



INFORMATION PROCESSING

TUTORS

These notes are designed to assist teachers of the course and are in a condensed format. Teachers should also consult the syllabus for this module and adapt these notes accordingly by using extra examples and by filling out this material with detail. Students will be expected to apply the material in a BUSINESS environment and with reference to the particular situation specified in the examination question. Repeating these or any other notes in a generalised form is unlikely to satisfy the required answers for any question.

DEFINITIONS

Hardware – the physical parts of the computer. e.g. keyboard, printer, memory boards

Program – a set of instructions to achieve ONE main task. e.g. analyse the customer file and display all customers who have not paid within a given time-scale.

Package – a set of programs designed to perform ALL aspects of a MAJOR business application. e.g. to deal with customer orders, purchase orders, stock control etc. The package probably would include features for automatic installation, user customisation, built-in help, manuals (on or off line) etc.

Software – a general name for a program or package. The distinguish between:

Application Software – software to solve a business application.

System Software – software to assist the computer in organising itself.

See notes in GENERAL PURPOSE SOFTWARE for more details and also below.

Data – The raw facts input into a computer. It can also refer to data held on computer file for later processing. See **Information**.

Information – strictly this refers to data that has been processed and therefore has an associated meaning. Computer output should be information.

Parameter – an item of data which shows HOW a process should be performed. e.g. an indicator to show a sort program which is the key field for sorting.

DATA CAPTURE

Candidates should be aware of

- *a range of data capture methods*
- *the advantages and disadvantages of particular methods of data capture.*
- *business needs in choosing a means of data capture.*

Data capture should be as automatic as possible with as little user intervention as possible.

The reasons for this are:

- humans make mistakes
- humans cost money because they must be paid
- data is captured at source and there is therefore less chance of it being lost in its passage to the computer

WARNING

Previous examination answers have been vague in this area. “Used in banks” (or supermarkets) will NEVER gain marks. Banks use a wide range of software and applications and the candidate must be specific. In fact, many ATM machines are actually outside of bank buildings!

KEYBOARD

Possesses the standard typewriter layout of letters/digits and punctuation. In addition it has function keys which are used by particular software for pre-arranged purposes. e.g. F1 is often used for HELP features. A separate numeric keypad is useful for pure numeric entry such as in retail data entry. Additional keys are available for Internet access. Candidates should NOT quote particular function keys in describing use of software because these keys may only apply to that particular package.

Uses – Text data entry where there is little structure such as names and addresses.

Corrections to data.

Entries to general purpose software programs such as word processing.

Advantages – Ideal where data is unpredictable or of a variable shape.

Limitations – Limited to the speed and accuracy of the keyboard operator.

Expensive because it is user intensive.

TOUCH SCREEN

This must not be confused with VDU which is an output device. The finger breaks vertical and horizontal cross beams just in front of the screen to pinpoint the position on the screen. It is suitable for data input where the data is limited in range and effectively represents a series of options. Some areas of the screen can allow a move to a submenu of options.

Uses – Tourist areas to pinpoint the locations of particular places of interest that could be buildings or items within a building.

Restaurants to select from a limited menu to transmit the order directly to the kitchen and to generate customer billing.

Choices for a bank customer at an ATM.

Advantages – provides information with little user action (or skill).

Cuts out the need for human information providers.

Limitations – only suitable where the data is a small FIXED range of items. The width of a human finger limits the number of items that can appear on the screen. If sub menus are used, this further limits the number of options on the main screen.

MICR (Magnetic Character Recognition)

Banks originally developed this data capture method with a view to processing huge numbers of cheques received AND to provide some security against possible fraud. If another business decided it wanted to use MICR for data input, it would find it difficult to obtain a supplier of the hardware because banks have the vested interest. The only hardware available would read cheque-format data. This method is very effective for a quite specific purpose.

Uses – Cheque reading ONLY.

Advantages – Fast

Cheques can still be read even if they have minor damage

Only human action is in feeding the reader.

Limitations – Reads only one line of data and a small character set.

Requires line-up characters for the hardware to location the position of data. Line of data is always in the same position on ALL cheques.

Stylised characters only can be read.

Requires additional software to convert the input signals to meaningful information.

OCR (Optical Character Recognition)

Standard shaped characters (typically output by printers) are read by optical scanning and interpreting the returned signals from patterns held on file. More sophisticated hardware will now read hand-written data provided it is well formed or in strict positions.

Uses – Utility bills (electricity, gas, water) sent to customers can be re-input to signal that the bill has been paid in full because it holds the customer number and amount due (turn around document).

Advantages – Eliminates the need for human keying of data

Limitations – only suitable for small quantities of data.

Requires line-up characters for the hardware to location the position of data.

Keyboard entry is still needed, in the case of utility bill, if the payment is for an amount other than that stated on the form or if the paper is folded and the data cannot be read.

Requires optical character recognition software to convert the input signals to meaningful information.

This is prone to errors if the input is hand-written.

BAR CODE READING

Bar codes are read by optical scanning and the returned signals are converted into binary. They are used purely for identification.

Uses – Supermarket checkouts to identify the product bought.

Libraries (2 uses) – to read the book identifiers and also the borrower's ticket code.

Hotel – to identify guest and enable internal charging to an account for all services used.
Ticket entry to events with pre-purchased tickets.

Forms issued (and returned completed) to identify the recipient.

Advantages – Eliminates the need for pricing every item in a supermarket because the manufacturer prints the item code on the package in the form of the bar code.

Less staff needed in a supermarket.

Enables the price of a product to be changed by a single entry on the main file.

Cuts down queues at supermarket checkout and less checkout errors.

Now enabling self-service in supermarkets with one supervisor to oversee several checkouts to resolve problems.

Better customer receipts at supermarkets and automatic stock updating.

Bar codes can be read at any angle.

Bar codes are also printed with numeric code for human interpretation.

Limitations – some codes will not read properly and need to be keyed in.

Requires line-up marks (centre of bar code) for the hardware to location the position of data.

NOTE: Bar codes do NOT contain item price on a product. The reasons for this are:

- Products may be made in other countries where the currency is different.
- Prices can be set by the seller and varied for "special offers".

OMR (Optical Mark Recognition)

Pencil/ink marks are placed at specific positions on paper and detected by the machine by optical scanning and testing the reflected signal. The position of the mark identifies its value and this is then converted into a binary value.

Uses – Multiple-choice examination input. The candidate places a mark to indicate his/her chosen answer.

Lottery tickets. The purchaser marks the chosen numbers.

Stock auditing. A warehouse clerk marks the number of items actually on the shelves.

Meter reading in private houses/businesses. The current value on each meter is recorded.

Street surveys. To record the views of people outside of buildings.

Advantages – The data is recorded in computer readable format at the place where the data occurs.

Limitations – Small Numeric data only.

Requires line-up characters (side of page) for the hardware to location the position of data.

Difficult to correct errors at the point of capture resulting in the need for keyboard correction/entry on faulty documents.

MOUSE / DIGITISER

A mouse CAN be used to enter data but the context needs to be explained. Where DATA appears in a question, the use of a mouse to selecting options is unlikely to earn any marks unless the options are clearly stated and refer to data and not just option selection.

Uses – Computer Aided Design/Drawing. The mouse can be used to select a standard object (circle) but its main function is in positioning the object and adjusting its size by dragging handles. Objects can then be turned, inverted, coloured under mouse control (this is the DATA element).

Advantages – Any other form of input device does not have the controls for drawing.

Limitations – Needs complex software to enable accurate drawings to be made.

VOICE RECOGNITION

Human voice signals are converted to waveforms and binary signals are then compared with standard patterns held on file.

Uses – Control of machinery remotely or in “dirty”/busy situations such as a machine room/warehouse.

Direct input to a text editor.

For handicapped people unable to use other means of input of data.

Advantages – Hands free.

Limitations – Needs complex software to enable accurate identification of words.

Small vocabulary and limited by regional accents.

Needs to be calibrated for particular users.

Subject to errors when voice changes (colds).

SCANNER

Paper documents are optically scanned and converted to electrical and then bit-map binary signals. Some scanners give higher resolution than others (dpi).

Uses – Archiving of documents – legal requirements.

Advantages – Documents can be entered into computer readable form without the need for any user input. These can then be edited.

Limitations –Resolution can be variable.

The documents take up a large storage capacity.

CHOOSING THE RIGHT CAPTURE METHOD

Points needed to be considered when choosing a method of data capture for a given application are:

- Human involvement in the input process – normally required to be minimum for cost purposes.
- Cost of data capture.
- Accuracy of the received data.
- A means of correction or alternative means of input when the chosen method fails.
- Can the data be captured at source – in the class room, on the street etc ?
- Volume of the data. Large data entry may need a different method from small data entry.
- Comparison of alternative methods of data capture.



DATA CHECKING

Validation – the process of checking data AS IT IS INPUT without reference to a file, to establish whether it is reasonable. Validation does not guarantee that the data entered is correct.

Verification – the process of double entry of data where the two versions are compared. This is now rarely used although still common when entering a new password to ensure the user has not made a typing error.

In both cases, if an error is detected, it is reported and the data rejected. If the data is entered online, the user is requested to re-input.

VALIDATION

Examination questions about validation rarely earn many marks because answers given are usually too general and do not relate to the specific data quoted in the question. Where data is given, the answer **MUST** indicate how the data can be validated – “range” is not enough.

Means of validation (with examples) :

Range – In a date, the day of the month must relate to the month in question so must be in the range 1 to 30 for June. Month must be in the range 1 to 12 if numeric.

Format – A code may consist of 2 letters followed by 4 digits. The answer must mention this.

Data Type – Quantity of goods purchased must be numeric. In some cases, it must be an integer.

Length – An identification code may have to be exactly 7 characters.

Presence – While some fields in data entry may be optional, some must be present. e.g. product code when purchasing and item.

Connected fields – The value of one field may determine the range of values in another. Clearly the order of two related fields is important.

Check digits – used only for REFERENCE numbers. e.g. customer code, product code, student identifier etc. It could not be used for other numeric data (e.g. number of items purchased) because who or how would the check digit be derived?

CHECK DIGITS

A reference number is extended by one digit (normally at the right hand end) to include a check digit.

A check digit system guarantees detection of:

1. transposed digits - e.g. 12345 input as 12435
2. the wrong repeated digit - e.g. 12234 input as 12334
3. any single digit error - e.g. 12345 input as 12845

There are two aspects of check-digit systems. Candidates **OFTEN** explain the first aspect when the question is asking about the second.

1. Determining the check digit to be added to the end of a reference number to make it computer checkable.
e.g. a 5 digit customer number is converted to a 6 digit number with its check digit. Each of the digits has a different **WEIGHT**. This normally runs from 1 upwards with the most significant digit using the highest weight and the check digit using 1.
e.g. Convert 13478 to 13478c where c is the check digit to be calculated.
Method: Multiply each digit by its weight and total the result.

| | | | | | | |
|----------|--------|------|------|------|-----|----------------------|
| | 1 | 3 | 4 | 7 | 8 | c |
| Weight | 6 | 5 | 4 | 3 | 2 | 1 |
| Products | 6 + 15 | + 16 | + 21 | + 16 | + c | = 74 + c = CHECK SUM |

Make the check sum up to a multiple of 11 (the **MODULO**). So 74 + 3 = 77. Hence c = 3
The full reference number is 134783

The modulo is often chosen as 11 because it is a prime number and in the multiplication process, it will not be possible for one (and usually 2) errors to cancel each other.

2. Checking a reference number entered during validation to determine its validity.
Repeat the above process except now the check digit is known. Suppose above 135783 is entered by mistake.

| | | | | | | |
|----------|--------|------|------|------|-----|------------------|
| | 1 | 3 | 5 | 7 | 8 | 3 |
| Weight | 6 | 5 | 4 | 3 | 2 | 1 |
| Products | 6 + 15 | + 20 | + 21 | + 16 | + 3 | = 81 = CHECK SUM |

The check MUST be a multiple of 11 to be valid. 81 is not a multiple of 11 so this number is INVALID.

There is one problem with stage 1. Suppose the check sum is $78 + c$. This would require the check digit to be 10 to make it up to 88. There are two possible courses of action:

1. Do not use this number and reallocate a new reference number for this customer/product/student.
2. Use X to represent TEN as the check digit. This method is adopted for 10-digit ISBN (International Standard Book Numbers) which gives each published book throughout the world, a unique number. The advantage of this is that no new number had to be allocated. However, it does mean the ISBN cannot be classified as numeric.

SPECIFYING VALIDATION FOR A PARTICULAR SITUATION

This will be required in the examination. An example is given below.

“An employee can select his /her holiday dates by filling in a holiday notification form (up to 1 year in advance) and giving the start and end dates of the holiday. No employee may take more than 2 weeks holiday in succession. The dates are recorded as day, month (1-12) and year (4 digits).”

Candidates often describe validation on OTHER possible fields that might be used in the application even though they are not mentioned in the question. These points earn NO marks.

Possible validation of these dates:

1. All six data items must be numeric (integer).
2. End date must NOT be before start date. Can be the same if 1 day is requested.
3. The difference between the dates cannot exceed 14 days.(*)
4. Day (start and end) must be consistent with the month. e.g. not more than 30 for June (month 6).
5. Months must each be in the range 1 to 12.(*)
6. Dates must be in the immediate future.
7. Dates must not be later than 1 year from today's date.(*)

The three items marked (*) are specific to this particular question. Candidates should note this.

DATA OUTPUT

Data can be output using a:

1. **VDU** – suitable where it is need only temporarily (perhaps for checking data from a file).
2. **Printer** – as a permanent visible record.
3. **Plotter** - for large complex diagrams/charts/drawings.
4. **Film/Microfiche** – photographic storage for large quantities of data..
5. **Sound** – for instructions to people in unusual situations. Voice simulation can sound artificial.

PRINTERS

The range of printers over the years has increased but must printers now used are from the following list:

- **Dot-Matrix printer** – Although of 1980's design, improved versions are still used. Print heads contain a grid of 24 pins. One character at a time is printed by hammering a selection of pins against a print ribbon onto the paper. The quality is poor but often adequate for internal memoranda in a company or for transaction slips. Pre-carboned paper enables multiple copies to be produced with one printing. Crude graphic characters are possible with different combinations of the 24 pins. Different models allow continuous stationary or single sheets to be used. The speed is measured in characters per second.
- **Laser printer** - Light is modulated by the computer and directed onto a drum by rotating mirror. The chemical coating on the drum is electrostatically charged. Where light shines on the drum, the charge is removed. Toner powder is attracted to the charged parts and this is transferred to paper during the rotation. Heating the paper helps to fix the printing. The quality is very high allowing graphics to be printed. The speed is measured in pages per minute. The cost depends on the quality printout. It is commonly attached to a LAN to service all work stations.

- **Ink-jet printer** – Ink is sprayed onto the paper through tiny nozzles by either heating for a short time or by pressure. It is slower than the laser with typical speed of 4 pages per minute. However, it is able to print graphics and in colour making it superior to the dot matrix. It is one of the cheapest printers now around and is often attached to stand-alone computers in the office or at home.
- **Line printer** – Two types are still available having been the printers of the mainframe as far back as the 1960's. The **barrel/drum** printer has a line of each character embossed on the drum so that all occurrences of the same character in a line of print are produced at the same time. The drum then rotates to print all occurrences of the next character. The hammers hit the paper and carbon roll onto the drum. The speed of the drum is so fast that one revolution of the drum produces one line of print. 2000 lines per minute are achievable. The **chain** printer has a chain revolving horizontally and carries several copies of all the characters. A line of print is achieved by hammering each character of the line onto a ribbon as one of the appropriate characters on the chain is in position. These types of printers are still used for large print runs such as printing many thousands of utility bills in one run. They are now being replaced by large scale laser printers

DISPLAY OR PRINT?

1. The screen gives a temporary output and is useful when a printed record is not needed. The limited size of the screen means that is limited for viewing long or wide documents where scrolling would need to be used continuously. The screen is inappropriate when it is necessary to be checking back and forwards through a lengthy document.
2. Printed output gives a permanent record with the option of multiple copies. The output may be needed elsewhere by people not producing this output. However, it does waste much paper. Despite the move towards a “paper-less society”, printouts are frequently produced which are unnecessary.

PLOTTERS

They are usually on a larger scale than printers for fine drawings. Lines can be drawn at any angle. Text is actually drawn rather than printed. There are basically two type:

Flatbed – an arm holds a pen and different colours are possible using different pens. The arm moves along the paper to give an X movement. The pen moves along the arm to give a Y movement. A combination of these enables inclined lines are shapes such as circles to be drawn.

Drum – the paper is wrapped around two rollers that rotate to feed the paper in the X direction. The pen moves across a bridge to produce the Y movement.

Uses – architectural drawings, circuit design, car design. Plotters can be very large – to plot the lines of a car FULL SIZE.

MICROFORM (Microfiche, Microfilm)

The output is effectively photographed and transferred to a film in the form of a sheet or roll. A projection device is needed to view the microform. Microfiche typically stores hundreds of “sheets” on one flat film card with each sheet being a rectangular frame. The first frames show the index of the contents of the whole card. The microfiche reader is a back projection screen to magnify the tiny images to a readable size. The user moves the card around physically to select the required frame. It is particularly useful for viewing large data files where conditions do not suit the use of a computer – dirty workshop.

Uses - archiving data, catalogues of parts in a garage where the data changes only infrequently.

THE COMPUTER PROCESSOR BOX

DEFINITIONS

Tower Main tall box holding the Processor unit and various storage drives. Its shape is designed to minimise the space taken up on a desk

Binary Unlike the decimal system which uses the digits 0 to 9, binary is a system which uses only 0 and 1 but uses the same basic rules of number systems. See Notes for Number and Logic. It means that lengthy binary numbers can represent

| | |
|--------------|--|
| | relatively small decimal numbers. The reason for the use of binary is that electrical/magnetic properties are usually two-state (on/off) and devices using them can be small, cheap and very fast. |
| Byte | 8 binary digits (BITS). |
| Mb | Megabyte. Size of memory or hard disc in millions of bytes. |
| ROM | Non volatile main/primary memory used for start-up. Contents are systems programs and parameters relating to running the computer system. Memory is retained when the power is switched off. |
| RAM | Volatile main/primary memory used to store user programs and data. It is for temporary storage and the contents will be lost when the power is switched off. |
| EPROM | Erasable Programmable ROM memory which can be reprogrammed with an EPROM blower using a ultra-violet light source to erase previous contents. Used mainly by software developers or institutions that have highly specialised needs and need a fast running program. EPROMs are susceptible to erasure by ordinary light and therefore need to be shielded from light. |
| PROM | Programmable ROM which can be programmed once only. Also used by developers. |
| SRAM | Static RAM for temporary user data programs which retains its data while the computer is on. It is faster than DRAM but more expensive to produce. |
| DRAM | Dynamic RAM is more densely packed onto a chip and is therefore cheaper but is slower to access. It uses less power than SRAM. It needs to be refreshed regularly otherwise its contents are lost. |
| Cache | Fast memory to store regularly used routines or data thus anticipating future calls to memory -some delays could otherwise occur accessing data from dynamic RAM. |

THE PROCESSOR

This is the “engine room” which provides the computer’s power. Its speed of operation is dependent upon the **clock speed**. All operations are synchronised with the system clock which acts rather like the beating of a drum with actions taking place on the beat. The **control unit**, synchronised with the clock, ensures that data and instructions move between the correct parts of the computer. It sends signals to component parts of the computer to activate them after deciphering program instructions brought in from memory. e.g. peripheral devices to receive or send data or ALU to perform an arithmetic operation.

MEMORY

Programs are loaded into memory in order to be run under the control of the control unit. Some areas of memory will be used to hold data as named **program variables**. Other areas will be used as **buffers** to hold blocks of data read in from a peripheral or secondary storage area. Buffers are also needed for output devices. The screen uses **video RAM**. The contents of memory are brought into the processor along **buses** under the control unit. (see Fetch-Execute cycle later).

ARITHMETIC LOGIC UNIT (ALU)

This performs basic calculations under the control of control unit. Arithmetic operations are performed using the **accumulator** as a temporary working area.

e.g. ADD P might be the instruction “Add the contents of P to the accumulator”.

Similarly, all decision instructions are processed here.

e.g. BNZ P might be the instruction “Branch to label P in the program if the accumulator holds a value which is non-zero. The ALU will set a 1-bit flag to 1 if conditions are met or 0 if not. In this case, when control detects a 1, it will change the program counter to the address of P so that the next instruction obeyed as at location P.

BUSES

Address Bus carries a memory address to and from memory.

Data Bus carries data to and from memory or an instruction from memory to the control unit. Both these buses are PARALLEL devices so that many bits can be transferred at the same time. Serial transmission would be too slow.

REGISTERS

Key single storage locations are used for particular purpose during the processing of instructions. Their precise uses are described in Fetch-Execute cycle section.

Accumulator – the temporary working area for calculations and logic operations

Program counter (PC) – this holds the address in memory of the next instruction. Normally this register is incremented (1 added) during the processing of an instruction to show that the next instruction will be found in the next physical location in memory. However, after a decision (see ALU above), the next logical instruction may be in a different part of memory and the program counter will adjusted to the appropriate address.

Current Instruction Register (CIR) – This holds the latest instruction brought from memory on the data bus which the control unit will decipher and then signal other parts of the computer to take action.

Memory Address Register (MAR) – This holds the address in memory for the next item to be fetched. It could refer to an instruction (from the PC) or a data location value (from or to memory).

Memory Buffer Register (MBR) – An instruction collected from memory is stored in this register before being moved to the CIR. Similarly, if an instruction requests access to a data value or it is to be returned to memory, the data is placed here. Some books refer to the MBR as the Memory Data Register.

FETCH-EXECUTE CYCLE

The execution of a single instruction in a program normally passes through 3 phases:

1. **Fetch** phase when the instruction is obtained from memory.
 - a. PC holds the address of the instruction
 - b. Address in PC is passed to MAR
 - c. Control places this value on address bus
 - d. Instruction at the memory address given in the address bus is placed in MBR
 - e. This instruction is transferred back to the processor along the data bus.
 - f. The contents of the data bus is placed in the CIR
2. **Decode** phase when the binary instruction is broken down ready for execution.
 - a. The control unit breaks up the instruction in the CIR into operation code and address parts (in the simplest instruction formats)
 - b. Control decodes the operation code. It could be arithmetic (e.g. ADD - see above), decision (e.g. BNZ – see above) or instructions to send or receive data from peripherals/storage.
3. **Execution** phase when the instruction is obeyed.
 - a. Appropriate signals are sent to other parts of the system.
e.g. ADD and BNZ would require a signal to the ALU indicating the type of action to take
A read/write instruction would mean a signal to the peripheral or storage device.
 - b. Arithmetic instructions will require a memory read. e.g. ADD P. The address P is placed in the MAR and sent down the address bus. The contents of P are placed on the data bus and returned to the ALU where the value will be added to the accumulator.
 - c. A store instruction (Store contents of accumulator to memory) will require the storage address to be placed in the MAR and then transmitted along the address bus. The value of the accumulator will be placed in the MBR and transferred down the data bus where it will then be placed in memory.
 - d. A decision instruction will require the appropriate flag to be checked. If it is found to be 1, the address in the CIR will then be transferred to the PC. (see ALU above)

FLAG REGISTER (or STATUS REGISTER)

Single bits are set to 1 if a particular event has occurred.

Sign flag – set to 1 if a test for negative has occurred, 0 if positive

Zero flag – set to 1 if the last arithmetic operation produced a zero

Carry flag – set to 1 if the last arithmetic operation resulted in a carry

Overflow flag – set to 1 if the last arithmetic operation overflowed the available memory

Other bits are used to show different types of interrupts, some would have priority over others.

CONFIGURATION DIAGRAM

From the above information, it is possible to draw a configuration diagram showing how the various parts of the computer are linked. Distinction needs to be made between data movements and signals sent to components of the computer to activate them. Note, in this sense, an instruction is data to the processor control circuits.

Candidates must mark the distinction clearly on diagrams and show precisely between which parts the link occurs and the direction of flow. Some links are in BOTH directions but not all (e.g. data from a keyboard). A number of incorrect diagrams submitted by candidates have shown a direct link between input and output peripherals without any intermediate processing!

SOFTWARE

Clear distinctions are needed between the broad types.

SYSTEM SOFTWARE

Operating System – This is the management software that oversees all operations in the computer. The main tasks it performs are:

- “Boot up” the computer when it is switched on. The term means to pull in the routines it needs from disc and to run start routines. Modern operating systems now save the settings of the computer when it was last used and in booting, the computer is returned to this same state (icons on the desktop screen etc).
- Load programs from disc into memory and run them ensuring there is no violation when several programs are in memory at the same time. Handling multitasking.
- Manage the use of peripherals and transfer data between peripherals and memory.
- Interpreting user keyboard commands and taking appropriate action.
- Deciding the priority of events when two coming together might conflict.
- Handling security of all elements within the computer and taking appropriate action when a violation occurs.
- Networking where computers are connected to a LAN.
- Logging the processes that have been run.

Utility Programs - Provide utility programs for common organisational use. This includes:

- Disc management routines - creating folders, copying/deleting/moving/renaming files
- Sort programs.

Language translators – these refer to computer languages (not French \leftrightarrow English)

- **Compilers** for HIGH level languages – they convert programmer’s source code in one stage to produce a runnable program and a machine code version which can be saved.
- **Interpreters** for HIGH level languages - convert programmer’s source code one line at a time and then executing that instruction before moving to the next. No runnable program is produced and an error in a particular instruction will only show when that instruction is translated – if it is not translated in a given run, the error will not be detected. Interpreters are used for development of new programs.
- **Assemblers** for LOW level languages (machine code languages). Low level languages use coded instructions similar to binary machine instructions. This means that although writing in these languages there are restrictions (slow and therefore expensive to generate, restricted to particular machines), there are some great advantages (code has less overheads and so much smaller final programs are produced which are fast to run). In certain circumstances, these advantages outweigh the restrictions such as in machine or industrial process control

APPLICATION SOFTWARE

The distinction between General Purpose and Special Purpose software needs to be stressed. Candidate answers suggest Special purpose software is produced “to perform a specific task” – ALL programs are produced for this reason! A “specific BUSINESS task” should earn the mark.

General Purpose Software – software produced of a general nature where the producer does not govern HOW the user actually uses it. The user decides the uses. Examples;

- **Word processing package** – This is a substantial text editing package enabling the user to create textual documents and choosing the editing features (underline, centring, margins, fonts etc). Sophisticated features include mail merge, some drawing/graphics and the inclusion of items created in other packages.
- **Spreadsheet package** – This package provides a grid of cells. Each cell can hold a variety of types of data and relationships (formula) can be generated to produce values in one cell calculated from values in other cells. A large selection of prewritten functions means the user can draw on the expertise of others.
- **Database package** – This package enables related data to be structured into records (e.g. customer details, student details). These records can then be amended, deleted and displayed. Searches enable records that satisfy the user conditions to be displayed. Elaborate optimisation of the way data is stored enables databases to be accessed very quickly. Teachers should emphasise that the purpose of a data base is to produce OUTPUT, not to accept input.
- **Graphics/Drawing package** – This package provides standard shapes (circles, rectangles, lines) to be positioned, rotated, stretched and coloured. Combining a number of these components builds up a picture/diagram which could be used in the design industry.
- **Presentation package** – This is a fashionable package that enables the user to produce a multi-media slide show with colour, animation and sound for showing to clients. This could be used to promote a new product, idea or procedure.

There are other general purpose packages but only these will be tested in examinations. None of these packages “do” anything when started. It is the user who dictates what is produced.

Special Purpose Software – software to administer/arrange/run a particular business/industrial process.

Once the software is started, the user is guided to select the appropriate options and enter the required data.

Most application software falls into this category. The APPLICATIONS module considers some in detail.

Three examples are given briefly

- **Retail Administration** – This deals with selling products to customers. Details of the order are entered and an invoice/receipt/packing note/picking list could be generated. The order details would be stored on a disc file and accessible when the customer pays at a later date. Associated with retail would be the administration purchasing goods from suppliers in bulk. This is effectively the same as the retail but in reverse where the company is now the customer. Stock control enables efficient ordering and minimum expenditure of goods awaiting purchase.
- **Medical Records** – A patient registers, his/her details are held on file including the various treatments offered. Later consultations by that patient with different doctors enable the previous medical history to be accessed quickly. Drug information could be held for the doctor to read before making a decision on treatment. Stock control of drugs could be included.
- **Estate Agency** – A person wanting to sell a house could register with the agent, the details stored on file. When a prospective buyer arrives, all the properties meeting his/her requirements could be accessed.

In all of these, once the program is started, the user may be faced with a menu but each menu option would perform a specific task. Each transaction would be saved automatically after it is completed, unlike with general purpose software where the user decides when this happens.

FILING SYSTEMS

DEFINITIONS

File = An organised collection of RELATED records. e.g. ALL the records of students in a college.

Record = ALL the details about a particular person or thing. e.g. details of one student.

Field = One indivisible item within a record. e.g. Date of Birth

Organisation = The way in which records are held on the file. This will dictate HOW the data can be accessed.

Access = The method of retrieving a record from file. This is limited by the method of organisation.

In trying to locate a record on file, the record can either be found and read into memory or a positive statement that the record is not present is output – because the user keys in the wrong access details or perhaps another clerk has not yet entered the relevant data.

Key field = The field on a file normally used to order that file. Student number might be used to ensure the records on the student file are in student number order.

Block = Records on all media are grouped in blocks (sometimes called buckets). A single record is NOT read from the file into memory but the whole block is read. This speeds up file access because the slowest part of disc reading is locating where a record is. By reading a block, the next record is also in memory and this is often the next needed.

Hit-rate = This is the proportion of records needing changing compared with the total number of records on the file. If this is high, serial or sequential organisation and access could be used. If low, then perhaps indexed-sequential or random organisation is needed.

Master file = This is a main/semi-permanent file holding all the details about a particular business area. e.g. All employee records would be held on an employee file and separate from the customer records on the customer file. Much of the data stays the same for a long time. Changes might be added to the master file from the transaction file.

Transaction file = This is a temporary file used to collect changes until they are updated onto the master file. This type of file is only used in batch processing. The file would first have to be sorted so that a new file is produced which is in the same.

Batch Processing Transaction data is collected as above over a period of time and then perhaps once per week/month, the master file is updated. Apart from keeping the file for a short while for security purposes, the data is no longer needed. (See below)

File Update

By Copying – All changes to the master file are achieved by creating a new file, changing those records that need changing but copying the others across, unchanged (Serial/Sequential organisation)

By Overlaying – Changes are implemented immediately by overwriting existing records (Indexed-Sequential/Random organisation)

Filing cabinet – This is included because of the confusion candidates have between manual and computer filing systems. Teachers are advised NOT to use the analogy. Candidates need to know the computer definitions of file and record. In a filing cabinet, a paper folder might be labelled as belonging to a particular employee but inside it would be a random selection of papers holding data (and odd notes) with little structure. Two employees might have the same types of documents but they could be on different paper (one a form and the other a hand-written note) and the order does not matter.

Each of these items on paper would, IN THE COMPUTER, be RECORDS and could even be in different files – weekly payroll data would not necessarily be with employee name and address but would be linked by employee number. Each employee would NOT have his/her own file but a series of records in different files.

WHY FILES?

Data cannot be retained in main memory of a computer because RAM is only temporary and too small. Using ROM would prevent the data being changed. Hence, secondary storage systems were developed to store data not currently needed in the computer. If a computer is currently running a retail system, there would be no need for employee details to be available until a payroll run was required. Data on file has to be organised.

FILE ORGANISATION

SERIAL (SER)

Records stored one after another in the order that they are entered into the computer. There are no gaps between the records. The only order is therefore chronological - Customers do not conveniently place orders in customer number order nor do they order in product number order.

Reading the file – Records can only be read in the order they are on file starting with the first. To find a particular record, it could be necessary to read the whole file if the required record is the last one.

Adding a new record to/Amending a record/Deleting a record from an existing file – Generally these processes require a new file to be created which is an exact copy of the original but with the one change. Records can be added to the end of a disc file but with tape, a new file is created.

Uses – Transaction data as it is received (processed LATER) – it may have to be sorted before processing.

SMALL reference files which can be read quickly from the beginning every time it is accessed. e.g. VAT rates, payment levels for staff.

Media – ALL types of storage media permit Serial file processing.

SEQUENTIAL (SEQ)

Records are stored one after another with no gaps and in KEY FIELD order.

Reading the file – The key field value is entered into the computer. The file is read from the beginning until the required record is located or until either a key-field value on file is higher than the required one OR end of file is reached. In either case, the record is not present and should be reported.

Adding a new record to/Amending a record/Deleting a record from an existing file – Generally these processes require a new file to be created which is an exact copy of the original but with the one change and the record placed in its correct key field order.

Uses – Transaction data as it is received (processed LATER) – it may have to be sorted before processing.

SMALL master files which can be read quickly from the beginning every time it is accessed.

Uses – Reference/Master files where instant access is not needed and batch processing is suitable.

Media – ALL types of storage media permit Sequential file processing.

INDEXED SEQUENTIAL (IS)

This organisation stores records effectively in sequential order but in blocks with a range of key values associated with each block. There are therefore some physical gaps in the file if records for some key values are not present. A small index file is associated with the main file showing the HIGHEST key field value permitted in that block. To reduce the size of the file so that there are not large physical gaps, the key values allocated to each block may mean that IF all records were present, they would not fit in. In practice, this is uncommon rather than usual. To overcome this rare situation, an overflow area is allocated. Where the record SHOULD have been is placed a small tag indicating the block number where the record actually is. Such a record would therefore need TWO reads instead of the normal one. Candidates often write that EVERY record is named in the index file – this is not so. It is NOT like a book index.

Reading the whole file – The file is treated as sequential and follows the same process as above.

Reading the one record - The key field value is entered into the computer. The small INDEX file is read sequentially until a key value is found which is equal to or greater than the required record. This then identifies the block where the record is stored. The whole block is read into memory and searched to find the required record. With very large files, there could be several levels of index so that the first indicates the **cylinder** in which the record should appear and the second the **block** within that cylinder. See **hard disk** (below).

Adding a new record to/Amending a record/Deleting a record from an existing file – The single record is located as for reading above, the change is made in memory and then written back to the file overwriting the original. Deleted records can be removed from the block to free up space for other record that should be there including those moved into the overflow area.

Uses – Reference/Master files where instant access is needed but also where the whole file might be needed for certain processes. i.e. find a customer record but also report on the while file for customers who have not paid their bills or are to be targeted for a sales promotion..

This is ideal for online access perhaps in telephone ordering system.

Media – Only disc-based media because of the need to overwrite existing records.

RANDOM (R)

Random organisation does not mean records are stored randomly - implying they could be anywhere on the file. The file “appears” to store records randomly. The key field is used to identify where the record is stored. The process is:

Key field → algorithm → block number

So, a mathematical process is applied to the key and the resulting number that comes out gives the block number. The size of the available storage space usually dictates the formula. If 1000 blocks of disc space are available, a simple algorithm could be DIVIDE key value by 1000 and use the remainder is used as the block address. This remainder would therefore be in the range 0 to 999. It follows that consecutive key values would be located in different blocks and helps to spread the records over the file. An overflow area is again needed where some bunching occurs and this would be totally unpredictable. Tags are used as with Indexed Sequential files.

Reading the whole file – This organisation is not suited to applications where the whole file would need to be read because of the apparent spread of the records because there would be no order in the records accessed.

Reading one record - The key field value is entered into the algorithm, the block address calculated and the record then read as for an indexed sequential block read.

Adding a new record to/Amending a record/Deleting a record from an existing file – The single record is located using the algorithm as above, the change is made in memory and then written back to the file overwriting the original. Deleted records can be removed from the block to free up space for other records that should be there but are located in the overflow area.

Uses – Reference/Master files where instant access is needed but where reading the whole file is NOT appropriate. This is also ideal for online access perhaps in telephone ordering system.

Media – Only disc-based media because of the need to overwrite existing records.

WHICH FILE ORGANISATION IS NEEDED FOR AN APPLICATION?

The designer will make the decision after considering the following questions.

- Will individual records need to be located or is the whole always used? Are both methods needed? How quickly must the data be accessed?
- How big is the data file? SERIAL gives quick very access if the file is small because there is no indexes to consult or algorithms to use.
- Record content and size?
- Is data permanent or temporary?
- Who is the data for?
- How frequent is the data collection?

- Will the data need to be amended/deleted/appended?
- How long will the data kept and what happens to it at the end of that period?

STORAGE DEVICES

Distinguish clearly between:

Storage device = hardware with read/write heads. e.g. disc drive

Storage media = material on which data is held. e.g. diskette

Magnetic Tape – Data is held longitudinally using 9 tracks recorded as magnetic dots on a plastic tape coated with magnetising material. Each transverse frame holds the 8 bits of a byte plus a parity bit. Many frames would make a record and like disc, several records constitute a block. Each block is separated by an **inter-block gap** that is actually there because the tape speeds up and slows down in the reading and writing of the blocks. The tape only runs when it is being read or written to. Tapes have largely been replaced by disc systems

Uses - data logging, transactions, small reference files, back up (tape streamer).

Diskette – Small enclosed circle of magnetic sheet with the data stored on circular tracks. Each track is split into sectors. The outer tracks are used for the **File Allocation Table (FAT)** which lists for each file, its name, type of file, date created and actual location(s) on the disk. A large single file can occupy several different areas when the disk has been heavily used – new files take up the gaps, some created by deleted files. Diskettes are slowly becoming obsolete being replaced by CDs.

Uses - transfer of small personal files, small backup files.

Hard Disc/Winchester – Can be likened to many diskettes placed one above the others. A single arm moves between the platters, with read/write head for each recording surface, above and below. It has a much larger capacity than diskettes. All the tracks accessible from ONE position of the arm constitutes a **cylinder** of data. A start of a file is normally recorded on the tracks of a cylinder because they are vertically above each other and reduce the need for the physical arm movement (mechanical and hence the slowest part of the reading process). Only when these are all used will the heads move into the next cylinder. **Seek time** is the time for the heads to move to the required track – a mechanical process which takes time. The disc is normally divided into named **directories (folders)** to group files relating to particular applications from others. Programs are in a separate directory.

Uses - application programs, random access business files

Optical/Laser – Similar in nature to a diskette with circular tracks but the data is burnt in with a laser. They give very fast access and are cheap.

CD-ROM - Data can only be read. **Use** - supply of software

CD-R - Write once then only read **Use** - archiving

CD-RW - Can read, write and delete as in magnetic discs.

Memory stick – Small semi-conductor memory which plugs into the USB port.

Uses – Similar to CDs but useful in transferring files from one machine to another. Does not require specific drives that CDs need.

PROCESSING METHODS

Interactive – Used in program development where testing produces instant results and the end-user can make adjustments accordingly until the program is correct.

Real-time – This is interactive with feedback. Originally it referred to automatic systems like the computer control of processes where instant feedback might be vital to prevent danger. It has been extended to business applications. Processing one transaction can result in the output being recorded instantly and may then affect subsequent input and processing. e.g. Airline seat reservations from many remote sites. A seat booked in one office is recorded so that the same seat cannot be booked elsewhere in another office, one second

later. The seat is marked as temporarily reserved by the first user. That user will either confirm the booking or release the seat for other users.

Batch - Data is collected into groups called batches and processed at a later time in one session without any operator intervention. This is becoming less frequent now as fast disc systems enable data to be updated at the time of processing. To ensure data is not lost, a **Batch Total** is manually calculated for each batch, input with the batch and then recalculated by the computer during the run to check the total tallies. Any failure and the whole batch is rejected and is resubmitted at the next run. Any numeric field can be totalled for the batch total. However, if the actual total is meaningless (e.g. total of employee numbers), it is normally called a **hash total**. A count of the number of records in a batch is NOT used for the batch total because it is common for each batch to have the SAME number of records.

SECURITY

Distinguish between

Security – ensuring hardware, software and data are not damaged or lost.

Privacy – ensuring data held is used properly and not accessed by unauthorised people.

ENSURING SECURITY

REALISTIC answers are required. Telling people NOT to do something does not prevent it happening. There must be monitoring and safeguards. Answers should avoid lists printed in books that do not relate to the real world. Restricting discs from magnetic fields is often quoted and is only relevant if there is any chance that there could be a magnetic field in a computer room – few computer rooms would have a television.

- Backup copies of data and software to recover against any loss. Backup copies should be checked on a regular basis that they actually do work.
- Adequate training of staff to ensure proper processes are in force.
- Monitoring of all hardware by a supervisor/head of section, of software and data use to ensure proper methods are being used.
- A regular meeting of responsible personnel to review of security and measure its successes and failures.
- Preventing data and software being brought into the organisation. Computers without disc drives are one realistic way to prevent this. Making it a company rule is not enough.
- Ensuring proper environmental conditions that may require air conditioning.
- Providing food/drink areas away from the computers and regularly monitoring that no drinks are taken from there.
- Having disciplinary procedures for any violation revealed by monitoring processes.
- Regular maintenance of hardware. Possible provision of emergency power supply.
- Prevention of access to sensitive areas by access controls. This can be password entry at doors.
- Security guards for LARGE institutions but this is clearly unreasonable for a small business with only a small number of staff.
- Having built-in virus checkers that automatically check any new input. Having a regular virus checker update is essential. Also, a policy is needed to make everybody aware of the procedures that must be undertaken when a virus is detected or even suspected.
- Implement an e-mail policy and prevent at source, e-mails entering the system except from valid sources.
- Any data or software on removable discs should be signed in and out with one person responsible for this process and these discs locked away when not needed. This media should be clearly labelled to prevent any accidental use of the wrong one.

ENSURING PRIVACY

- Provide adequate and ongoing training in the importance of privacy and the procedures to be adopted.
- Monitor the privacy actually practised.
- In open areas, ensure screens are not left on unattended. Terminals should be well apart to avoid eavesdropping.
- Encrypt data.
- Set up a firewall to reduce external unauthorised access to data.
- Allocate complex passwords to data users. Ensure passwords are regularly changed with the system warning of the deadline for changes. No passwords are to be written down. Supervisors must monitor the use of passwords.
- Use passwords that are difficult to guess. Software can check the strength of a password.
- After three unsuccessful attempts at entering a password, the user should be logged out and a relevant supervisor immediately notified.
- Set up different levels of password access on a “need-to-know” basis.
- Log all access to sensitive data with user name, time and date and computer identifier.
- Set penalties for staff proved to have violated any privacy requirement.
- A regular meeting of responsible personnel to review of privacy and measure its successes and failures.

Candidates often suggest the use of passwords but do not state the reason for them and how they are implemented. There is an obvious difference between software passwords preventing illegal access to data and room passwords achieved by entering an access code on a keypad. The detail is important to prove that the candidate understands and is not just recalling a list he/she has learnt.