interesting	Somewhat interesting	Not at all interesting
E-mail task	1	2	3	4
Database task	1	2	3	4
Web search task	1	2	3	4
Yeast simulation	1	2	3	4

#### 11. How clear were the instructions for each task?

	Very clear	Somewhat Clear	Confusing	Very confusing
E-mail task	1	2	3	4
Database task	1	2	3	4
Web search task	1	2	3	4
Yeast simulation	1	2	3	4

12. Did you have enough time to carefully complete each task?

<u> </u>				
	More than enough time	Enough time	Almost enough time	Not enough time
E mail table	diough time		2	4
E-mail task	1	2	3	4
Database task	1	2	3	4
Web search task	1	2	3	4
Yeast simulation	1	2	3	4

13. How easy did you find each task?

10. How casy ara ye	ou illia cacii tao	orx :		
	Very easy	Easy	Difficult	Very difficult
E-mail task	1	2	3	4
Database task	1	2	3	4
Web search task	1	2	3	4
Yeast simulation	1	2	3	4



The following questions are about the e-mail task.



14.	Have you ever used e-mail programs before taking this survey? □ Yes □ No
	If 'No', please skip to Question #19
15.	The e-mail program in this survey may have worked differently from others you have used. How much did any differences interfere with your ability to complete the task?  Not at all Very little Somewhat Very much There were no differences
	Additional Comments
16.	The e-mail program in this survey may have looked different from others you have used (for example icons/buttons may have been different, the inbox may have been presented differently). How much did any differences interfere with your ability to complete the task?  Not at all Very little Somewhat Very much There were no differences
	Additional Comments
17.	In this task you were asked to use e-mail to communicate with friends and arrange a social event. How well does this reflect the way you typically use e-mail? □ Very well □ Somewhat □ Very little □ Not at all
	Additional Comments
18	In what other ways do you use e-mail?



The following questions are about the database task.



ave you ever used database programs before taking this survey?  Yes No If 'No', please skip to Question #24  The database program in this survey may have worked differently from others you have used. How much did any differences interfere with your ability to complete the sk?
ne database program in this survey may have <u>worked</u> differently from others you live used. How much did any differences interfere with your ability to complete the
we used. How much did any differences interfere with your ability to complete the
<ul> <li>□ Not at all</li> <li>□ Very little</li> <li>□ Somewhat</li> <li>□ Very much</li> <li>□ There were no differences</li> </ul>
ditional Comments
ne database program in this survey may have <u>looked</u> different from others you have led. How much did any differences interfere with your ability to complete the task?  Not at all  Very little  Somewhat  Very much  There were no differences
dditional Comments
this task, you were asked to use a database to sort information. How well does is reflect the way you typically use databases?  Very well Somewhat Very little Not at all
dditional Comments
i



The following questions are about the Web search task.



24.	Have you ever used search engines (such as Google or Alta Vista) to search the Web before taking this survey?  ☐ Yes ☐ No  If 'No', please skip to Question #29
25.	The search engine in this survey may have worked differently from others you have used. How much did any differences interfere with your ability to complete the task?  Not at all Very little Somewhat Very much There were no differences
	Additional Comments:
26.	The search engine in this survey may have looked different from others you have used (for example, the search page and/or results page may have looked different.)  How much did any differences interfere with your ability to complete the task?  Not at all  Very little  Somewhat  Very much  There were no differences  Additional Comments:
27.	In this task were you asked to use a search engine to look for information on the Web. How well did this task reflect the way you typically use search engines?  Very well Somewhat Very little Not at all
	Additional Comments.
28.	In what other ways do you typically use search engines?



# Follow-Up Questionnaire

The following questions are about the yeast simulation.



29.	Have you ever used simulations or other similar learning tools before taking this survey?  ☐ Yes ☐ No
	If 'No', please skip to Question # 34
30.	The simulation tool in this survey may have worked differently from others you have used. How much did any differences interfere with your ability to complete the task?  Not at all Very little Somewhat Very much There were no differences
	Additional Comments:
31.	The simulation tool in this survey may have <u>looked</u> different from others you have used. How much did any differences interfere with your ability to complete the task?    Not at all   Very little   Somewhat   Very much   There were no differences
	Additional Comments:
32.	In this task you were asked to use a simulation to collect and display information.  How well did this task reflect the way you typically use simulations or other learning tools?  Uery well Somewhat Very little Not at all  Additional Comments:
33.	In what other ways do you typically use simulations or other learning tools?



# Follow-Up Questionnaire

34.	Is there anything else you want to tell us that might help us improve this survey?

Thank you very much for your answers and comments.

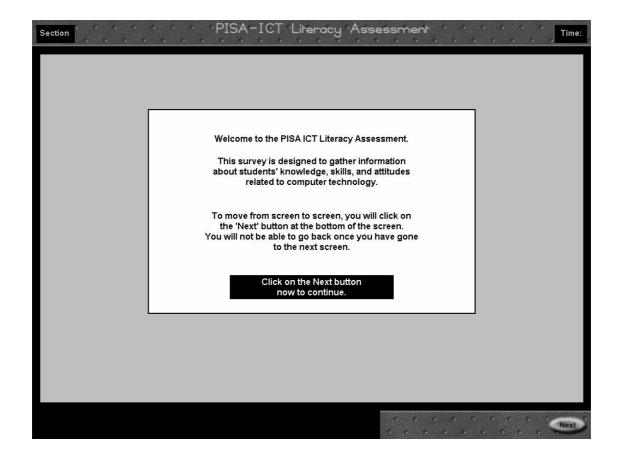


#### SECTION 3 - OBSERVATION/RETROSPECTIVE INTERVIEW MATERIALS

Some core issues and questions to guide observations

- What process did the student use to complete a task?
- · Was there anything unusual or unexpected about that process?
- Was the student able to complete tasks easily and with some automaticity?
- Did the student seem to be familiar with ICT terminology used in the survey (e.g., URL, in-box, bookmark)?
- Were there any problems with the task design that interfered with the student's performance?
- Were there any problems with the interface design that interfered with the student's performance?
- Was the student comfortable with unfamiliar interface elements (could they change their approach if it did not yield the intended results)?





#### **Observation**

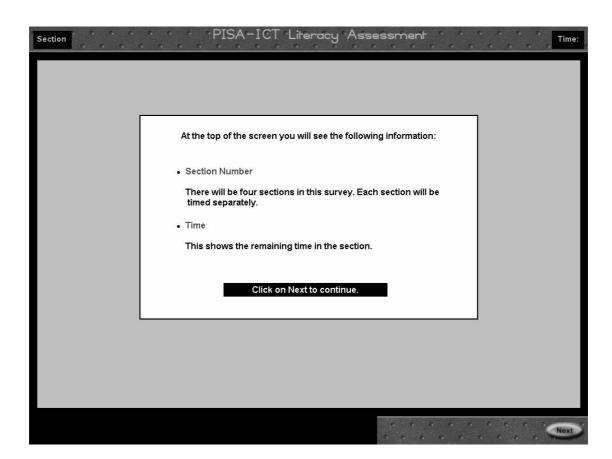
Does the student understand to click on 'Next' to continue?

Yes No

#### Follow up

Were these directions clear or was anything confusing?

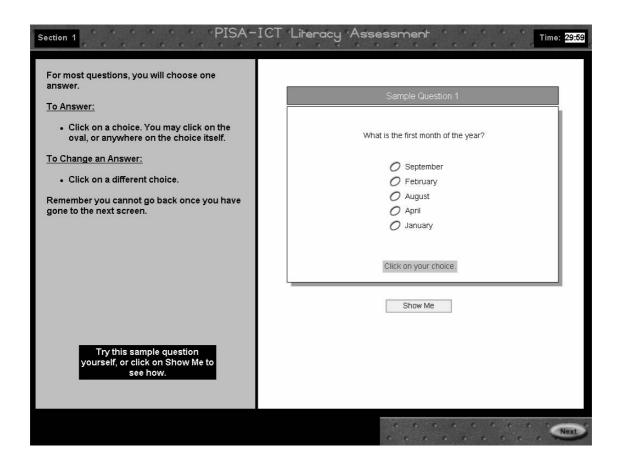












$\sim$	hser	1:	
( )	nsen	van	m

Reads directions in left pane

If student does more than one o	f the following, indicate order if possible.
☐ Clicks on a choice	
Tries changing an answer	<del></del>
☐ Clicks on 'Show Me'	

Is the student able to easily complete the task?

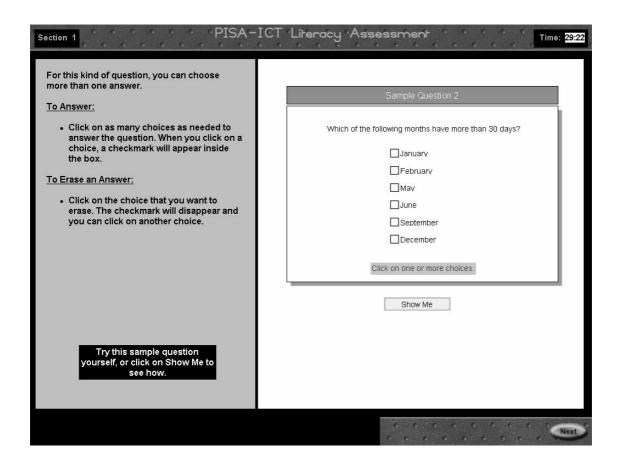
Yes

No

#### Follow up

If the student clicked on 'Show Me', probe for reason - to help understand how to answer, to see what happened? (No need to ask this all three times if the student clicks 'Show Me' on each sample.)



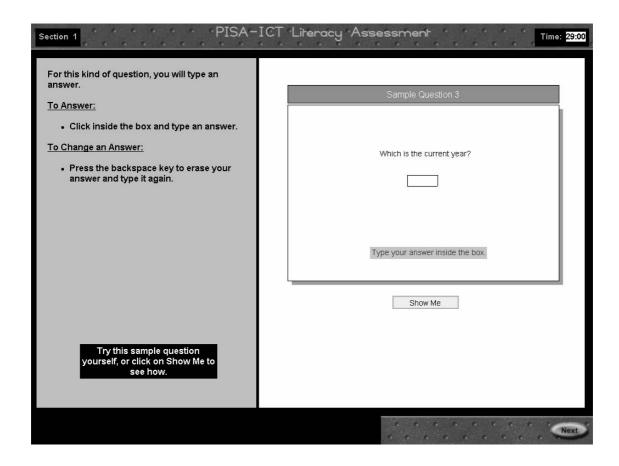


<u>Observation</u>				
If student does more than one of the fol ☐ Clicks on one choice ☐ Clicks on more than one choice ☐ Tries changing an answer ☐ Clicks on 'Show Me'	lowing, indicate	order if po - - -	ossible.	
Is the student able to easily complete th	ie task?	Yes	No	

#### Follow up

If the student clicked on 'Show Me', probe for reason - to help understand how to answer, to see what happened? (No need to ask this all three times if the student clicks 'Show Me' on each sample.)





#### **Observation**

If student does more than one of the following, inc  ☐ Types a response ☐ Clicks on 'Show Me'	licate order if possible.
Is the student able to easily complete the task?	Yes No

#### Follow up

If the student clicked on 'Show Me', probe for reason - to help understand how to answer, to see what happened? (No need to ask this all three times if the student clicks 'Show Me' on each sample.)





Observation - Background Questionnaire

Mark any questions that cause difficulty or confusion for the student.

## Follow up

Probe any questions that presented a problem for the student - what did you find confusing? How might the question be improved?

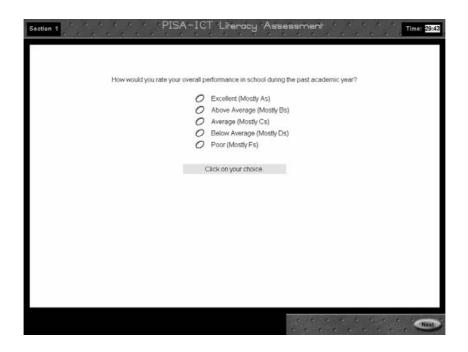


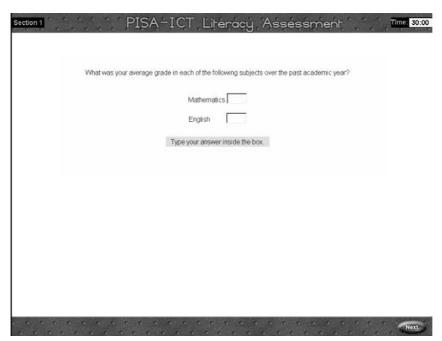


[Note - The text entry box will accept a 4-digit number or a 2-digit number, including a 2-digit Showa year in Japanese.]









[Note - students may enter a '+' or '-' in addition to their grade. Numbers are accepted as well as letters.]

#### Follow up

Ask the student to explain any unfamiliar or unusual grade designations.





If students answer 'No' to this question, they are branched to the following question:

- How often do you use other technologies?

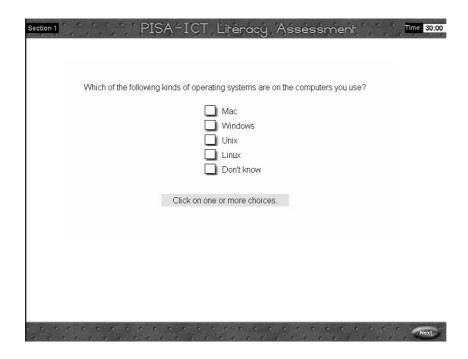






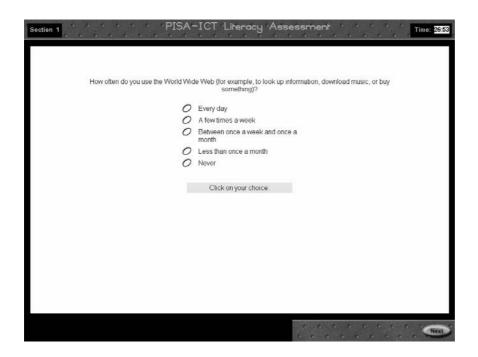








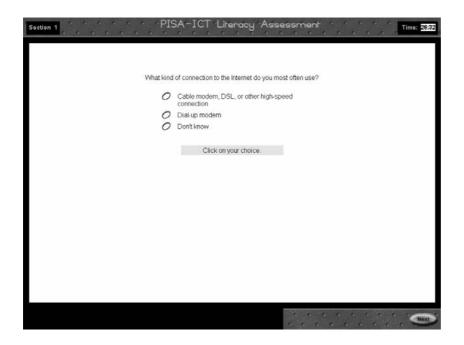






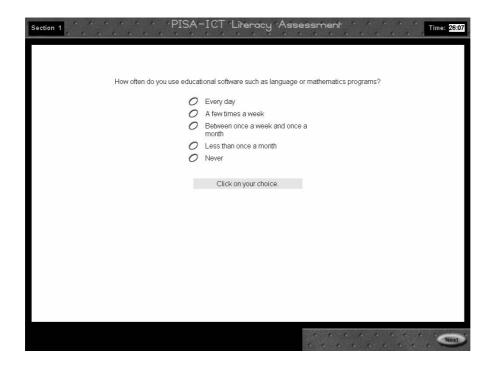
[Note - if the student answers 'Never' the software will skip the next question.]





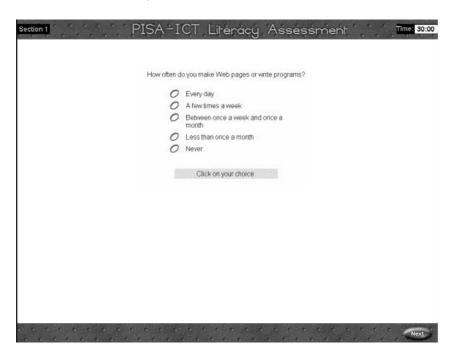








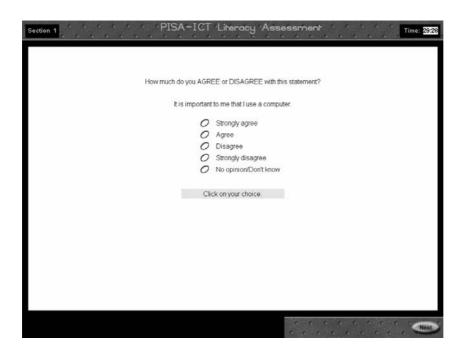




[Note - may probe to find out if the student uses HTML or programs such as Dreamweaver to create Web pages; also if the student uses other programming languages to write programs or solve problems]













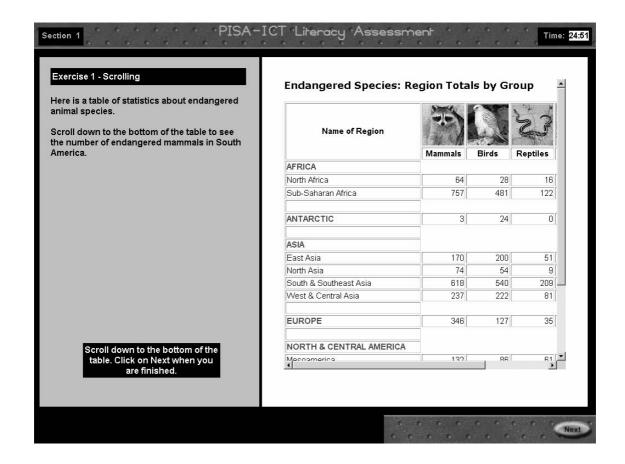
[Note - we intended to have the software skip the questions above and below for students answering that they have not used computers before (page 10). This fix did not get implemented. The test administrator can tell those students to click 'No opinion' on both these questions if they ask.]











## **Observation**

Does the student have any difficulty with vertical scrolling? Yes If the student scrolls, is 'South America' still visible on screen when the student clicks on 'Next'?

#### Follow up

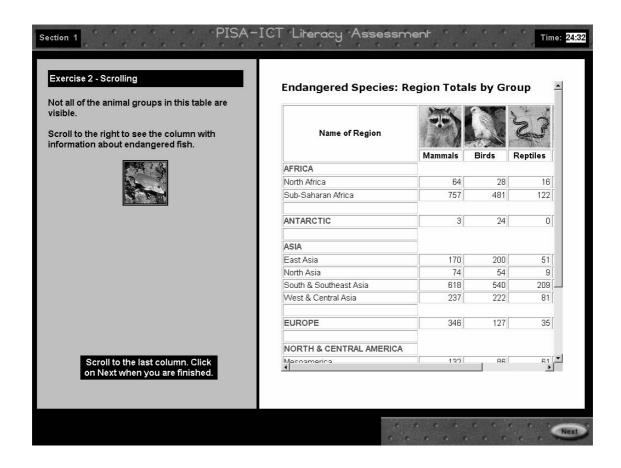
(If student had difficulty with the task) Is the term 'scroll' one with which you are familiar?



No

No

Yes



## **Observation**

Does the student have any difficulty with horizontal scrolling?

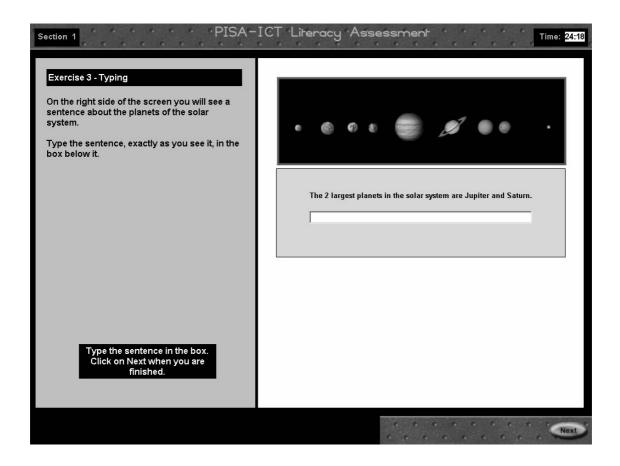
Yes
If the student scrolls, is the fish column still visible on screen
when the student clicks on 'Next'?

Yes



No

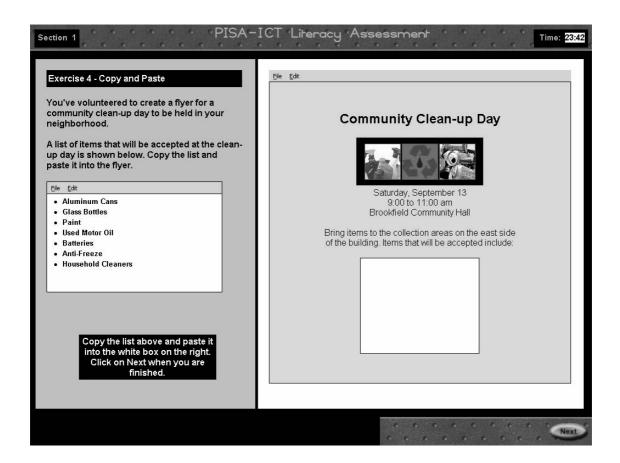
No



# Observation

Does the student know to click in the typing box? Yes No Does the student type easily? Yes No

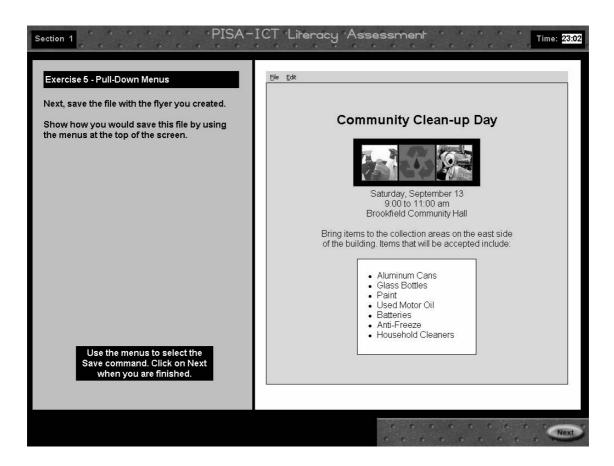




Observation
General process followed
Copies and pastes entire block of text
☐ Copies and pastes line by line
□ Retypes list
☐ Tries to use 'Cut'
How does the student copy/paste?
☐ Highlights entire block
☐ Highlights line by line
☐ Uses pull-down menu bar
☐ Uses short-cut keys (Ctl+C and Ctl+V)

Is the student able to easily complete the task? Yes No





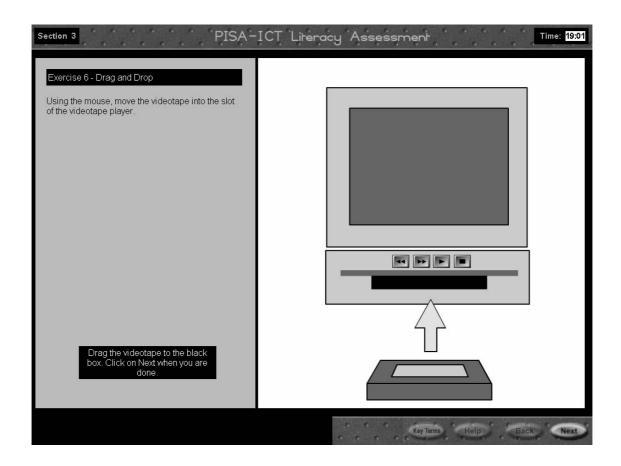
#### Observation

Is the student able to use the pull-down menu?

Yes

No



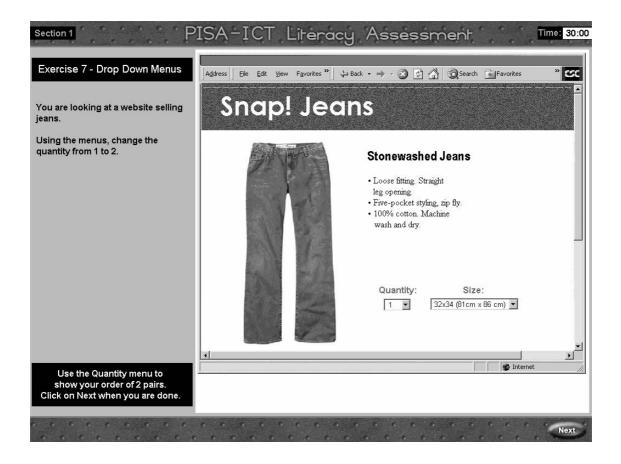


#### Observation

Is the student able to smoothly drag the videotape to the correct location?



No

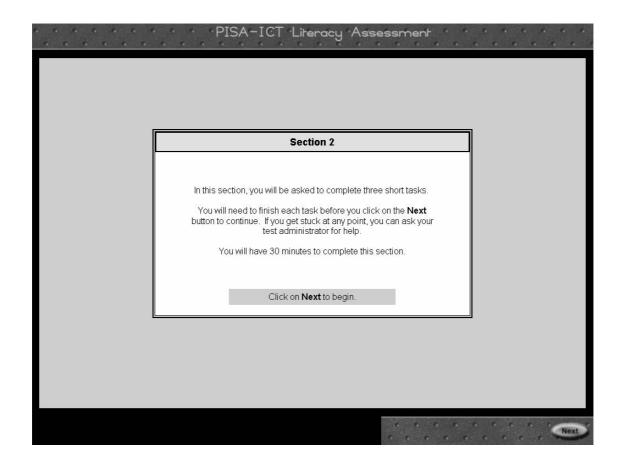


#### Observation

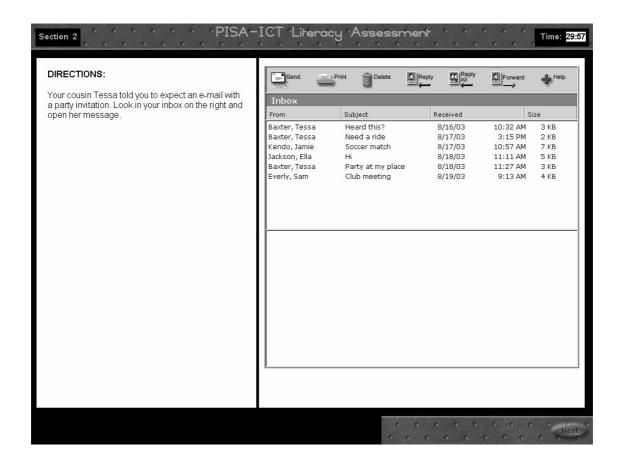
Can the student easily use the drop-down menu?

Yes No









#### Observation

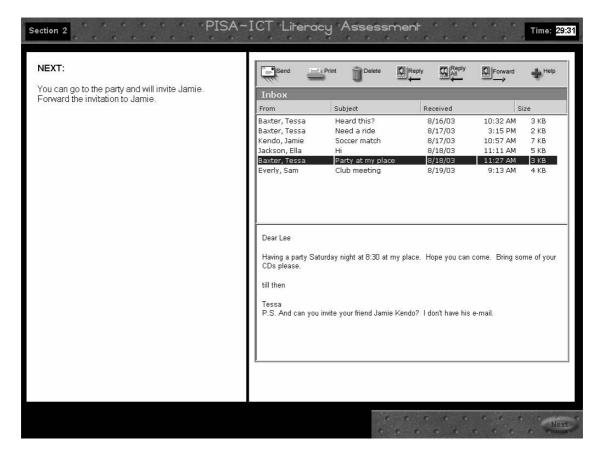
- ☐ Clicks on correct e-mail message (Baxter, Tessa Party at my place)
- ☐ Clicks on incorrect e-mail messages
- ☐ Clicks on e-mail icons circle all that apply: Send, Print, Delete, Reply, Reply All, Forward
- ☐ Clicks on 'Help'

#### Follow up

Probe if student clicked on incorrect e-mail message(s), probe (unsure of which email was correct, just curious to see what they said?)

Probe if student clicked on e-mail icon(s) (thought icons would help complete task; just curious?)





#### Observation

General process followed

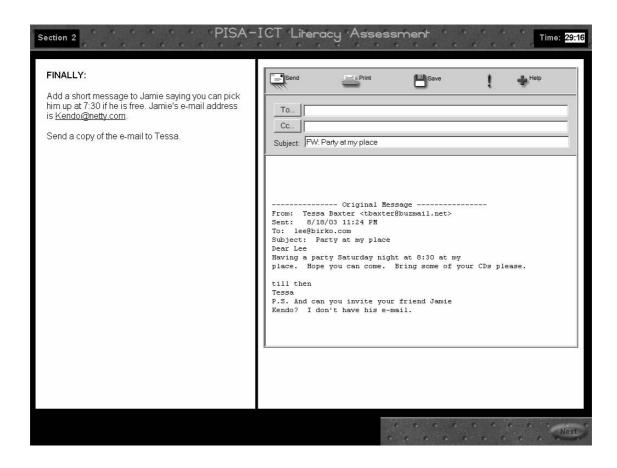
- ☐ Clicks on 'Forward' icon
- ☐ Clicks on incorrect icons circle all that apply: Send, Print, Delete, Reply, Reply All
- ☐ Clicks on e-mail message(s) in in-box
- ☐ Clicks on 'Help'

#### Follow up

Did you notice that the directions on the left had changed?

Probe if student clicked on any incorrect icons (thought icons would help complete task; just curious?)





#### Observation

☐ Types e-mail message - includes appropriate information (pick up at 7:3	0)? Yes	No
☐ Includes Jamie's e-mail address - circle: cuts/pastes types		
Includes Tassa's a mail address in CC circle: outs/pastes types		

☐ Includes Tessa's e-mail address in CC - circle: cuts/pastes types

☐ Clicks 'Send' button

Clicks on other icons - circle all that apply: 'Print' 'Save' '!'

☐ Clicks on 'Help'

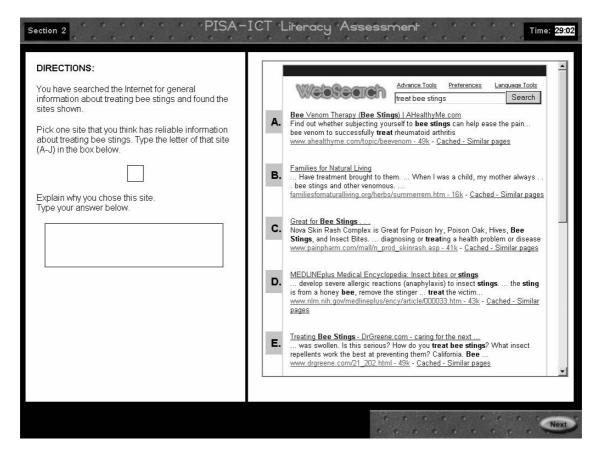
Is the student able to easily complete the e-mail task? Yes No

#### Follow up

Tell me how you completed this part of the task.

Probe if they did not include Tessa's address (not notice the directions? Unable to locate address?)





[Note - links on this task do not function.]

Correct responses for this first part include D, G, H, J - .gov and .edu sources with information about treating bee stings.

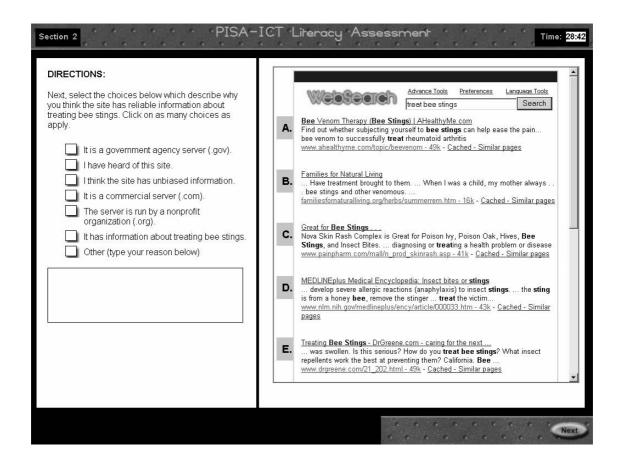
Observat	<u>ion</u>
	Scrolls to see all choices
□ '	Types explanation
	Tries to click on links to open Web sites
Student's	answer

How did you figure out your answer (what criteria did you use?)
Do you search for information on the Web?
Do you usually think about which sites might have reliable information?

Do you usually think about which sites might have reliable info (If yes) How do you usually decide that?

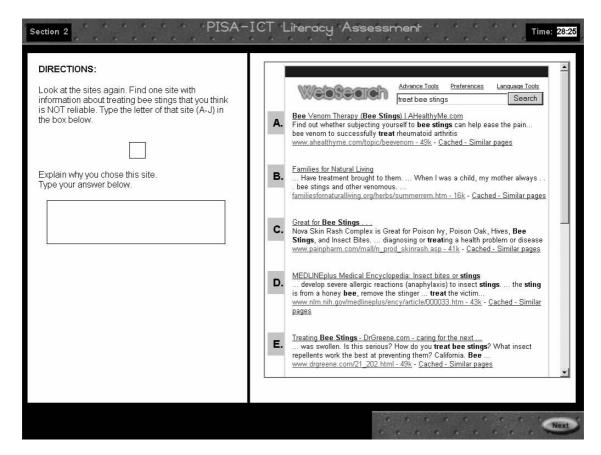


Follow up



[If student asks, this multiple-choice question is an experiment to see if we can get information that is similar to what we get if we allow students to type their answers - free response.]





Correct responses for the second part include B, C, E, I - .have information about treating bee stings, but with more questionable sources - commercial interests that could be biased, etc.

#### Observation

_	Sere	ıll	to.	200	all	aha	Nin.	٠.

☐ Types explanation

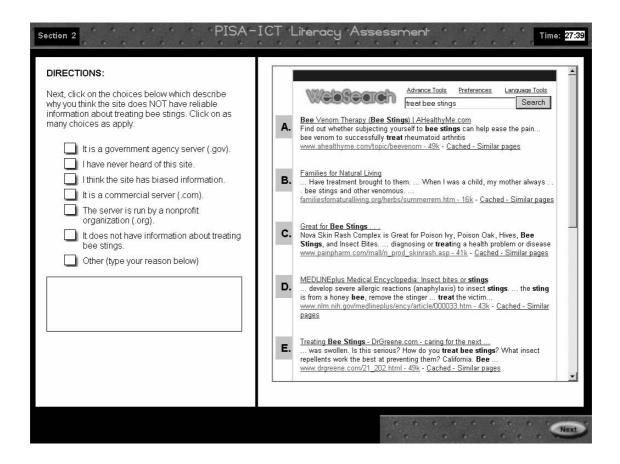
☐ Tries to click on links to open Web sites

Student's answer \_\_\_\_\_

#### Follow up

How did you figure out your answer for this question (what criteria did you use?) Did you find it any easier/harder to focus on finding sites that were not so reliable?





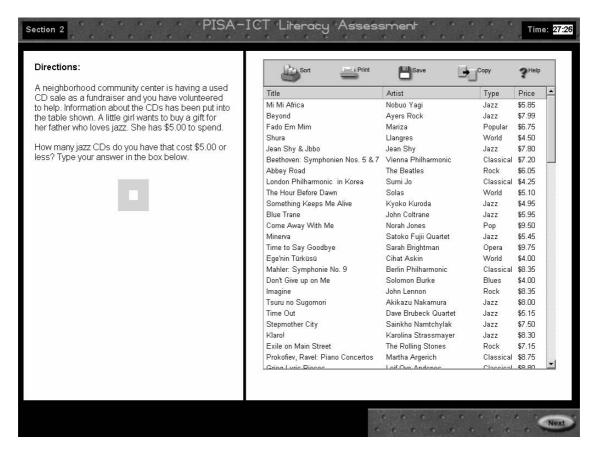
#### Summary

Was the student able to easily complete the web abstracts task?

Yes

No





Correct answer = 3

0	b	s	е	r١	V	а	ti	0	r	١

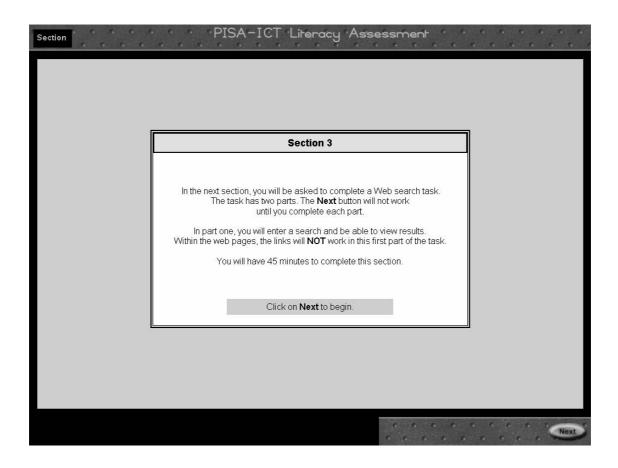
General process followed		
☐ Scrolls down list and counts to arrive at answer		
☐ Clicks on 'Sort' and sorts on one variable		
☐ Clicks on 'Sort' and sorts on two variables	and	
☐ Sorts more than once		
☐ Clicks on incorrect icons - circle all that apply: Print☐ Clicks on 'Help'	Save Copy	
Student's answer		
s the student able to easily complete the database task?	Yes No	

#### Follow up

Tell me how you found your answer.

If student did not sort, probe for why - didn't notice icon, not familiar with function, didn't think it was necessary for this task.

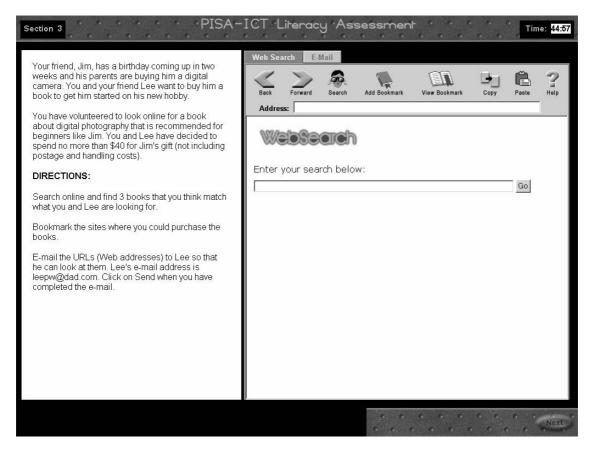




# Follow up

Were these directions clear? (Interested to know if student noticed that links do not work in first part of task.)





Correct sites: Booksellers.com (Booksellers: Digital Photography: A Basic Guide for Beginners); Armingtonpress.com (Learning Digital Photography); Narmada.com (Using Your Digital Camera)

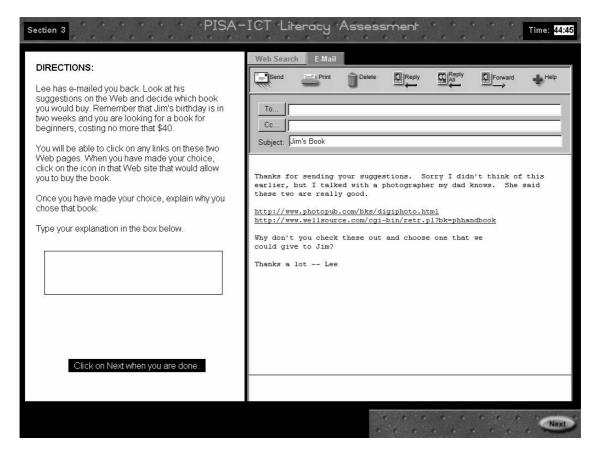
Observation			
☐ Opens one or more sites			
☐ Clicks on 'Next 15 documents' to go beyond 1 <sup>st</sup> page of hits			
☐ Tries links on web pages			
☐ Uses bookmark feature for (number of) pages			
☐ Copies and pastes URLs into email			
☐ Types URLs into email			
☐ Uses e-mail/web tabs to toggle back and forth			
☐ Clicks on 'Help'			
☐ Revises query			
☐ Clicks on 'Send' when finished			
Is the student able to easily complete the web search task? Yes			

Follow up

Explain how you went about this task. Have you used bookmarks before?



No



Correct choice - All You Need to Know about Digital Photography (site #1). Second choice ('The Digital Photography Handbook') is not available within two weeks as needed for Jim's birthday.

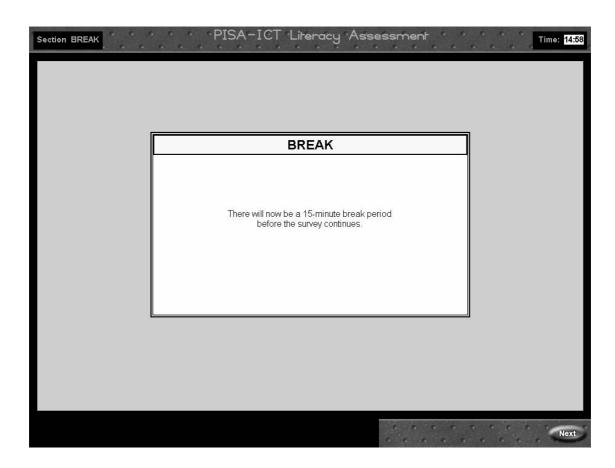
#### Observation

- ☐ Clicks on link to view web site #1 (All You Need to Know about Digital Photography)☐ Clicks on link to view web site #2 (The Digital Photography Handbook Wellsource)
- ☐ Clicks on 'Check Availability' link on Wellsource page
- ☐ Clicks on 'Add to Shopping Cart' icon to select 'All You Need to Know' (correct choice)
- ☐ Clicks on 'Buy Now' icon to select 'The Digital Photography Handbook'
- ☐ Types explanation

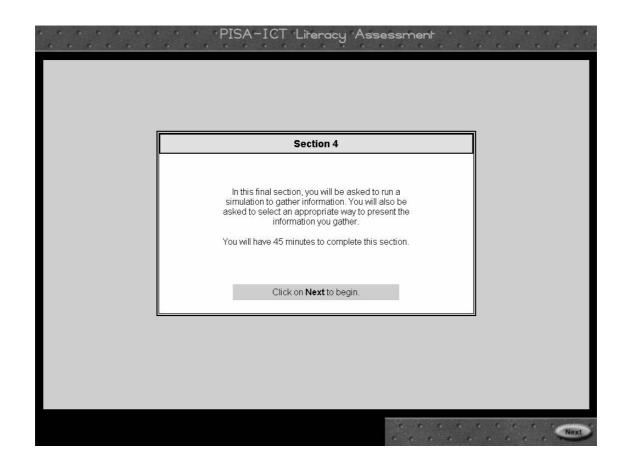
#### Follow up

If student did not type explanation, probe for reason (didn't notice type-in box, didn't choose to respond, etc.)

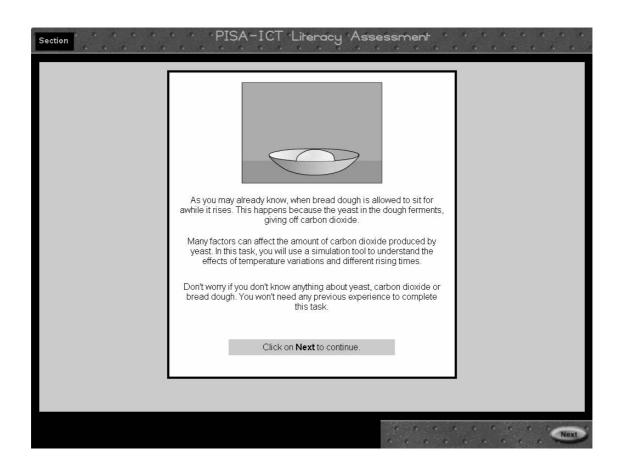






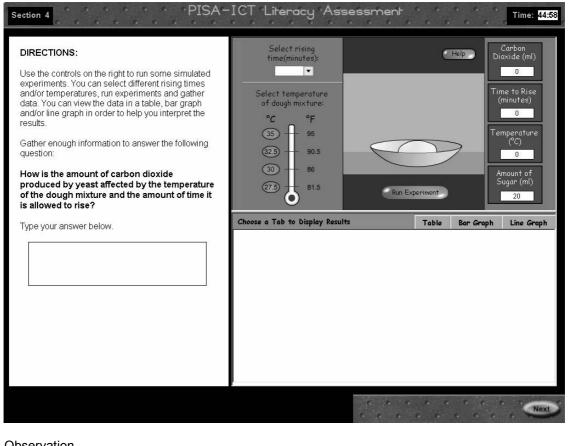






Follow up
Was this explanation helpful?





Observation	
Actions	Initial sequence (1 for first action taken, 2 for second, etc.)
☐ Selects rising time	
☐ Selects temperature	
☐ Runs experiment	
□ Makes table	
☐ Makes bar graph	<del></del>
⊟ Makes line graph	<del></del>
☐ Clicks on 'Help'	<del></del>

Does the student appear to run the experiments in any particular sequence? Examples include:

Holding temperature constant and running experiments across multiple rising times
Holding rising time constant and running experiments across multiple temperatures
Other:

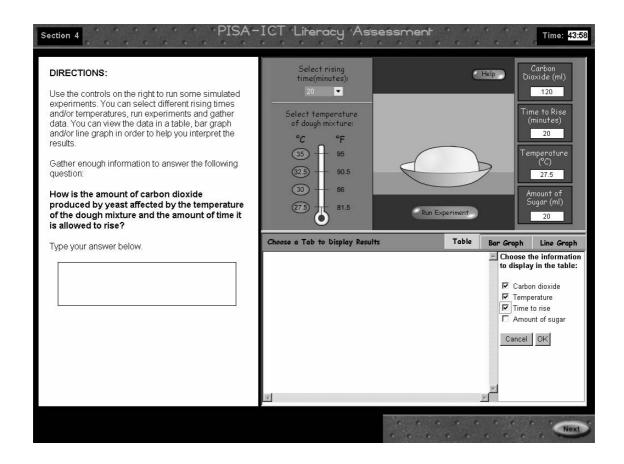
Appears to be random

Does the student make new table or graphs as more experiments were run? Yes No Any pattern? (e.g., made new table after each experiment or after several experiments, collected all information in a table before displaying in graph, etc.)

#### Follow up

How did you decide where to start with this task? How did you decide what experiments to run? Did you find any part of the task confusing?





#### Observation - Creating a table

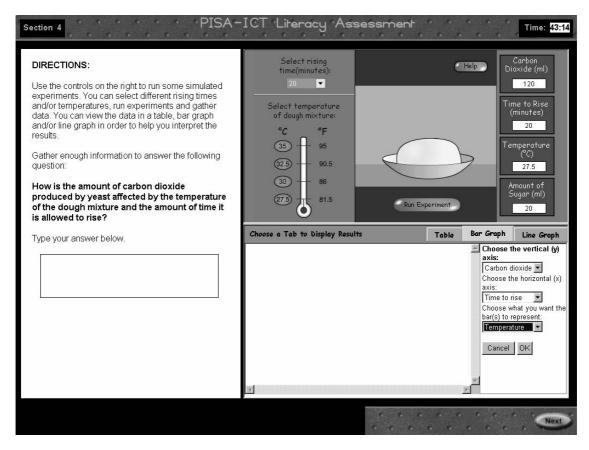
- Selects variables for graph: Carbon dioxide Time to rise Temperature Sugar
- ☐ Sorts information in the table (by clicking on column headers)
- ☐ Updates table as more experiments are run

Comments about how the student appears to use the table function:

#### Follow up

[If the student made one or more tables] Could you explain how you made the table and how you used the information in the table (e.g., to track which experiments you'd run, to decide what additional information you needed, to make a judgment about the results)?





Graphs - one helpful graphical display requires selecting 3 variables (carbon dioxide, time to rise and temperature) and having carbon dioxide on the y axis, time to rise on the x axis, and the bars or lines representing temperature.

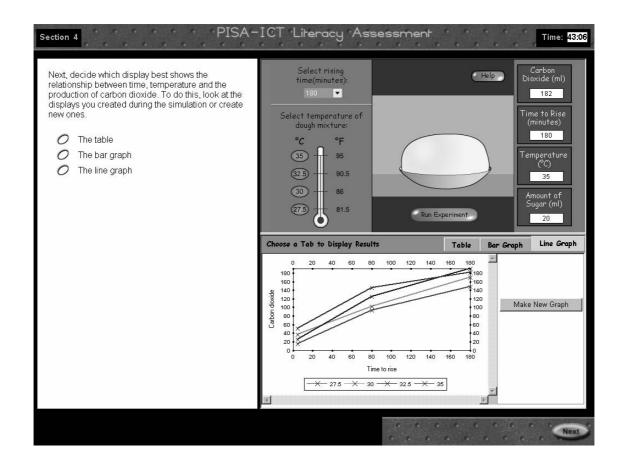
#### Observation - Creating a bar graph ☐ Selects variables for graph: Carbon dioxide Time to rise Temperature Sugar ☐ Selects variables for axes (circle choice) X axis Carbon dioxide Time to rise Temperature Sugar Y axis Carbon dioxide Time to rise **Temperature** Sugar Bars/lines Carbon dioxide Time to rise Temperature Sugar Observation - Creating a line graph ☐ Selects variables for graph: Carbon dioxide Time to rise Temperature Sugar ☐ Selects variables for axes (circle choice) X axis Carbon dioxide Time to rise Temperature Sugar Y axis Carbon dioxide Time to rise Temperature Sugar Bars/lines Carbon dioxide Time to rise Temperature Sugar

Comments about when and/or how the student appears to use the graph function (after each experiment, after all experiments are finished, to decide what experiments to run next, etc.):

#### Follow up

[If the student made one or more graphs] Could you explain how you made the graph and how you used the information in the graph (e.g., to track which experiments you'd run, to decide what additional information you needed, to make a judgment about the results)?





Observation - multiple-choice question (display)

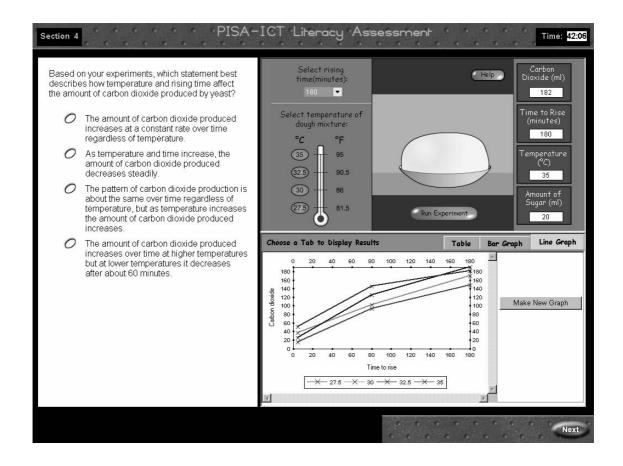
☐ Answers question based on existing experiments ☐ Runs more experiments before answering

Student's answer - circle: table bar graph line graph

#### Follow up

How did you decide on your answer?





Observation - multiple-choice question (display)

Answers question based on existing experiments

☐ Runs more experiments before answering

Student's answer - circle: A B C D

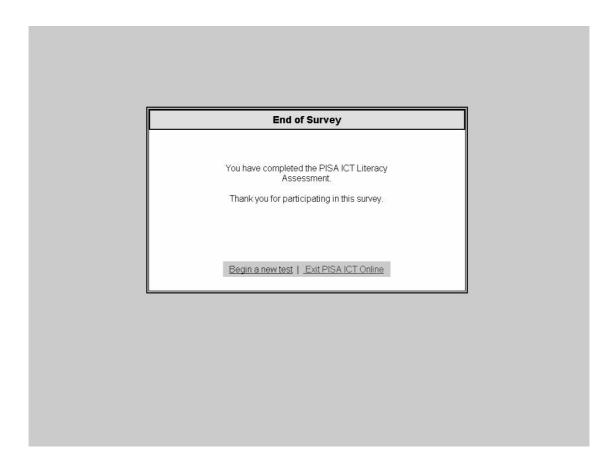
#### Follow up

How did you decide on your answer?

#### **Summary**

Was the student able to easily complete the simulation task? Yes No





# **SECTION 4 - TEST ADMINSTRATOR'S MANUAL**

PISA-ICT Literacy Assessment Feasibility Study Important Information at a Glance				
Important Phone Numbers and E-mail Addresses				
Mike Wagner (ETS, Software)	mwagner@ets.org			
Marylou Lennon (ETS, Project Direction)	mlennon@ets.org			
Altamese Jackenthal (ETS, Operations)	ajackenthal@ets.org			
Data Collection				
Marten Koomen – ACER (Australia)	koomen@acer.edu.au			
Ryo Watanabe – NIER (Japan)	ryo@nier.go.jp			
Holly Knott – ETS (United States)	hknott@ets.org			
Important Surv	Important Survey Information			
Administrator Password	pisaICT			
Student ID Number Range (Australia)	1001-1035			
Student ID Number Range (Japan)	2001-2035			
Student ID Number Range (United States)	3001-3035			
Important Softw	Important Software Information			
F5	Refresh			
Alt + F4	Close browser window and Reattach			
Ctrl + Alt + Del	If ALT+F4 fails to close the browser window			
Alt + S	Pause the Survey			
Alt + R	Skip Background Questions			



# PISA-ICT Literacy Assessment Feasibility Study Session Materials and Procedures

#### Materials for each testing session

- Administration Record
- Follow-up questionnaires for each student
- Event log(s)
- Payment receipts (if applicable)

#### **Providing Help**

As a general rule, test administrators should try not to intervene during the testing session. However, it is likely that there will be instances in which questions or situations may arise. In order to ensure a consistent response in these situations, please see the table below.

Problem	Response
Student has difficulty reading directions or other text.	Ask the student to try and continue the task.  If that is not successful, provide the minimum amount of help possible to allow the student to continue. Note the text that presented a problem in the event log.
Student cannot figure out how to move off a screen. For example, on the last screen of the e-mail task if the student does not click <b>Send</b> , <b>Next</b> is not active and the student cannot continue.	<ol> <li>Ask the student to take another look at the screen and see if he/she can figure out something else to try.</li> <li>Tell the student how to complete the task (e.g., "Remember that you have to send the email when you are finished.")</li> </ol>
Student sits for several minutes without doing anything (appears not to be reading or thinking about a task).	Ask the student if there is a problem and try to help resolve it.
Student has never used a computer and is unable to navigate through the system.	<ol> <li>Assist the student through the introductory screens and the background questionnaire (click on 'Next', click or type responses.)</li> <li>Encourage the student to try the remaining tasks in section 1 (scrolling, etc.) so that we collect some data on basic technical skills.</li> </ol>

Be sure to note any specific problems as well as the help provided in the event log. Test administrators who are observing a retrospective interview student may choose to record this type of intervention in the retrospective interview packet. At some point after the session is completed, please transfer such information into the event log so that we have a single record for data collection purposes.



#### Materials for retrospective interviews

- Master set of screen print outs (for reference during interview if needed)
- One packet (annotated screen prints outs) to record observations and followup discussion for each student participating in a retrospective interview. It is suggested that the student's ID# be typed into the space provided in the document's heading before printing each packet. This will identify each page in the packet should materials be misplaced.

#### Procedures - Retrospective Interview

- Each student participating in a retrospective interview will be observed throughout the testing session by a test administrator.
- The test administrator will record observations in the packet as the student goes through the testing material.
- Once the student has completed the test, he or she will be asked to complete the follow-up questionnaire.
- Finally,\_the test administrator will interview the student, going though each of the tasks as noted in the packet. Additional questions may be asked as warranted. The test administrator may also want to review the follow-up questionnaire responses with the student and probe for additional information.



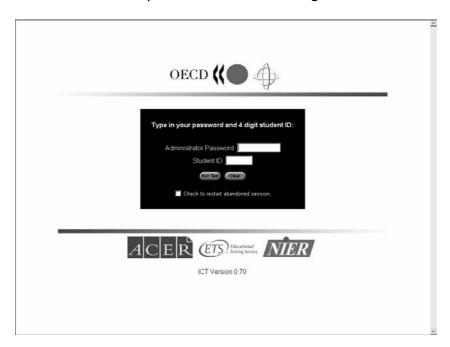
# PISA-ICT Literacy Assessment Feasibility Study How to Run the Software

#### Starting the Test

- 1. Connect to the internet and run **Internet Explorer** (Version 6.0).
- 2. Type one of the following URLs:
  - For the English version: <a href="http://naeptba.ets.org/ICTE.asp">http://naeptba.ets.org/ICTE.asp</a>
  - For the Japanese version: <a href="http://naeptba.ets.org/ICTJ.asp">http://naeptba.ets.org/ICTJ.asp</a>

#### Logging On

After you connect to the site you will see the following screen:



1. Type the Administrator Password: pisaICT Note that this password is case sensitive and must be typed exactly as it appears above.

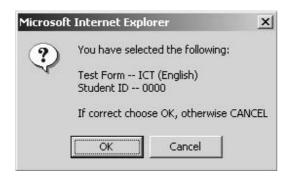


2. Type the four-digit Student ID number. Each student must be assigned a unique ID number from the following number ranges:

Country	Student ID Number Range		
Australia	1001-1035		
Japan	2001-2035		
United States	3001-3035		

If you have more than 35 students participating in the study, the number range can be extended as needed (for example, in Australia the numbers 1036, 1037, etc. could be added).

3. Click on **Run Test**. You will see the following message:



4. Click on **OK** to begin the test. If you have made an error, click on **Cancel** and then **Clear** to erase the password and student ID number.

Clicking on **Check to Restart Abandoned Session** will resume the test at the point you left it if you have to exit the test before it is completed.

#### Pausing the Test

If you need to end the test early, press ALT+S. You will see the following message.



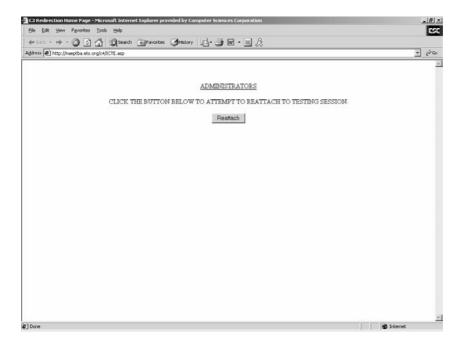
The timer stops until you resume the test. Click on OK to continue with the test.



Test Administrator's Manual – How to Run the Software

# **Ending the Test Early**

If you need to end the test early, press ALT+F4 or Ctrl+W. You will see the following message.



If you want to end the test session, close the window. <u>Note that if you end the test early, the student's data will not be saved.</u>



Test Administrator's Manual – How to Run the Software

#### **Break Period**

Between sections 3 and 4 of the test there will be a 15-minute break period. The break is mandatory and cannot be bypassed. The time will count down on the upper left of the screen.



After the time has elapsed, the following message will display.



Click on **Next** to continue the test.



Test Administrator's Manual – How to Run the Software

# Closing the Testing Session

After you have completed all four test sections, the following screen will display:



Click on **Begin a new test** to start testing another student.

Click on **Exit PISA ICT Online** to exit.



# PISA-ICT Literacy Assessment Feasibility Study Troubleshooting

If there are any software problems that prevent you from moving forward, try the three "Rs" in this order.

- **Refresh** Press F5 to refresh the screen. If this is not successful, try the reattach process.
- **Reattach** Hold down ALT while pressing F4 to close that browser window. As a result the Reattach browser window will be displayed. Click on REATTACH to continue the session.

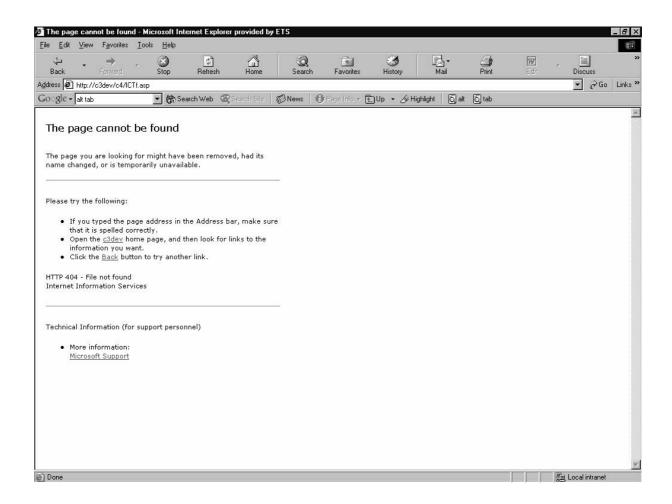
If ALT + F4 is not successful in closing the browser window, use CTRL + ALT + DEL to shut down and restart the computer. Follow the instructions for restarting a session.

• **Restart** – If the reattach process was not successful, reboot the computer and restart the survey. At the Login Screen type in the Student ID for the session you want to return to, then check the box for restarting an abandoned session.

The following pages contain examples of screen shots that you or your students might encounter during the survey.

- Usual and expected messages and warnings
- Pop-up messages or dialog boxes caused by user error (e.g. inadvertent keystrokes)

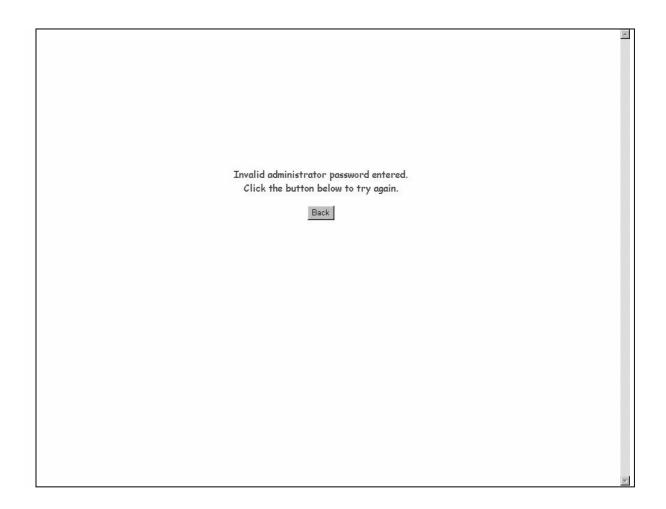




**What happened?** An incorrect URL was typed.

**How do I fix it?** Correct the URL and press enter.

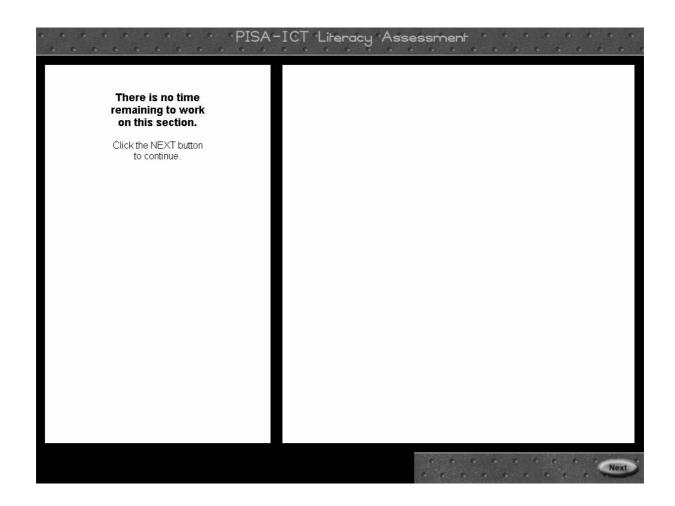




**What happened?** An incorrect password was typed.

**How do I fix it?** Click on the BACK and try again.

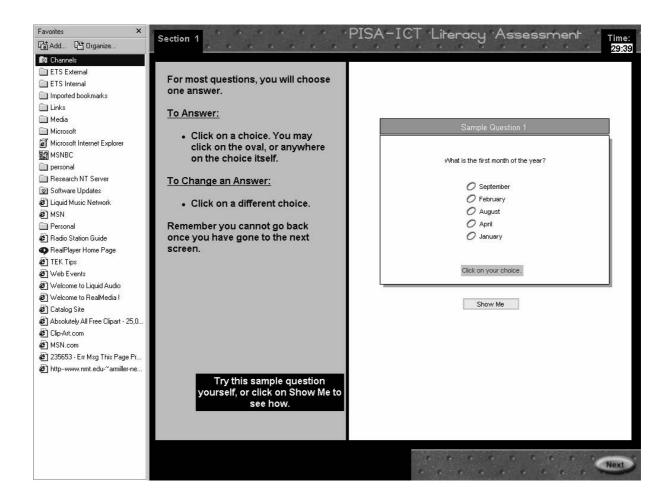




**What happened?** Section time has expired.

**How do I fix it?** Click on NEXT to continue.

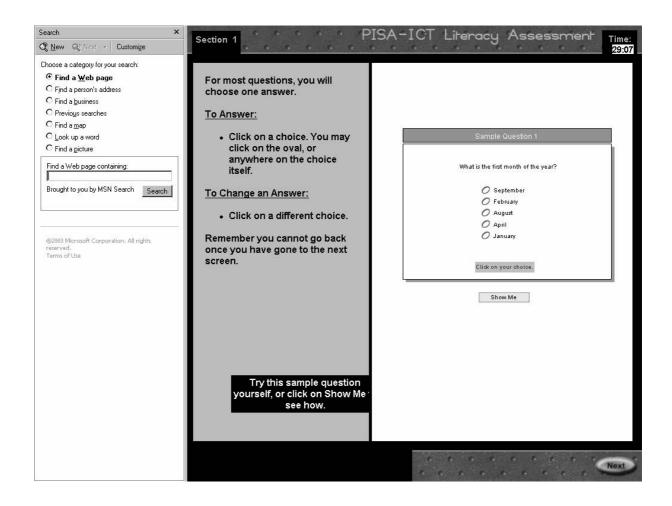




What happened? The Favorites list displayed as a result of the keystrokes "CTRL + I".

**How do I fix it?** Close the Favorites window by clicking on the X in the upper right corner of the window.

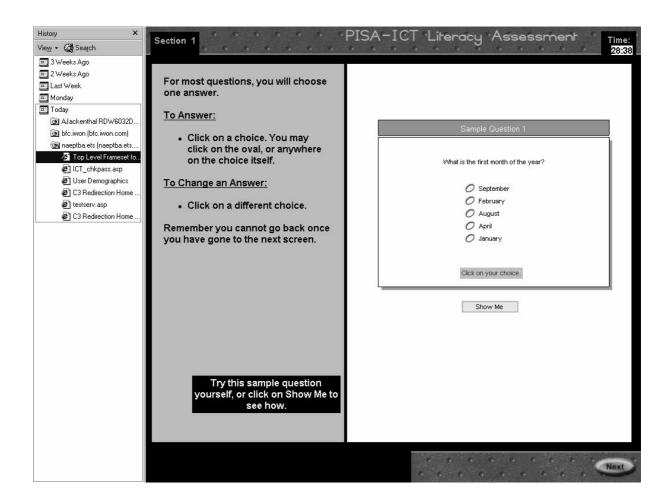




**What happened?** The Search window displayed as a result of the keystrokes "CTRL + E".

**How do I fix it?** Close the Search window by clicking on the X in the upper right corner of the window.

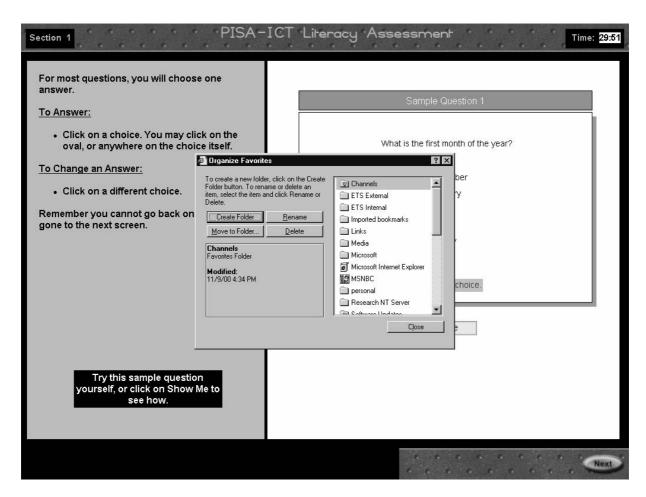




**What happened?** The History list displayed as a result of the keystrokes "CTRL + H".

**How do I fix it?** Close the History window by clicking on the X in the upper right corner of the window.

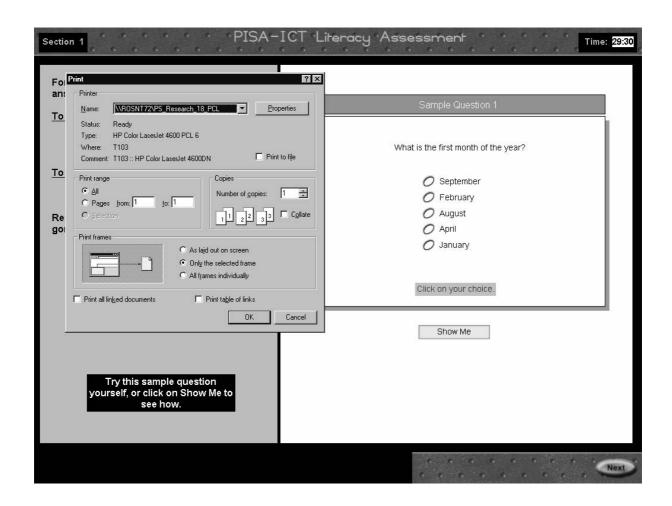




**What happened?** The Organize Favorites dialog box displayed as a result of the keystrokes "CTRL + B".

**How do I fix it?** Close the Organize Favorites dialog box by clicking on the X in the upper right corner of the window or click on CLOSE.

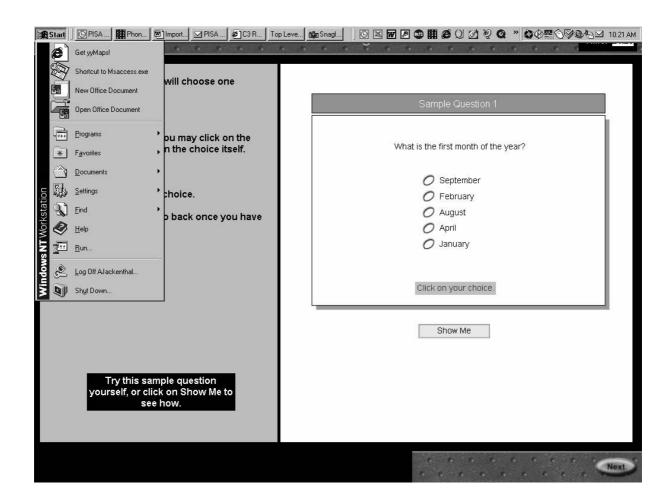




**What happened?** The Print dialog box displayed as a result of the keystrokes "CTRL + P".

**How do I fix it?** Close the Print dialog box window by clicking on the X in the upper right corner of the window or click on CLOSE.

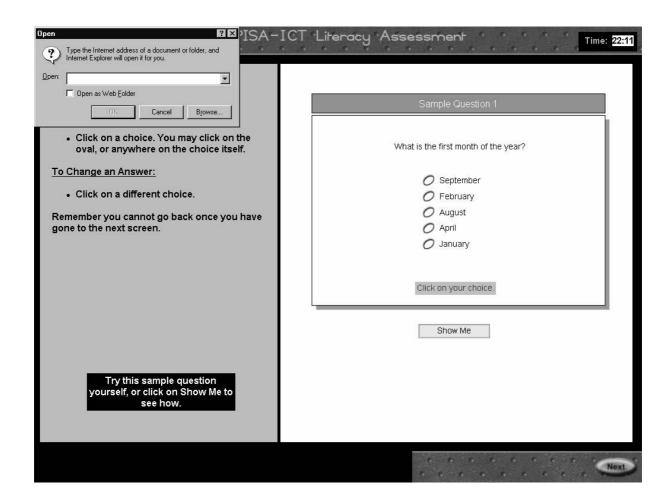




**What happened?** The Window's Start Menu displayed as a result of the keystrokes "CTRL + ESC".

**How do I fix it?** Close the Window's Start Menu by pressing ESC.

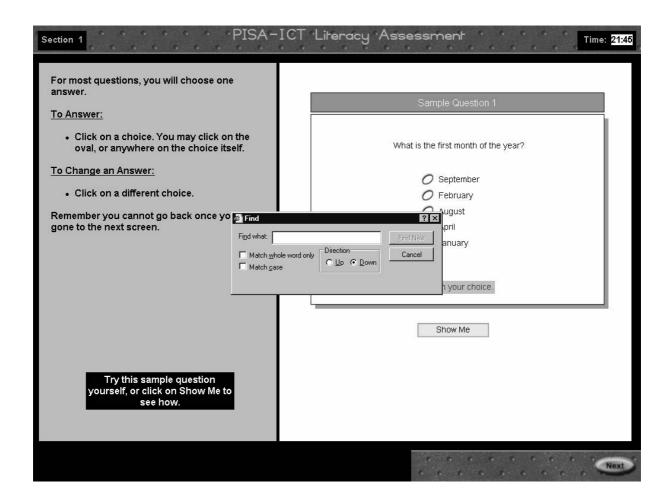




**What happened?** The Internet Explorer Open dialog box displayed as a result of the keystrokes "CTRL + L".

**How do I fix it?** Close the Internet Explorer Open dialog box by clicking on the X in the upper right corner of the window or click on CANCEL.

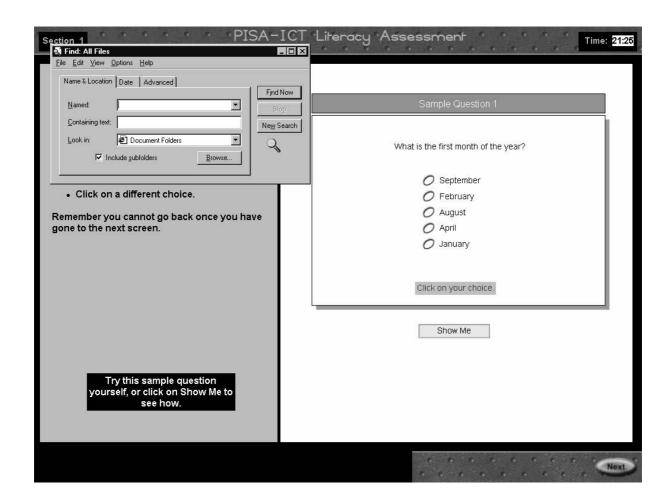




**What happened?** The Internet Explorer Find dialog box displayed as a result of the keystrokes "CTRL + F".

**How do I fix it?** Close the Internet Explorer Find dialog box by clicking on the X in the upper right corner of the window or click on CANCEL.

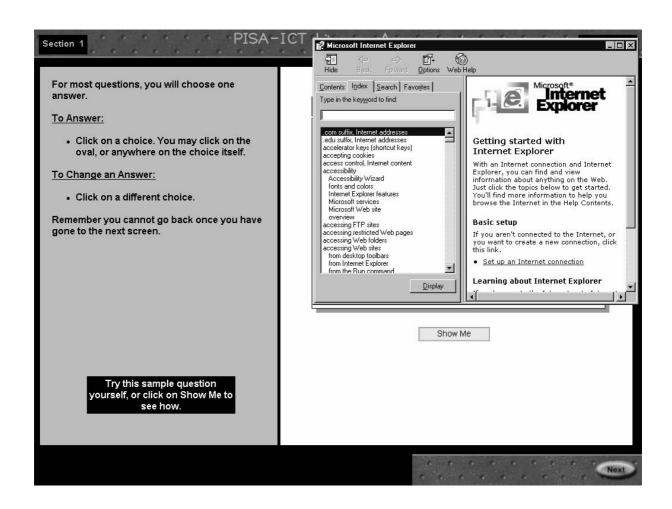




**What happened?** The Internet Explorer Find All Files dialog box displayed as a result of the keystroke "F3".

**How do I fix it?** Close the Internet Explorer Find All Files dialog box by clicking on the X in the upper right corner

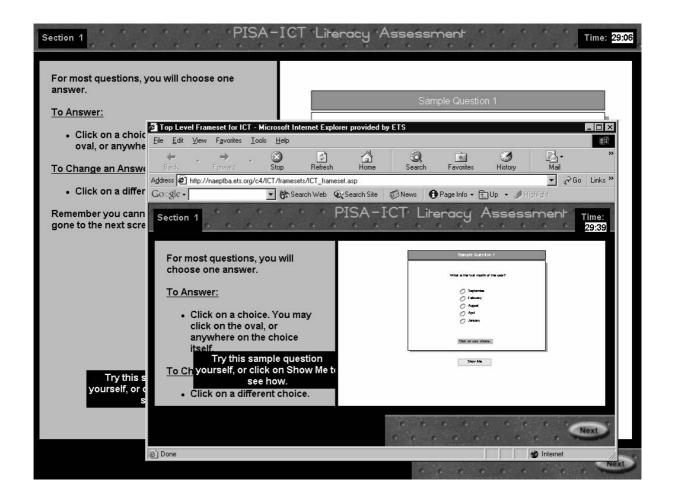




**What happened?** The Internet Explorer Help window displayed as a result of the keystroke "F1".

**How do I fix it?** Close the Internet Explorer Help window by clicking on the X in the upper right corner of the window

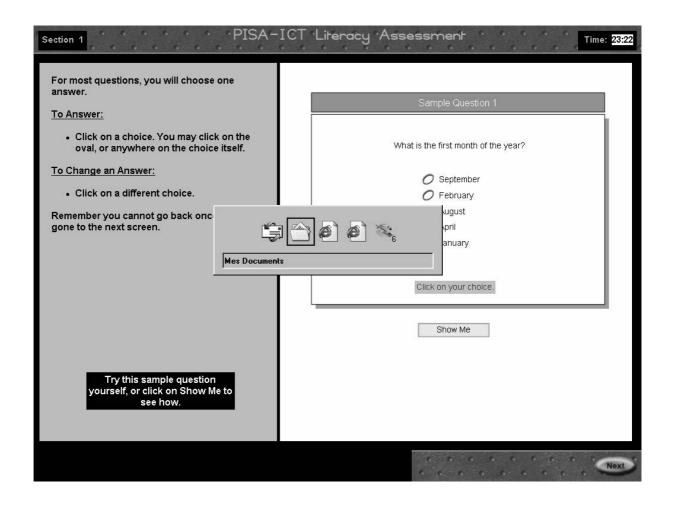




**What happened?** The current page is displayed in a new browser window as a result of the keystrokes "CTRL + N".

**How do I fix it?** Close the new browser window by clicking on the X in the upper right corner of the window.

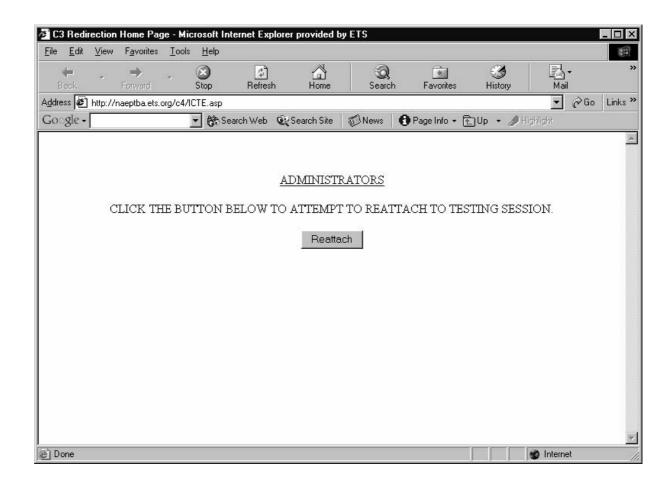




**What happened?** The Windows Tasks pop-up is displayed as a result of the keystrokes "ALT + TAB".

**How do I fix it?** While holding down ALT, press TAB until the Internet Explorer window labeled "Top Level Frameset for ICT" is selected.





**What happened?** The PISA-ICT reattach window displayed as a result of the keystrokes "CTRL + W" or "ALT + F4".

**How do I fix it?** Click on REATTACH to continue.



# PISA-ICT Literacy Assessment Feasibility Study Event Log Instructions

An Event Log should be completed whenever an irregularity occurs. This would include:

- Computer problems or incidents
- Providing help to a student that is important to note in order to properly interpret the data (for example, telling a student where to click so that he or she can continue with a task).
- Student questions about the survey, during the survey
- Ending a session early (before the survey has been completed)

If there is something to report, complete the following:

- Test Administrator Enter your name (for follow up, if necessary). If more than one administrator has something to report, each should do so on a separate log.
- Date/Session Enter the date and session (if applicable).
- Student ID Enter the number placed in the "Student ID" box of the Log In Screen.
- Computer Number/Type Enter the computer number or type (i.e., #7, laptop, etc.)
- Test Section Enter the test section number located in the upper left corner of the screen.
- Screen Briefly describe the screen (i.e., forward e-mail screen, second e-mail screen, etc.).
- Description of Problem, Action Taken, and Resolution Giving as much detail as possible record the problem, action taken and resolution. Use as many boxes as needed to record the incident.



DIGA TOTAL		
PISA-ICT Lite	eracy Assessme	nt Feasibility Study Event Log
Date/Session		Description of Event, Action Taken, and Resolution (if any):
Student ID		
Computer Number/Type		
Test Section		
Screen		
Date/Session		Description of Event, Action Taken, and Resolution (if any):
Student ID		
Computer Number/Type		
Test Section		
Screen		
Date/Session		Description of Event, Action Taken, and Resolution (if any):
Student ID		
Computer Number/Type		
Test Section		
Screen		
Date/Session		Description of Event, Action Taken, and Resolution (if any):
Student ID		
Computer Number/Type		
Test Section		
Screen		

