Outgrowing the Mode Effect Study of Paper and Computer Based Testing

Kentaro Yamamoto
Educational Testing Service, USA

Introduction

Changes are inevitable as we adjust to a new environment, to meet the needs expressed by others or sometimes just have fun, and besides nothing has ever stayed the same forever. Not all of these changes are frivolous and without a functional purpose. A new idea or product gets introduced and if it is useful it spreads, and then supersede the old one. Sometimes such changes result in major societal changes. Examples are numerous, printing press over hand writing, guns over bows and arrows, chopsticks over fingers, list can go on and on. Often happens is that when you examine these changes, changes introduce new dimensions to the old practices and we start seeing something entirely new, you can cook food with chopsticks without burning your fingers. What we are seeing now with written materials on a static paper being replaced by electronic media resembles one of the many historical changes in the past. While activity of reading remains the same in many respects in terms of visual information differentiated by the positioning and the sequences of symbols, many aspects have changed.

I would like to present coarser perspectives than my usual data driven knowledge building work, this is a perspective about forest instead of trees. I think it is very important to have clear idea about who the messenger is and perspectives I have. I have been building Psychometric models that represent a particular understanding of skills and interest over 30 years. After finishing my Mechanical Engineering training I started my Psychometric journey with professors of latent structure models such as Jim Paulson of Portland State University and Kikumi Tatsuoka at PLATO lab of University of Illinois and at ETS I have lead and leading various large scale surveys of children and adult populations in US such as NAEP, NALS and also as international population surveys such as IALS, ALL, PIAAC and PISA. I am interested in individual differences while recognizing overwhelming

common structures among us. I am a student of Psychometric modeling that capitalizes this common structure of human cognition and development.

I believe that we are living at the time in the history where we are entering more digital world and harnessing something we did not imagine before.

Proliferation of digital media

Let us consider characteristics of electronic text and electronic information in general for learning ideas and increasing knowledge. Nearly all youths in industrialized countries are ever approaching close to 100%, i.e., nearly all who wish to access Internet can do so for nominal fees. They are spending increasingly more hours outside of school accessing Internet for whatever reasons. Not only previously print information are adapted to digital format, completely new materials are being developed in digital format. Sheer volume of information is available at our fingertips and growing every day. New type of information is invented allowing higher level of interaction and also making the detail of information more adaptive. Major textbook publishers are switching from paper-based text to digital publishing in increasing numbers. Digital textbooks are organized differently and have links among all concepts and terminologies in the book as well as links to ever-current knowledge on the WEB. There are many web based resources on traditional academic subjects such as biography, history, science, mathematics, arts, music and as well as hobbies, games and entertainment. Also there are many lessons and instructions produced by many authoritative institutions and organizations on many subjects including very traditional academic subjects. And often they are constantly updated to reflect current thoughts when they are supported by an organization or shared openly.

Where learning takes place

Historically, formal education in many educational institutions throughout the country followed a curriculum or a set of curriculums often developed centrally and uniform applications were desired and generally achieved. The students come out of such educational institutions share the common experiences of learning as well as more or less common knowledge reflecting the well-prepared and formulated curriculum. It can be

said that the epicenter of learning has been in school, and teachers worked hard to stimulate students' curiosity and focused on the transfer of well-packaged knowledge from a teacher to students. Students learned how to communicate to each other and work together. What was difficult under this scenario for a teacher is to teach what each student needs to know beyond what he/she knows already. Everyone in the classroom would receive identical series of lessons for example on "Roman History" regardless of what each one uniquely knows due to their own learning outside of classes. Differentially prepared students may not be accommodated very well.

Increasing differentially prepared students in a classroom

This widely available and extensive academic information on the digital media to everyone enables an interested student to learn its own pace on whatever he/she chooses and as deeply as one wishes without any limitation imposed by the school building or school schedule. That means not all students educational/learning experiences would be as uniform as they have been in the past. Students can pursue own interest or needs whenever and wherever. Students may not be the only driver but also parents who can monitor children's' learning experiences. Classroom teachers should expect more variability in preparedness and prior knowledge and skills as they receive new students. Should the role of teachers be the same as predominantly paper and pencil world as primarily digital world in regard to information?

The real place of learning

The location of student learning in regard to the cognitive knowledge and skills may be overlapping with outside of school walls. It may not be limited to home but probably including everywhere else due to access to digital information can be done from anywhere. There is a practical limitation to increase access to digital media by student at school due to physical, financial or temporal reasons. Already primary place of interaction with digital information for student is outside of schools. However this does not limit to the domain of Internet but also recent and not so recent expansion of schools after formal schools, often called "Juku", originated in Asian countries in 1960s. Even in late 1990s, nearly 2/3 of lower secondary students in Korea, Japan, Hong Kong, and Singapore were attending "Juku"

throughout the country, and in higher rate in urban areas. Expenditure for Juku in Japan in 2010 surpassed the cost of formal education for elementary, lower secondary and secondary schools. Average expenditure for these schools can run from $2500 to $7000 a year depending on the age of students. Some elite Juku cost over $17000 a year to prepare for High School entrance examinations. So this looking outside to supplement or enhance own formal education as well as in preparation for entrance examinations is not anything new and great deal of efforts in terms of time and money have been expended. Some of these Juku franchises have been spreading recent years.

Quality and authenticity of information

Consistency and quality of information on the WEB is neither predictable nor stable. How the web looks visually is no proof for the authenticity and accuracy of information. Evaluation of information cannot be over-emphasized in regard to the internal consistency among the information presented as well as the organization that brought the information. Any information may need to be cross checked Students should acquire not only the information they are seeking but also skills to conduct evaluation of what they learn, this is akin to meta-cognitive skills. Students should exercise judicial skepticism on whatever they may encounter on the internet. Security and safety of Internet not only for the digital information but also social connection should be monitored.

What we can learn from surveys

We all know that even among printed texts such as novels, textbooks, letters, newspapers, telephone directories, and scientific journal articles can vary great deal in organization and structure. These unique textural characteristics of genre are optimized for information transmission from a writer to readers. Some ideas can be presented easily on tables and text compared to text alone such as describing the bus schedule without table is very complicated and nearly impossible. On the other hand, a story would be best told in series of proses.

Since early 1990s, we have been assessing literacy proficiencies defined in terms of human behavior as "Using printed and written information to function in society, to achieve one's

goals, and to develop one's knowledge and potential." Reading expert group defined literacy in terms of three domains, each encompassing a common set of skills relevant for diverse tasks:

Prose Literacy – the knowledge and skills needed to understand and use information from texts including editorials, news stories, poems and fiction:

Document Literacy – the knowledge and skills required to locate and use information contained in various formats, including job applications, payroll forms, transportation schedules, maps, tables, and graphics: and

Quantitative Literacy – the knowledge and skills required to apply arithmetic operations, either alone or sequentially, to numbers embedded in printed materials, such as balancing a chequebook, figuring out a tip, completing an order form or determining the amount of interest on a loan from an advertisement.

We applied these constructs on the special samples of US National Assessment of Educational Progress in 1984, National Adult Literacy Surveys in 1992, International Adult Literacy Survey in 1994, international Adult Literacy and Life skills Survey in 2008. All are household surveys of adult population and data collection took place at respondent's house with an interviewer present. Items appeared developed for these surveys all share similar characteristics by not requiring much external knowledge, it is not a memory test of what you have learned elsewhere. Rather they are to locate the relevant information and integrate or manipulate information and make inferences. They emulate our daily experiences to seek the information we need and use for our purposes. Two examples of Prose and Document literacy items are shown below.

Example 1:

IMPATIENS

Like many other cultured plants, impatiens plants have a long history behind them. One of the older varieties was sure to be found on grandmother's windowsill. Nowadays, the hybrids are used in many ways in the house and garden.

Origin: The ancestors of the impatiens, *Impatiens sultani* and *Impatiens holstii*, are probably still to be found in the mountain forests of tropical East Africa and on the islands off the coast, mainly Zanzibar. The cultivated European plant received the name *Impatiens walleriana*.

Appearance: It is a herbaceous bushy plant with a height of 30 to 40 cm. The thick, fleshy stems are branched and very juicy, which means, because of the tropical origin, that the plant is sensitive to cold. The light green or white speckled leaves are pointed, elliptical, and slightly indented on the edges. The smooth leaf surfaces and the stems indicate a great need of water.

Bloom: The flowers, which come in all shades of red, appear plentifully all year long, except for the darkest months. They grow from "suckers" (in the stem's "armpit").

Assortment: Some are compact and low-growing types, about 20 to 25 cm. high, suitable for growing in pots. A variety of hybrids can be grown in pots, window boxes, or flower beds. Older varieties with taller stems add dramatic colour to flower beds.

General care: In summer, a place in the shade without direct sunlight is best; in fall and spring, half-shade is best. When placed in a bright spot during winter, the plant requires temperatures of at least 20°C; in a darker spot, a temperature of 15°C will do. When the plant is exposed to temperatures of 12-14°C, it loses its leaves and won't bloom anymore. In wet ground, the stems will rot.

Watering: The warmer and lighter the plant's location, the more water it needs. Always use water without a lot of minerals. It is not known for sure whether or not the plant needs humid air. In any case, do not spray water directly onto the leaves, which causes stains.

Feeding: Feed weekly during the growing period from March to September.

Repotting: If necessary, repot in the spring or in the summer in light soil with humus (prepacked potting soil). It is better to throw the old plants away and start cultivating new ones.

Propagating: Slip or use seeds. Seeds will germinate in ten days.

Diseases: In summer, too much sun makes the plant woody. If the air is too dry, small white flies or aphids may appear.

Question 1: According to the article, what do the smooth leaf surfaces and the stems suggest about the plant?

_____

_____

Example 2:

Questions 12 - 14. **Use the charts on the opposite page to answer questions 12 through 14.**

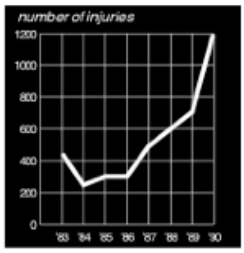
12. In what year were the fewest number of people in the Netherlands injured by fireworks?

13. According to the charts, what was the value, in US dollars, of fireworks sold in the Netherlands in 1991?

14. Describe the relationship between sales of fireworks and injuries due to fireworks.

Fireworks in the Netherlands

Victims of fireworks



Survey results reflected on instruction

We found in all surveys of adult literacy that not all good readers in one genre can read as well in the other genre, and vv. This variability was greater among middle to lower level of readers. Also some of the poorer readers are not aware of the organization and structure of text especially in document and quantitative text. We developed multi-media, group-based system including over 100 hours of instruction and exercises focusing on prose, document and quantitative literacy in mid 1990s called PDQ: Building Skills for Using Print. The PDQ was very unique in a sense that it taught more about how information was organized and structure of text as well as strategies to understand such texts. The PDQ was used for some of industries, community college students, as well as some staffs at ETS. We measured pre and post instructions and found substantial improvement and reviews from students and supervisors were quite positive and long lasting.

Printed texts represent unique organizational and structural conventions that arose from optimizing efficiency of representing particular information. Teaching about the organization and the information structures enhanced comprehension as well as application to the other areas. This notion can be extended to the electronic text.

Learning about a certain subject or understanding ideas often require not only novel like sequential reading but also checking references to verify what you are not very certain about.

Differences between readings on print vs. on screen

Most recently, the concept of literacy was expanded again with the Programme for the International Assessment of Adult Competencies (PIAAC). As the first computer-based, large-scale literacy assessment, PIAAC reflected the changing nature of information, its role in society and its impact on people's lives. As explained below, the scope of the prose, document and numeracy domains was broadened in PIAAC and the assessment incorporated two new domains.

- For the first time, this assessment addressed literacy in digital environments. As a computer-based assessment, PIAAC was able to include tasks that required respondents to use electronic texts including web pages, emails and discussion boards. These stimulus materials included hypertext and multiple screens of information and simulated real-life literacy demands presented by digital media.
- In PIAAC, the definition of numeracy was broadened once again to include the ability to access, use, interpret, and communicate mathematical information and ideas in order to engage in and manage the mathematical demands of a range of situations in adult life. The inclusion of "engage" in the definition signaled that not only cognitive skills but also dispositional elements, i.e., beliefs and attitudes, are necessary to effectively meet the demands of numeracy in everyday life.
- PIAAC included the new domain of problem solving in technology rich environments (PS-TRE), the first attempt to assess this domain on a large scale and as a single dimension. PS-TRE was defined as follows:

    "Problem solving in technology rich environments involves using digital technology, communication tools and networks to acquire and evaluate information, communicate with others and perform practical tasks. The first

PIAAC problem solving survey will focus on the ability to solve problems for personal, work and civic purposes by setting up appropriate goals and plans, accessing and making use of information through computers and computer networks."

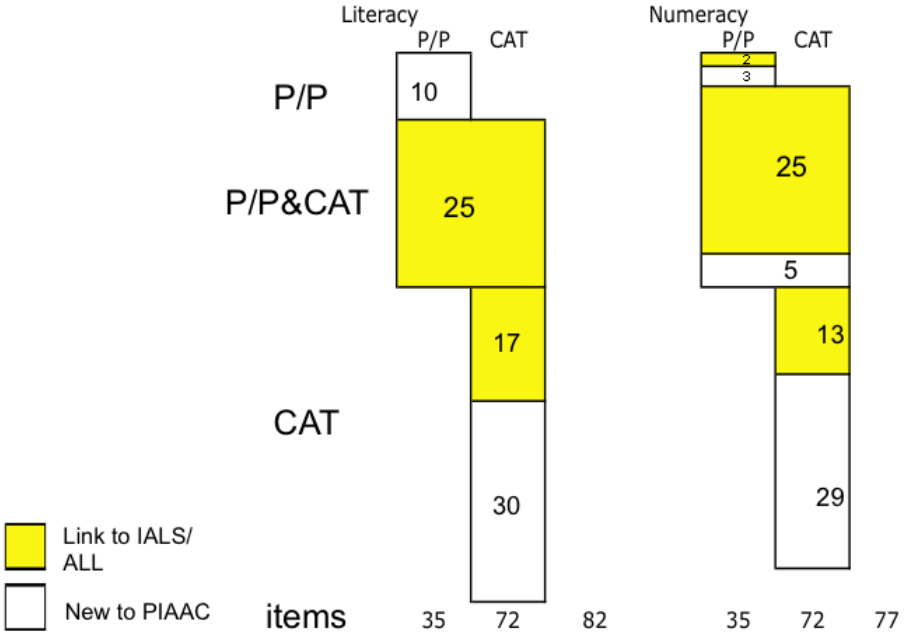
PS-TRE presented computer-based simulation tasks designed to measure the ability to analyze various requirements of a task, define goals and plans, and monitor progress until the task purposes were achieved. The focus was not on computer skills per se, but rather on the cognitive skills required to access and make use of computer-based information to solve problems.

- Finally, PIAAC included a reading components domain, which included measures of vocabulary knowledge, sentence processing and passage comprehension. Adding this domain was an important evolution because it provided more information about the skills of individuals with low levels of literacy proficiency than had been available from previous international assessments. To have a full picture of literacy in any society, it is necessary to have more information about these individuals because they are at the greatest risk of negative social, economic and labor market outcomes.

The PIAAC field study in 2010 was an opportunity to evaluate the equivalence of item characteristics between paper-and-pencil and computer formats. For the field, 82 literacy and 77 numeracy items altogether 159 items were authored in both paper-and-pencil format as well as in computer format. Following graph shows the classification of items. Out of 82 literacy items, 72 items were administered in computer format and 35 items were administered in paper format. Forty-two items in yellow came from the previous surveys thus linkable to IALS and ALL surveys and 40 items in white were developed new for PIAAC. Out of 42 items 25 items were included in both paper and pencil format and computer format. Remaining 17 items appeared only in the computer format. Out of 40 new PIAAC items, 10 items were administered in Paper format only and 30 items were administered in computer format.

For the numeracy items, there were total of 77 unique items and 72 were administered in computer format and 35 were administered in paper format. Forty items came from the previous surveys and 37 items were authored specifically for PIAAC. Out of 40 items, 25 items were included in both paper and pencil format and computer format. In addition to the linking items, 5 new PIAAC items were also included in both formats.

Number of Items in the Field Study



A subset of items was administered to statistically equivalent groups of samples following the design that balanced the administration order and incorporating a classical experimental design. Each respondent received these items in either paper-and-pencil or computer formats. Since groups are statistically equivalent, i.e., aggregate ability distributions at group level are statistically equivalent. Thus performances of any two groups of respondents on common items for both groups on either format can be compared and the difference should be only due to the mode of item presentation.

Altogether 30 countries/language groups and 37,337 respondents participated and submitted to the identical survey design. Based on this data, we were able to examine; 1) comparability of inferences among countries, 2) comparability of inferences across two formats of paper and pencil vs. computer based, 3) comparability of inferences of PIAAC results against previous IALS and ALL surveys, and 4) harvest item characteristic information in preparation for the PIAAC main survey.

Comparability of inferences was examined by using item response theory (IRT) psychometric model. In contrast to classical methods, which operate on sums or aggregated scored responses to tasks; IRT methods use all the information from each individual's pattern of response to develop a common scale for the proficiency variable. An IRT based measurement model describes the relation between the proficiency of a test taker and the response of a test taker to a task. More specifically, a measurement model describes the probability of an examinee with a particular proficiency level giving a correct answer to any given task within the pool of items on the scale.

In order to examine equivalence of item characteristics across countries, common set of item parameters of 2PL IRT model was estimated and found to fit quite well to all countries for all three scales in both modes of presentation. Deviation was fairly small and non-systematic and early all countries and items were found to be conforming to the international parameters. Sample size was too small for each country to estimate country specific item parameters for PIAAC FT data, but large enough to evaluate magnitude of deviation from the common international item parameters.

Equivalence of item characteristics among the literacy and numeracy items common to IALS and ALL on the paper-and-pencil version was examined. Equivalence of IALS/ALL item parameters to the CBT items adapted from IALS and ALL were also evaluated. Previously estimated IALS and ALL item parameters on P/P fit very well to the P/P items adapted for both scales of Literacy and Numeracy. For the IALS/ALL items adapted for PIAAC CBT, previously estimated item parameters fit quite well for the Numeracy scales with a few items showing noticeable deviation from the IALS/ALL item parameters. For the Literacy scales, more items showed clear deviation from the IALS/ALL item parameters.

Examine equivalence of item characteristics of literacy and numeracy items common to paper-and-pencil and computer-based formats. Several items were freed to estimate CBT only item parameters while majority of linking items shared the common item parameters between P/P and CBT items.

What you can concluded from this is that the skills required to read and comprehend printed materials in the area of literacy and numeracy skills on a computer screen are nearly identical to those required for the materials in the paper and pencil format. ICT skills of highlighting, scrolling, and clicking for the extent used in the PIAAC linking items did not have significant impact on the difficulty of items. This further implies measurement constructs remain intact on both formats. So the relatively simple interfaces of computer that can be taught in a few minutes such as scrolling, using mouse, and clicking can be ignored for skill assessment.

Understanding and learning from the written materials on paper or on the screen is identical within a limited boundary of what you can see the static representations. As far as static representations are concerned the same amount of information is contained and structures of information organization can be identical on both modes. Thus difficulties associated with encountering new vocabulary, referencing past relevant knowledge and dealing with complex sentence structures are all processed are the same. If you can understand ideas and process information on paper you can do the same on the electronic screen. This might be most prevalent for reading books on a computer or tablet. A novel is to be read sequentially and aside from e-reader remembering where you left off, the experience of reading book is pretty much identical whether you read a book or an e-book. It requires knowledge of symbolic representation, vocabulary, grammar, and sufficient prior knowledge to comprehend new ideas. Instruction and training of how to read focused on these aspects.

Rudimental ICT skills

PIAAC being designed for assessing literacy, numeracy and problem solving skills of adult population, it was imperative to accommodate different level of ICT skills present among adult population. Field study included rudimentary ICT skills performance tasks as well as self-reported experiences on ICT tasks, such as:

- Have you ever used a mouse?
- Have you ever used a mouse to point and click on word or objects on the computer screen?
- Have you ever used a scroll bar?
- Have you ever used a mouse to drag an object from one area of the screen to another?
- Have you ever used a mouse to highlight information on the screen?

When these ICT performance tasks and self reported ICT experiences are compared against skill measures.

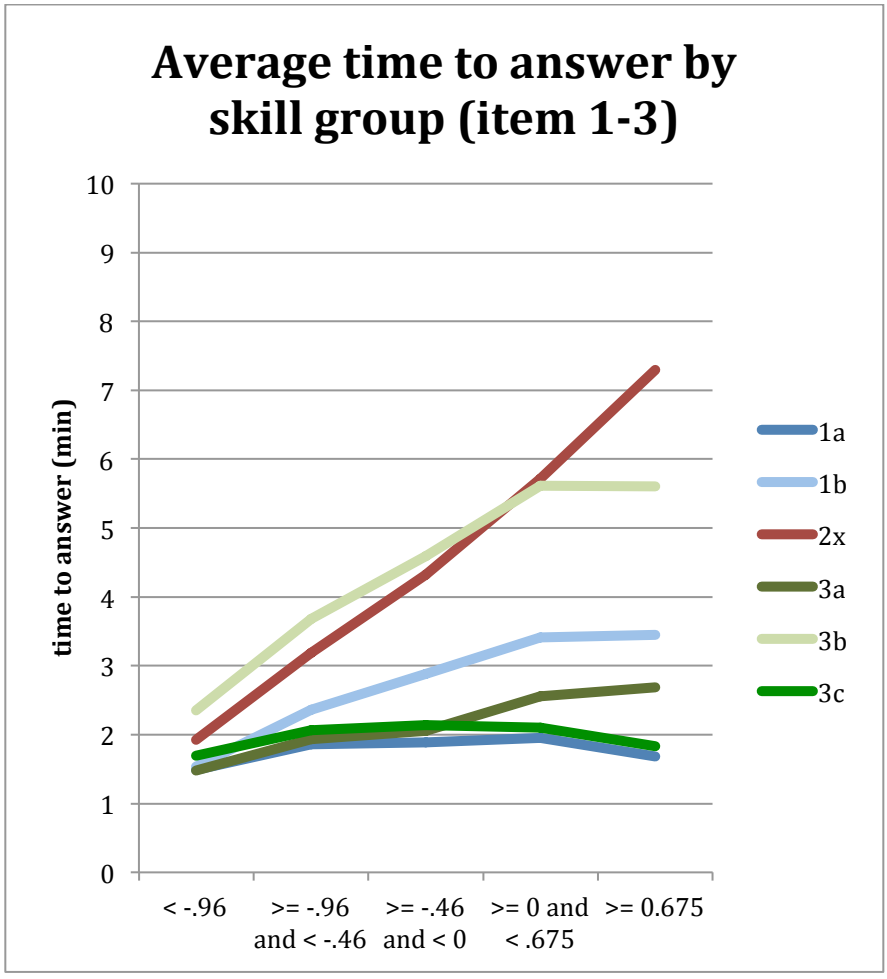
- 12.5% of adult population do not have ICT skills
- All who say do not used computer do not have ICT skills
- 25% of those who say use computer do not have ICT skills
- 50% of those who do not have ICT skills say use computer somewhat
- 95% of those who have ICT skills do use computer
- Mean skill measure of those who do not have ICT skills is .7 sd lower.

New item types used for Problem Solving TRE

The expert group for the PIAAC Problem Solving in Technology-Rich Environments developed a framework saying " goal-directed activities involving the abilities to locate, organize, evaluate, and communicate information through computers and computer networks for personal, professional and civic purposes." The group further defined the environments to be Web sites, emails & discussion boards, and spreadsheets. Cognitive dimensions being measured included; goal setting and progress monitoring, planning and self organizing, acquiring and evaluating information, and making use of information. Responses on altogether 23 PS items were examined in terms of accuracy/correctness and efficiency in terms of number of actions to accomplish task as well as speed.

It is our common experience that a person with higher skills answer a question more accurately and also more quickly. It has been true for out surveys on paper format. These relationships did not hold on the majority of Problem solving items on regard to the

speed. On average respondents with highest skills spent nearly 50% more time to answer questions and they produced more correct responses. Respondents with less skills spent roughly same amount of time on easy items as well as on difficult items, between 1-3 minutes per item regardless of difficulty of items. The respondents with higher skills spent between 1-9 minutes to answer an item, more time on difficult items but much shorter time on the easy items. Average time to answer an item for the first 6 items is presented on the following figure. Respondents were categorized into 5 ordered groups by standardized skill measures representing, 17.3%, 17.3%, 17.3%, 25%, and 25%.



The relationships presented above were quite robust across countries in general, although some countries tended to respond more quickly than others. First impression of suspecting the respondents may not be attempting PS items is not correct. They spend

ample time to study items yet not able to answer items. Also PIAAC being a household survey and each respondent answering questions on screen with an interviewer present in close proximity makes patterned non-response rare occurrence in compared to a group administration in a school setting. We are not quite certain either they are attempting many solution strategies and still cannot find the right strategies, or not even able to start difficult items. We need to wait for the current main survey results to answer such question. The more complex cognitive tasks within ICT context do require not the specific knowledge about the task but rather more generalized architecture of information one is encountering.

Knowledge transfer and acquisition have been through restricted channels such as schools, programs, facilities, and regions mostly these are physical contexts. The ICT environment allows anyone who is interested in learning to access information nearly freely regardless of school they may attend. Limitation of books, school buildings, classroom structures, learning style, scheduling become less restrictive than before.

There are many obvious benefits for using ICT as an aid to instruction to list a few:

- Presentation can be adjusted for the environment, thus enhancing clarity of information in terms of sight and sound,
- Presentation can include animation and film clip to further describe temporal relationships,
- Magnify hard to see activities and objects,
- Can simulate dangerous activities and infrequent or rare events,
- Allow modification to needs of students,
- Repetitious learning can be performed and monitored without intervention,
- Students can participate more independently, and
- Enable to use the resources on internet.

Carrying the analogy of experiences with PDQ instructional systems, we should teach conditional structures and architecture of information. As seen on the WEB each genre has a structure that is conducive to particular information often hierarchical and linked to aid bird view to detail as reader need for a specific purpose.