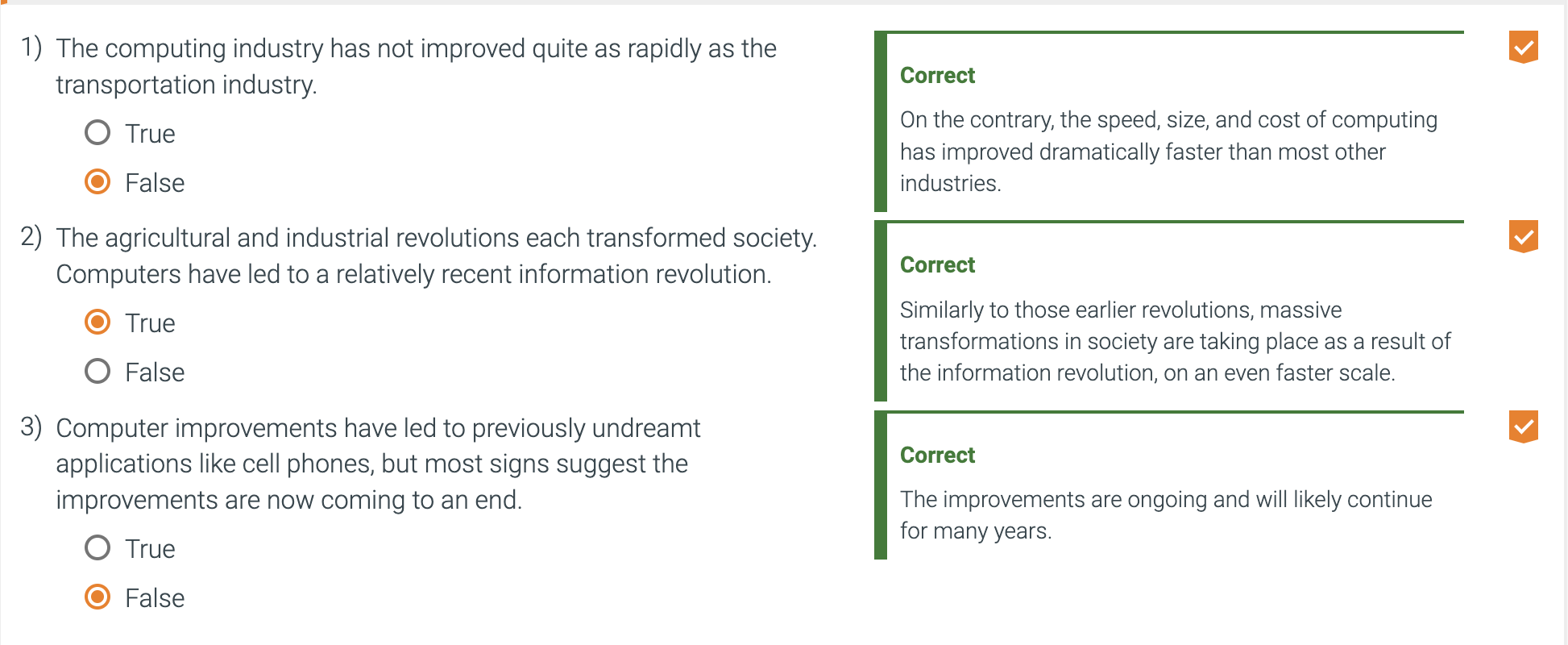
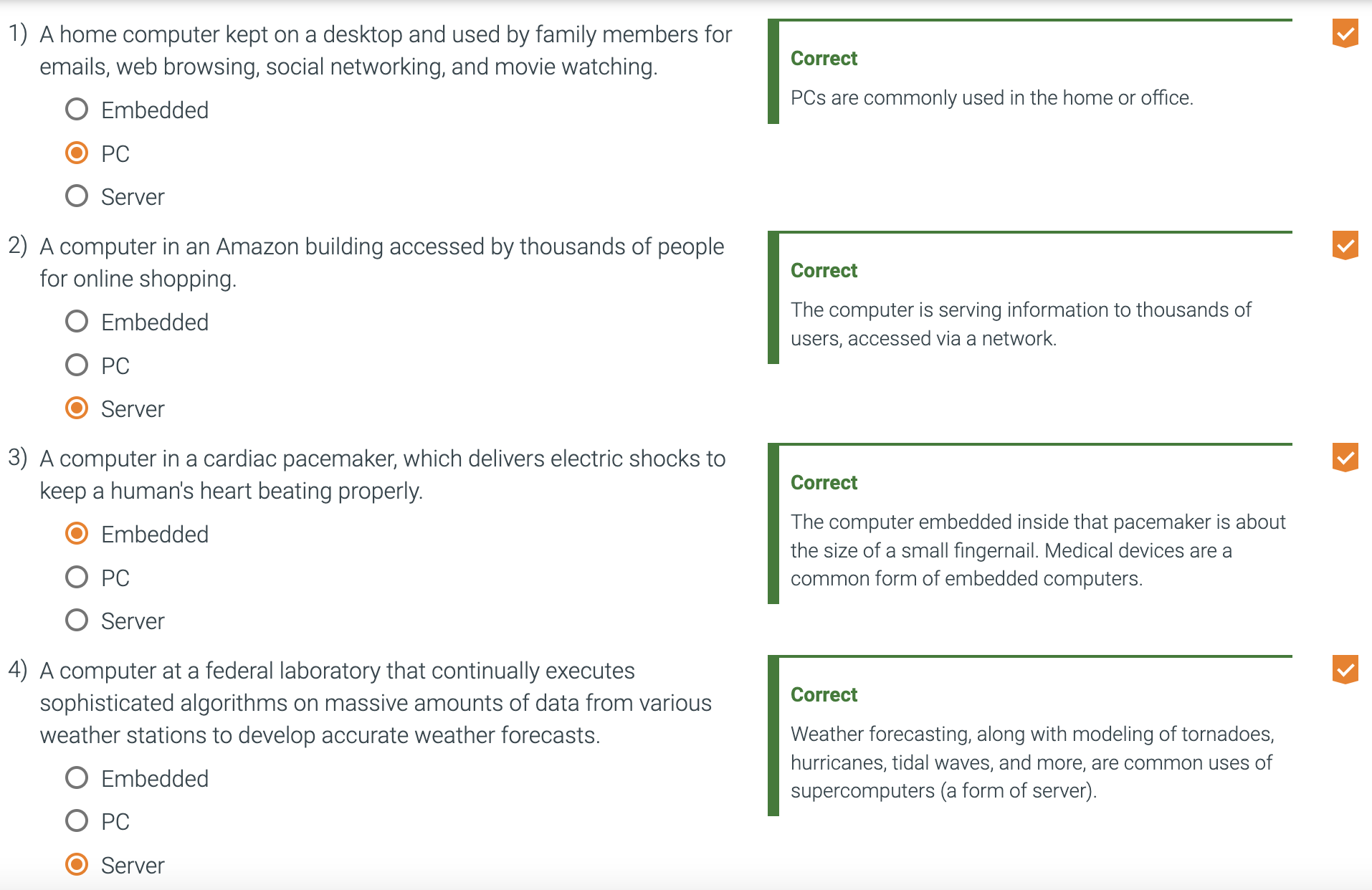
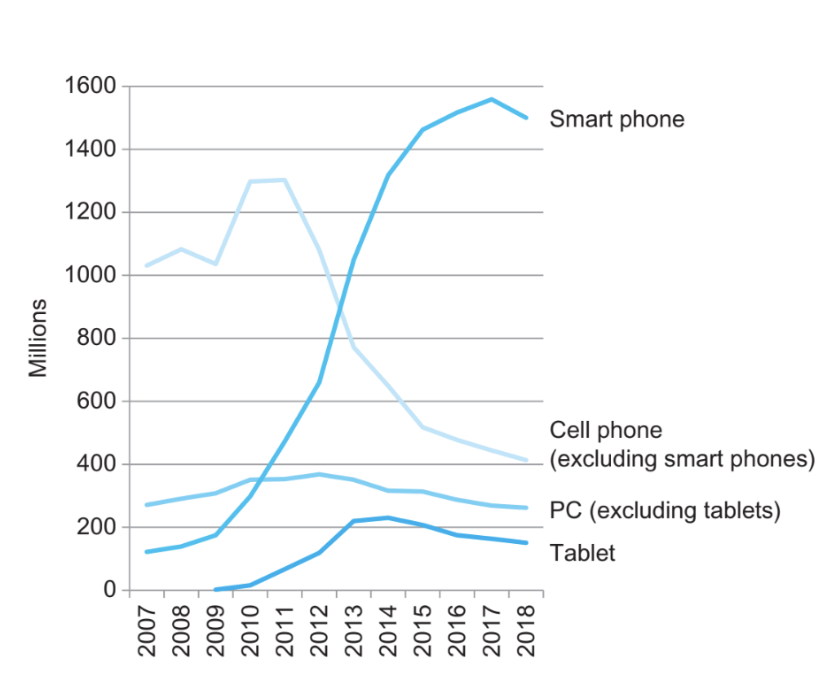
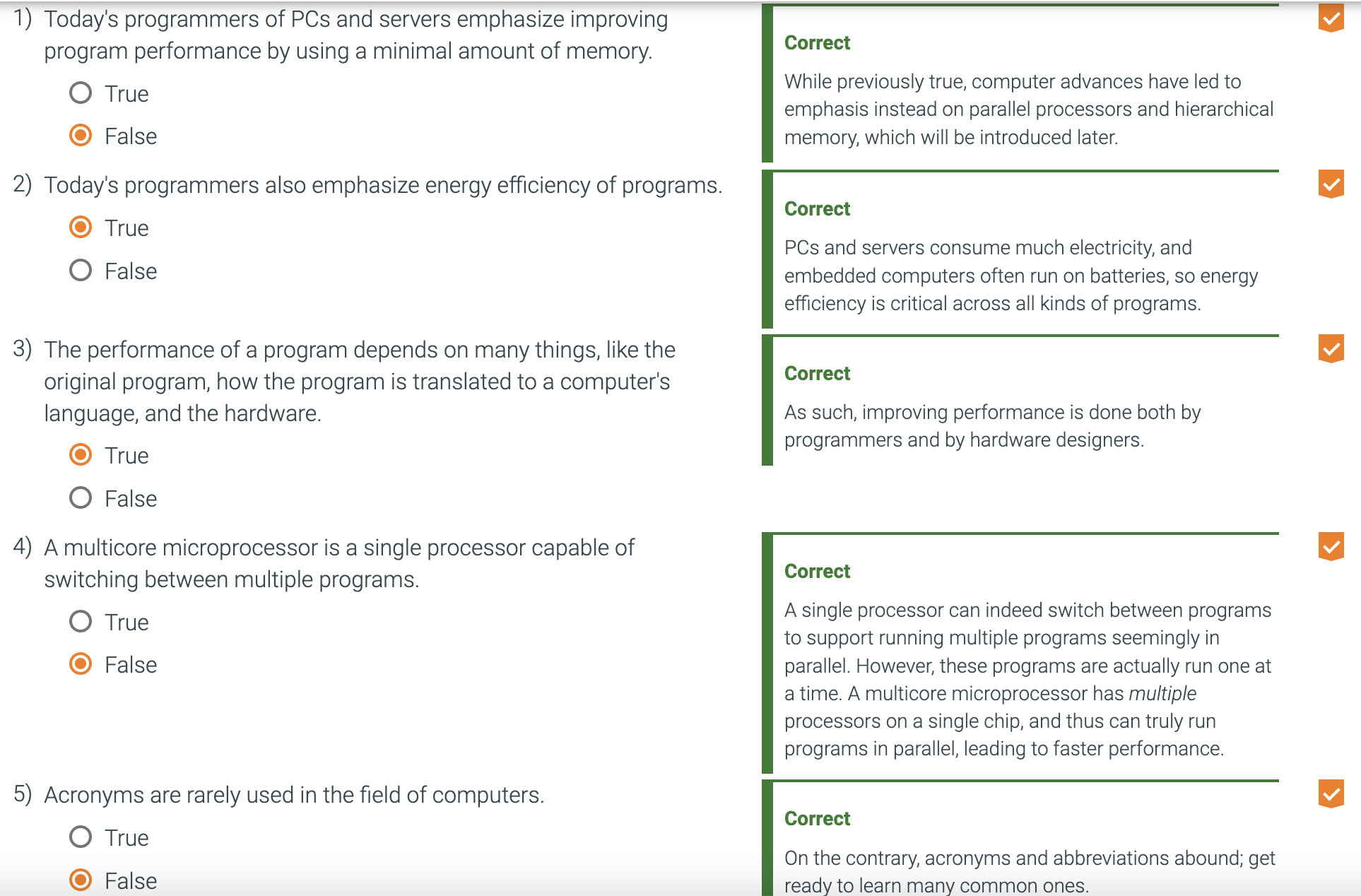
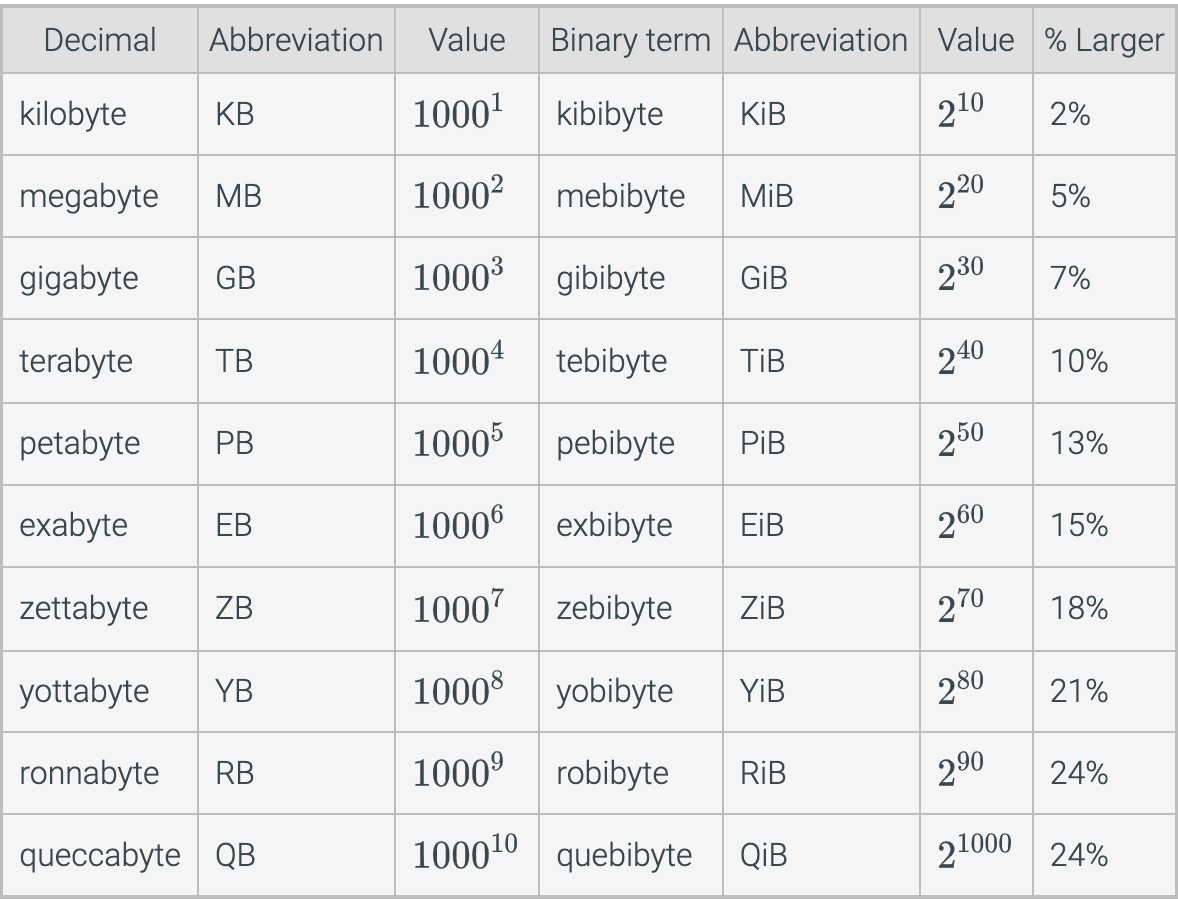
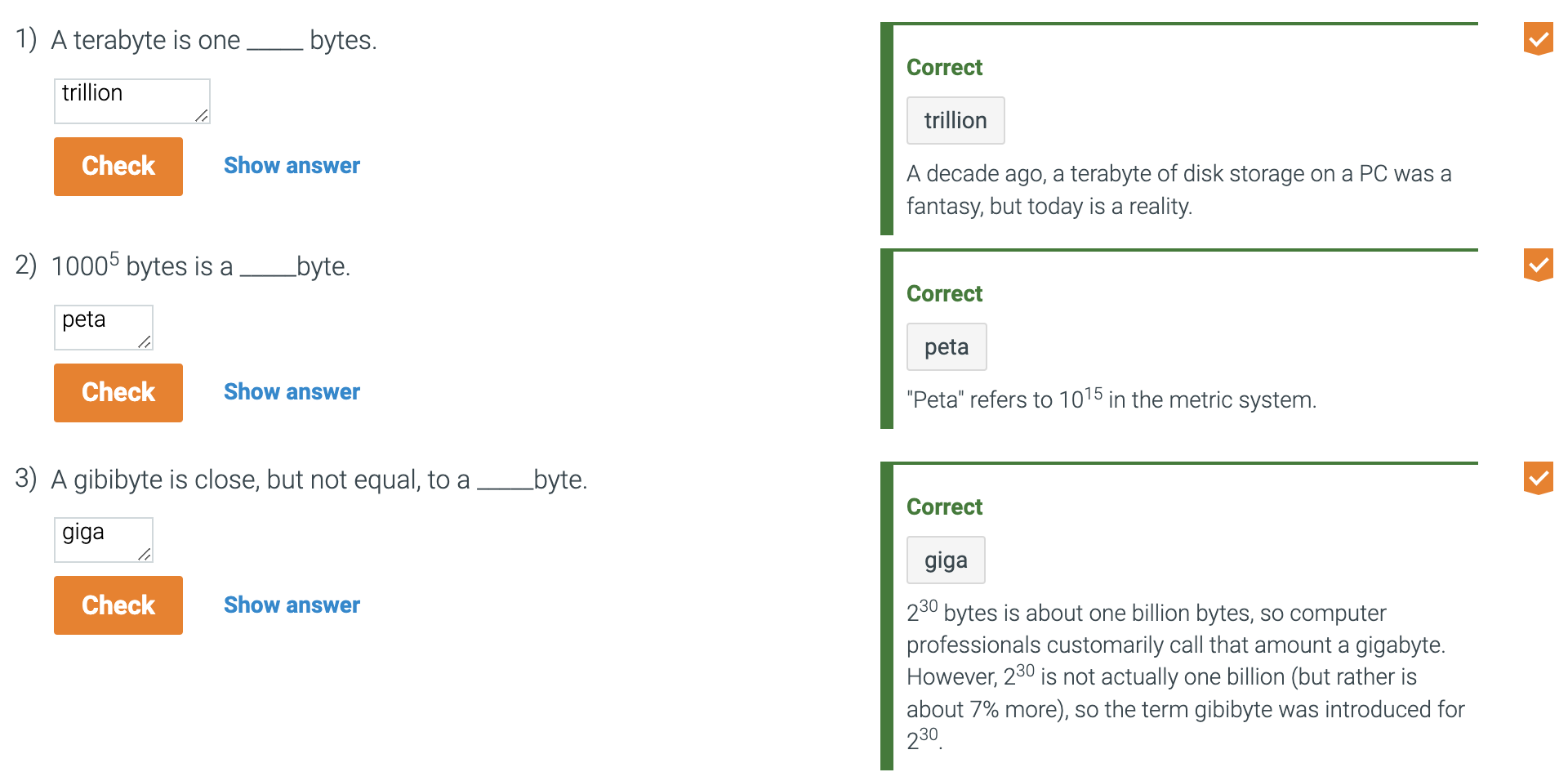
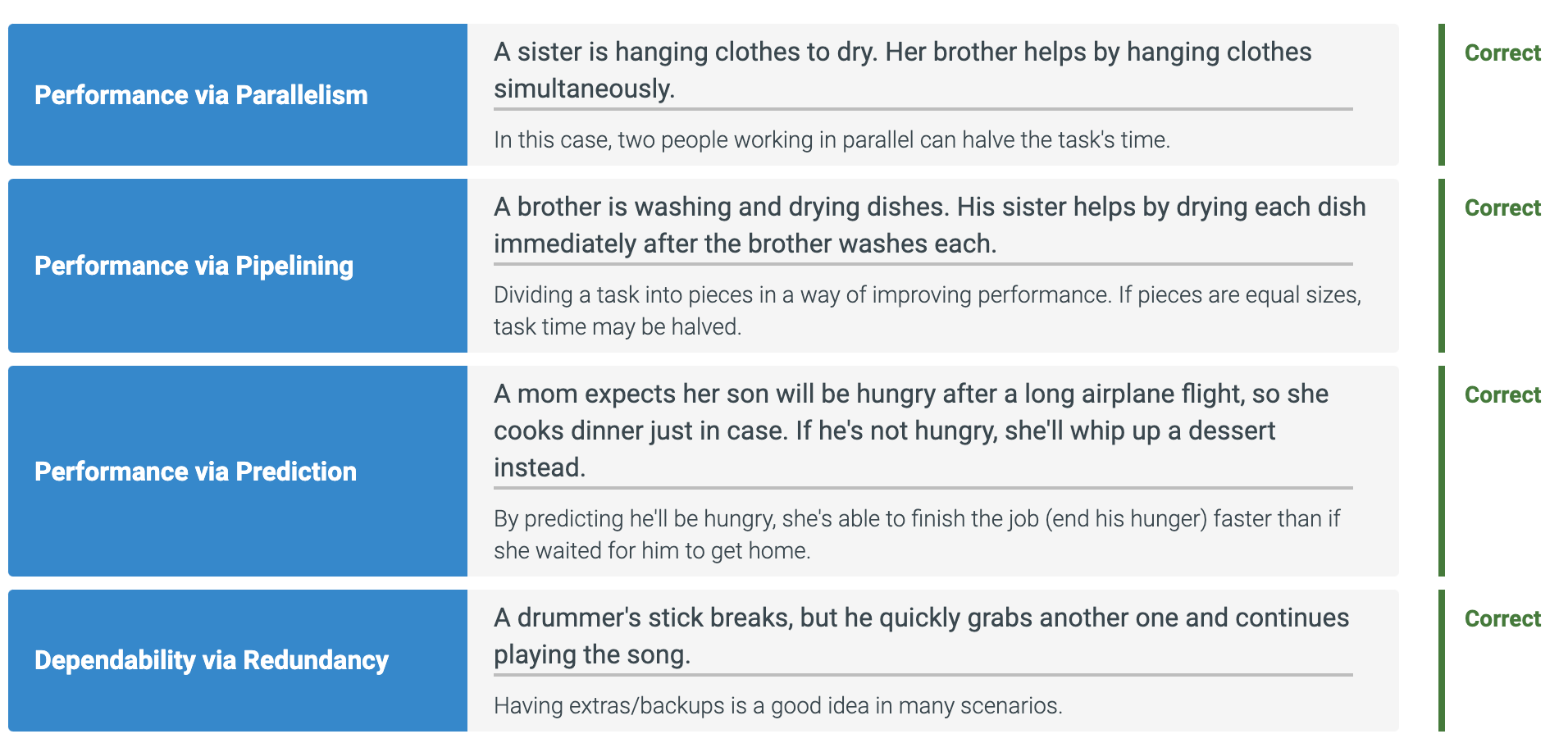
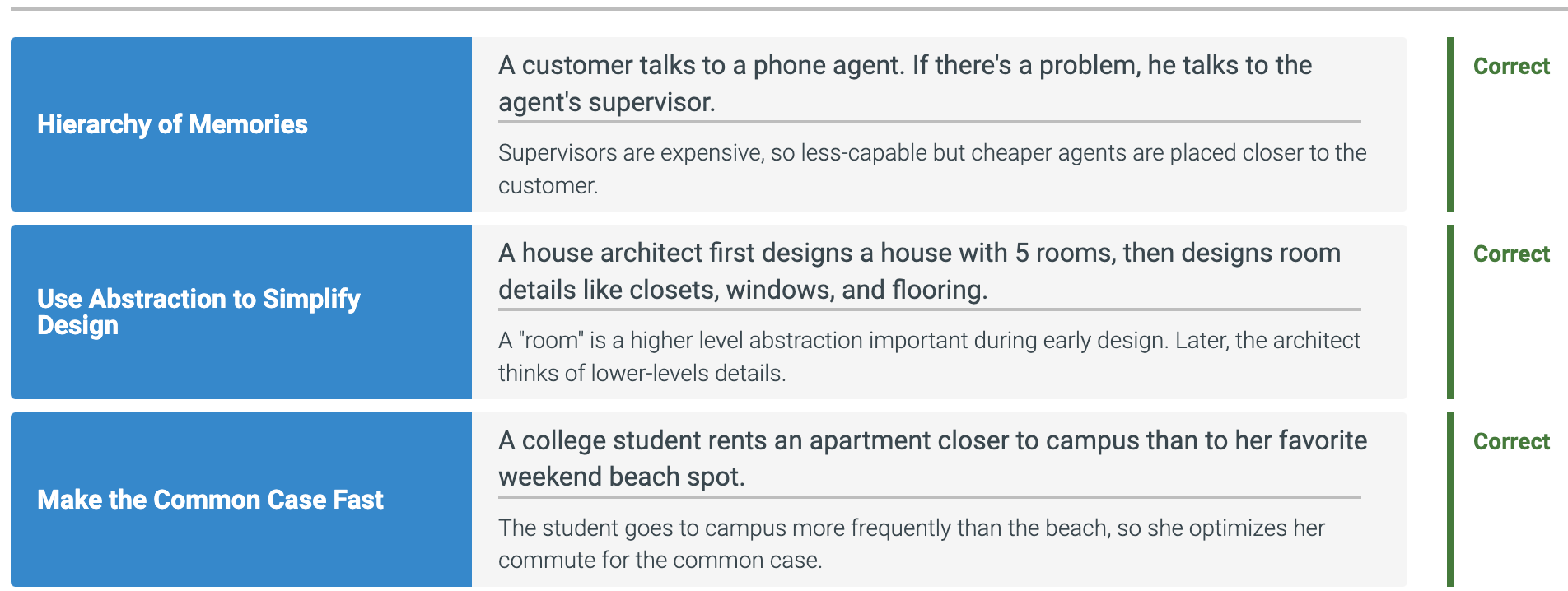
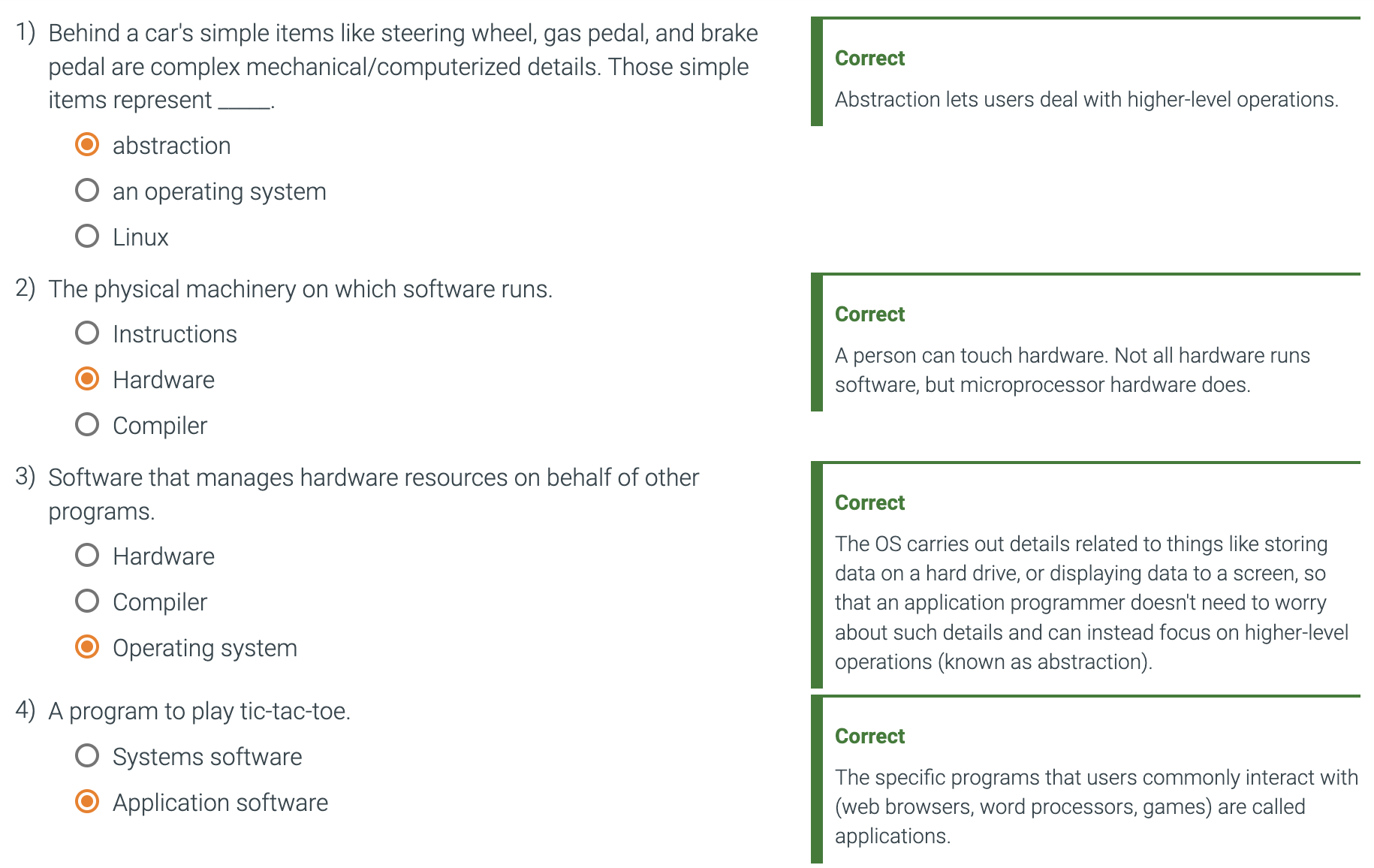
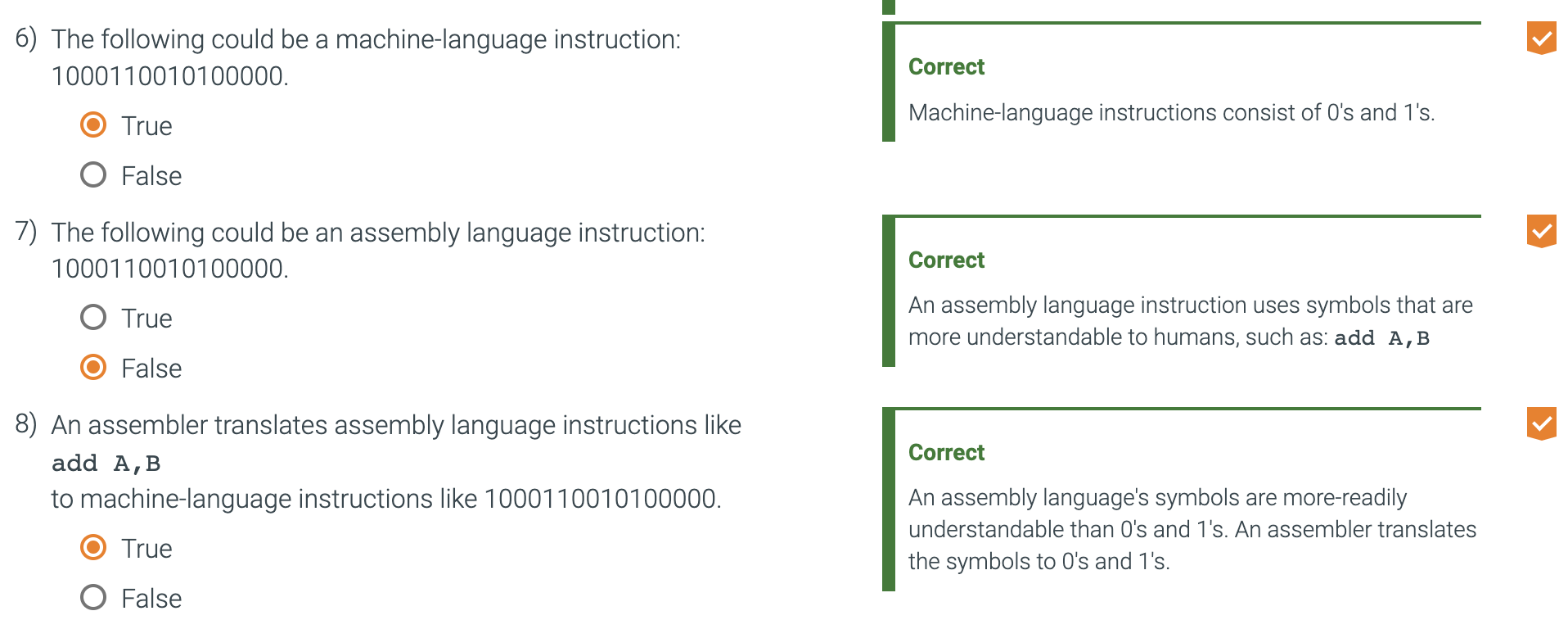
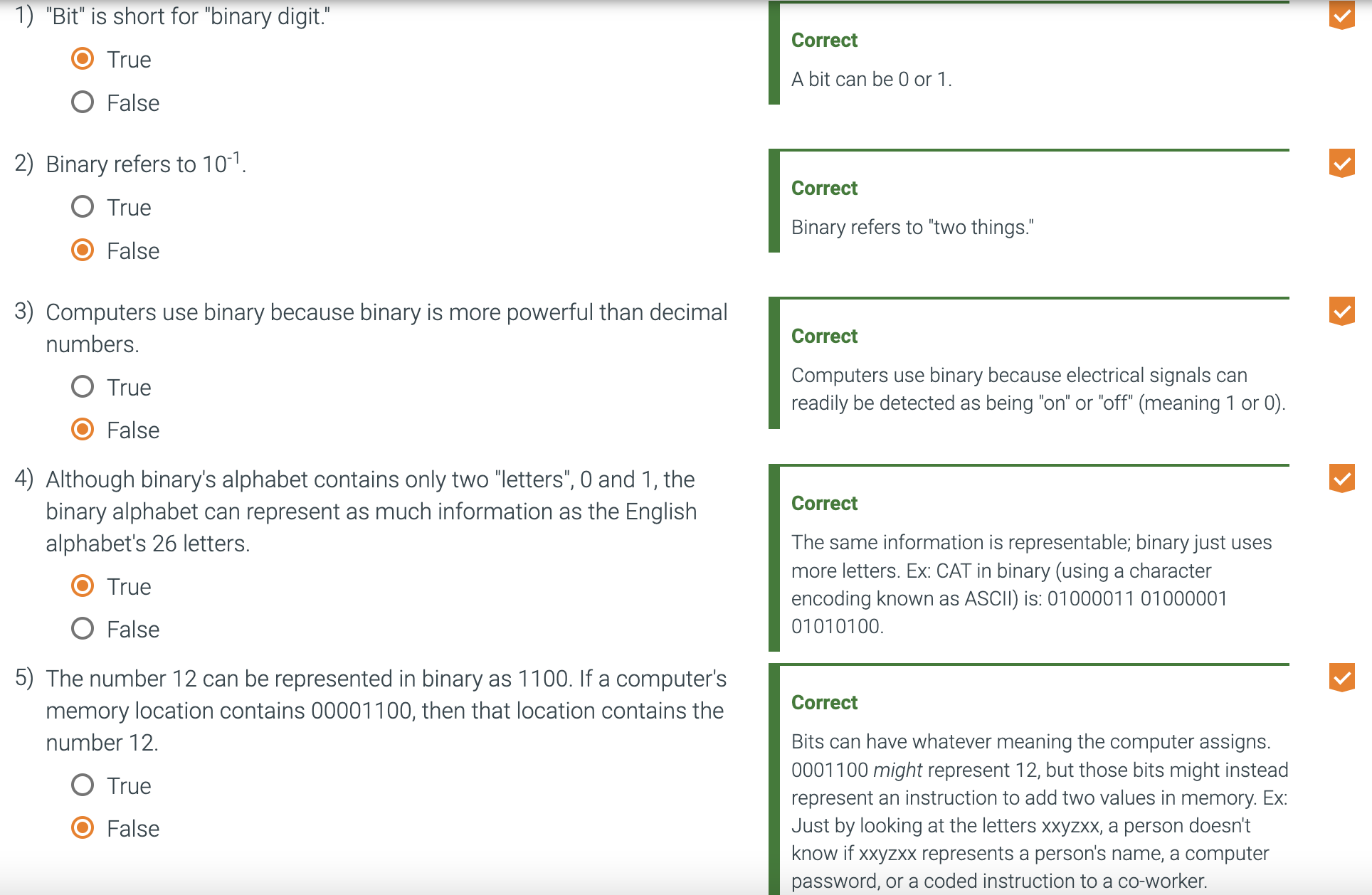
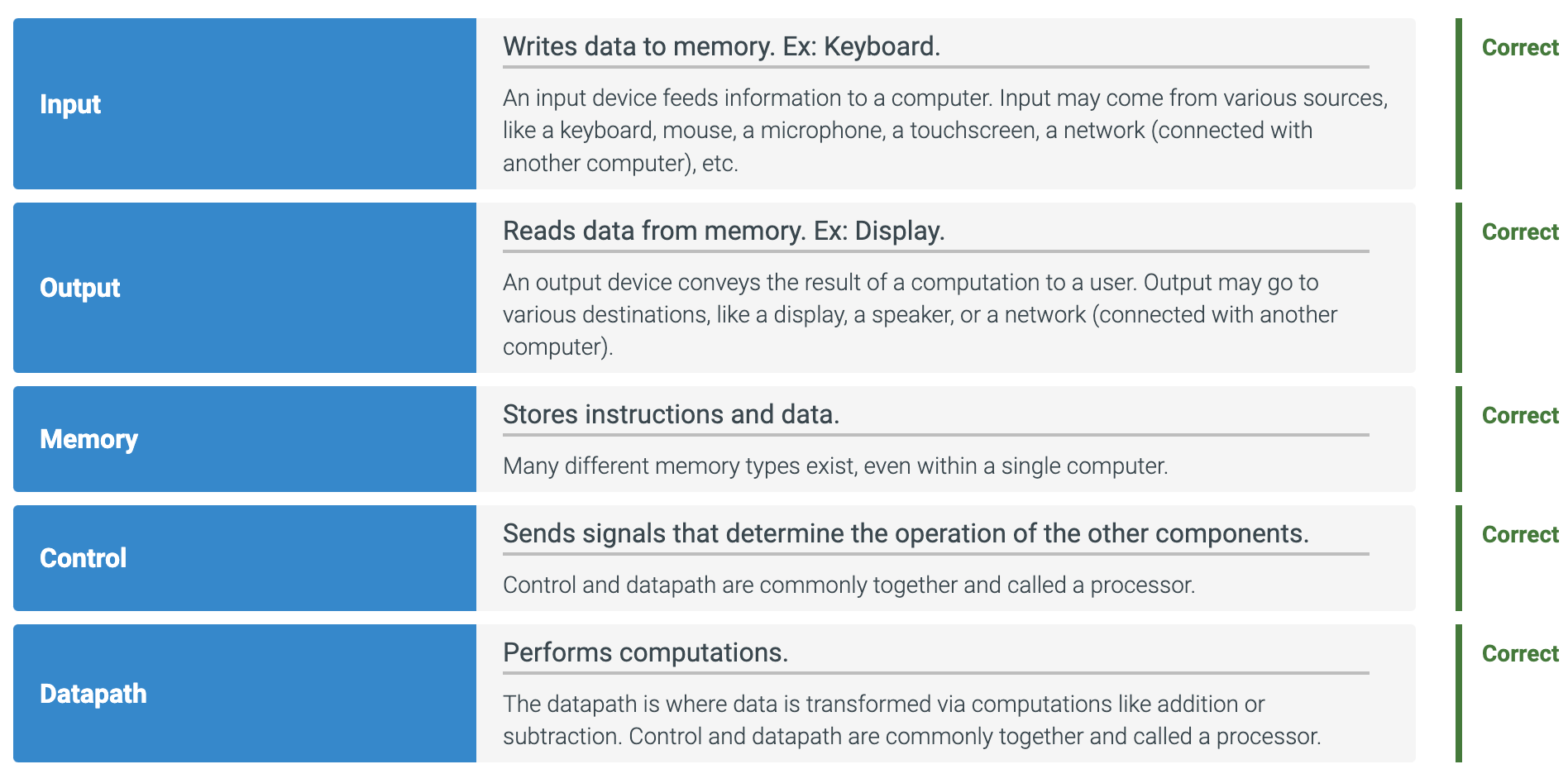
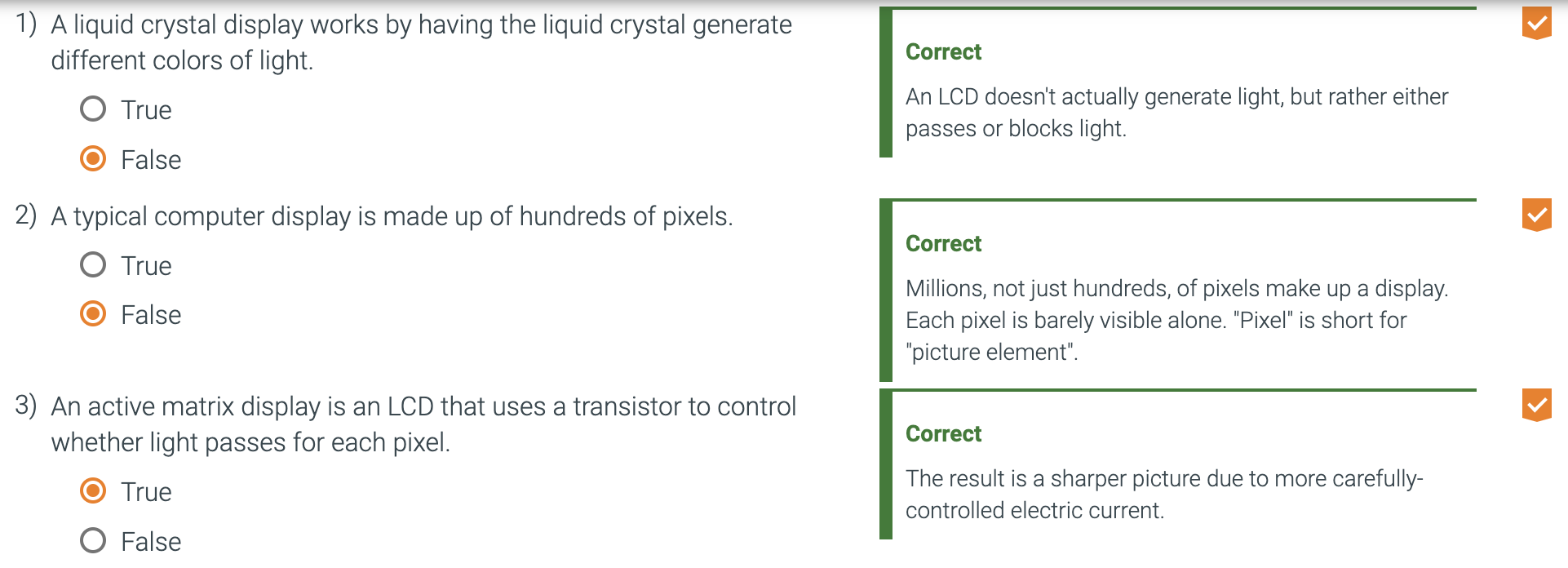
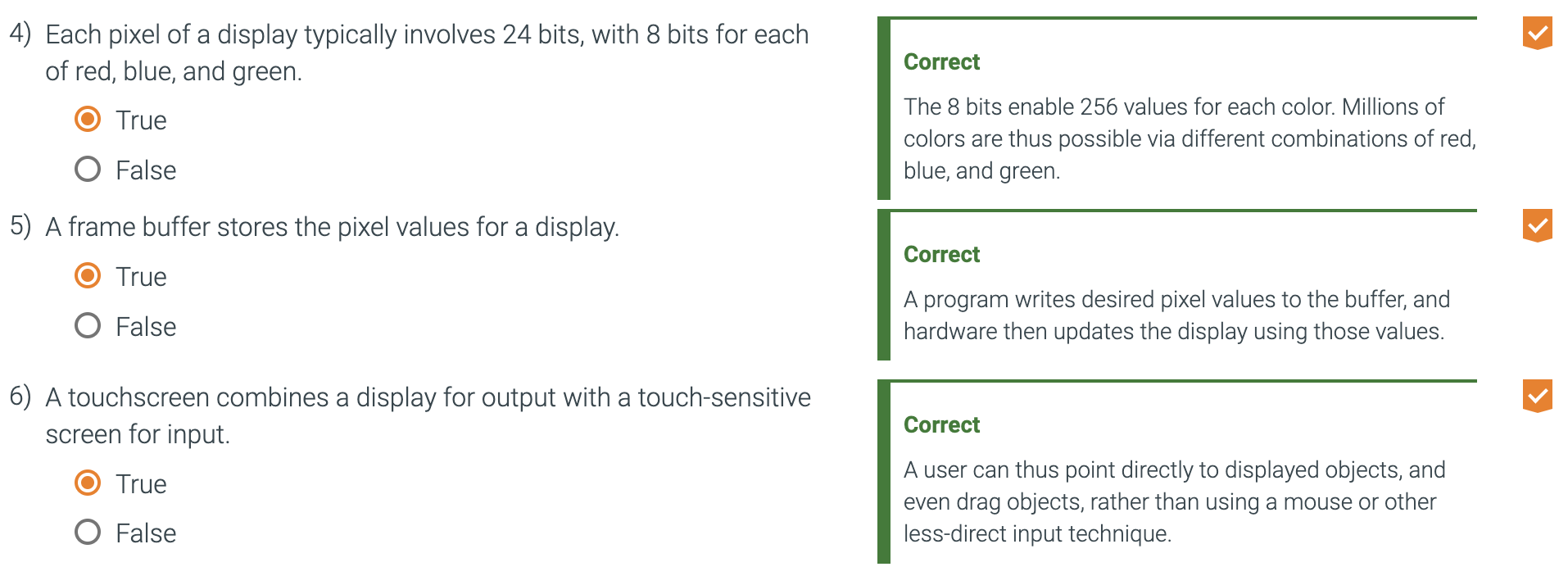
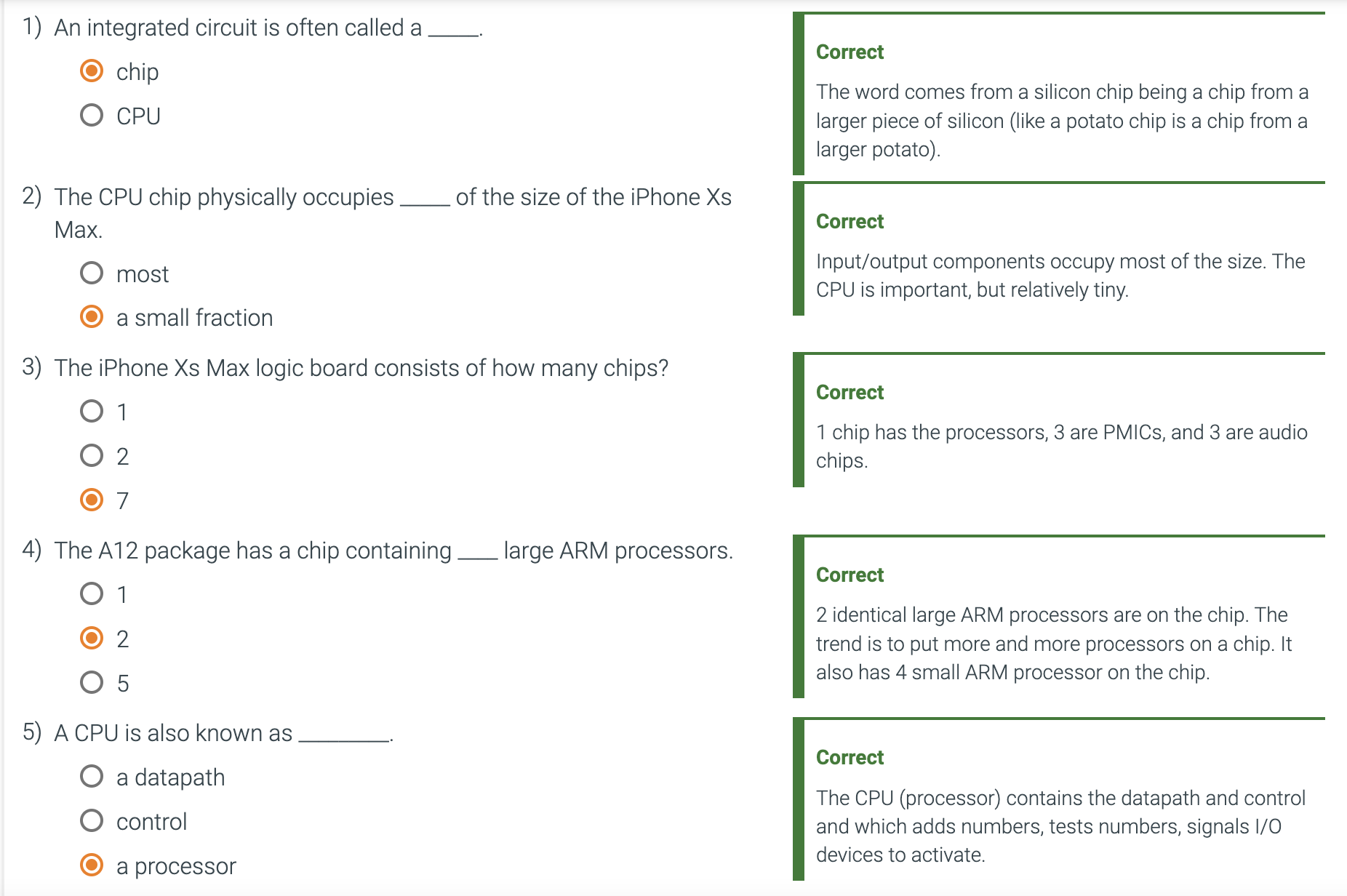
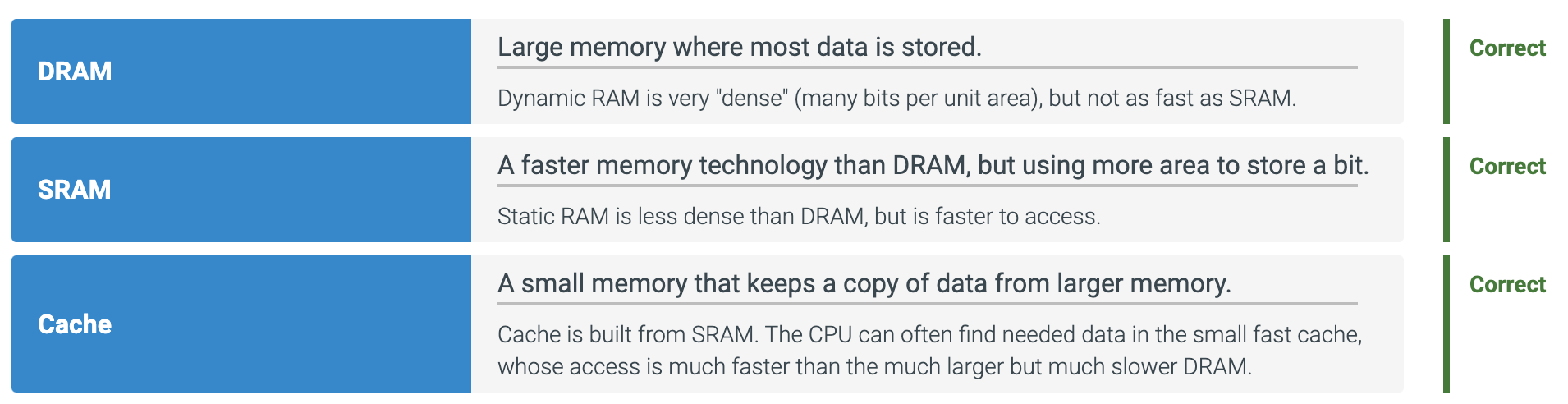
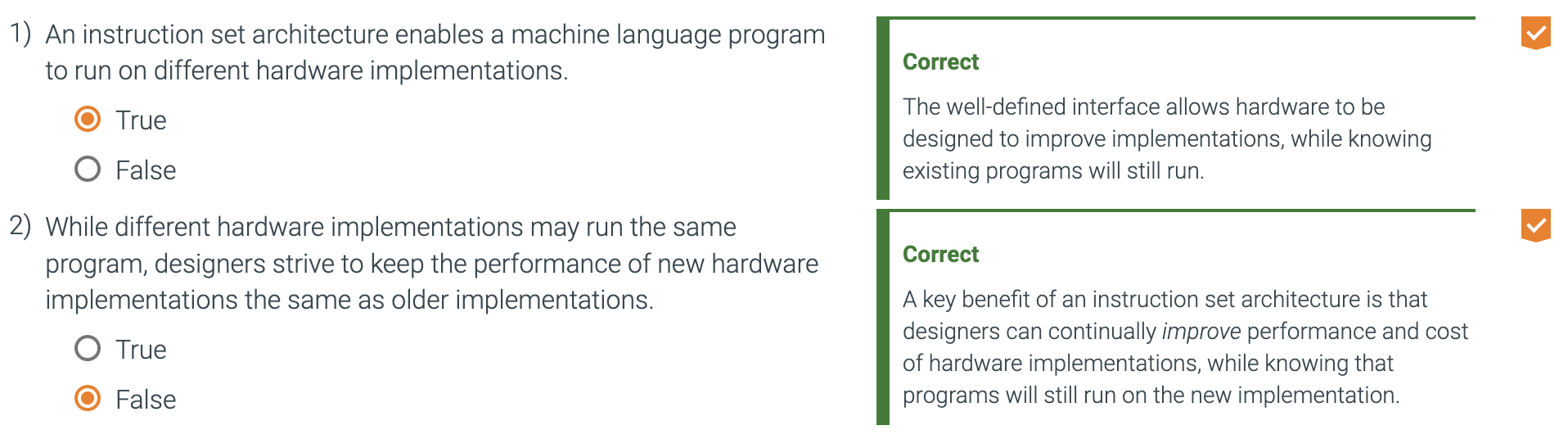
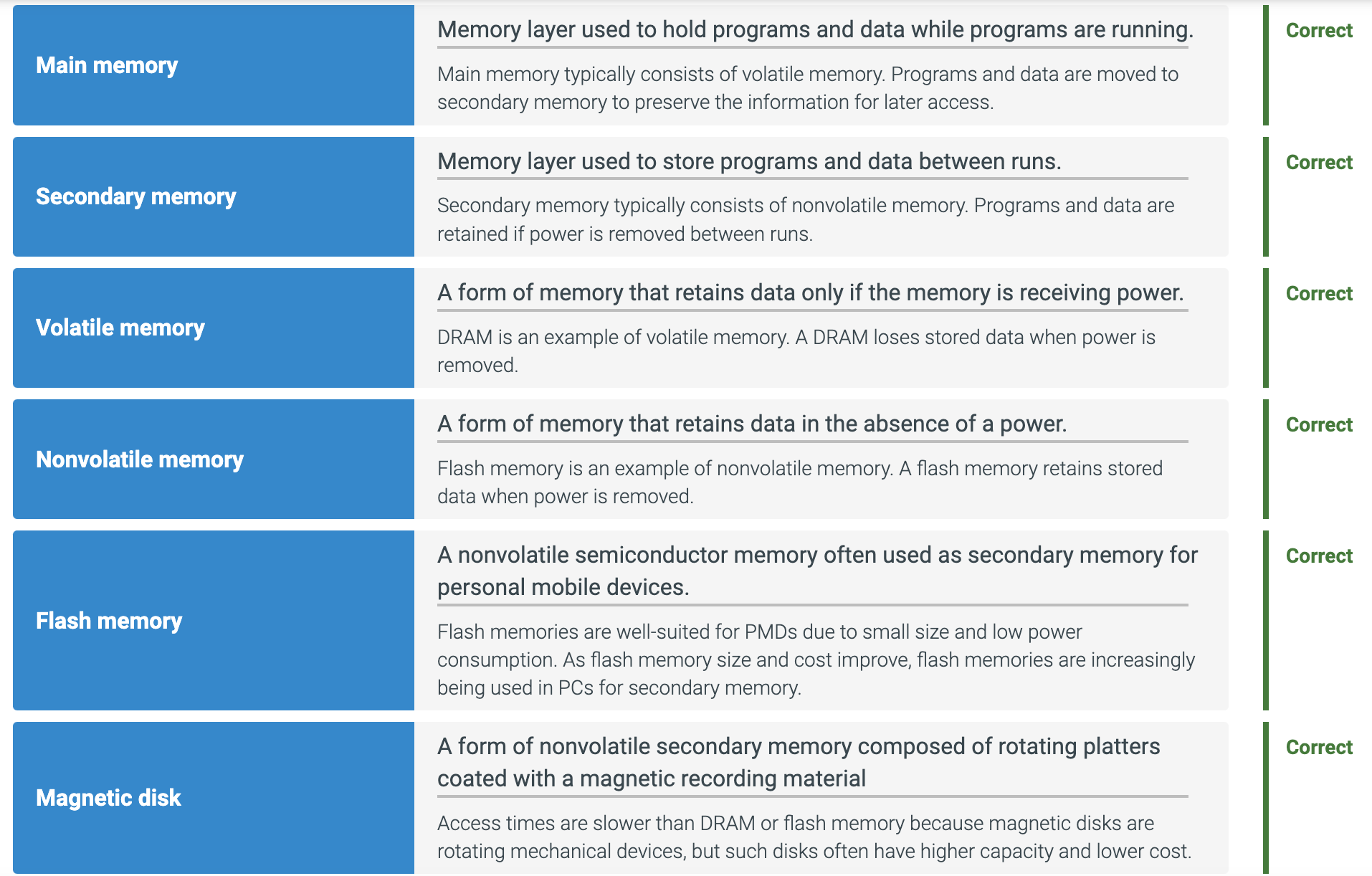
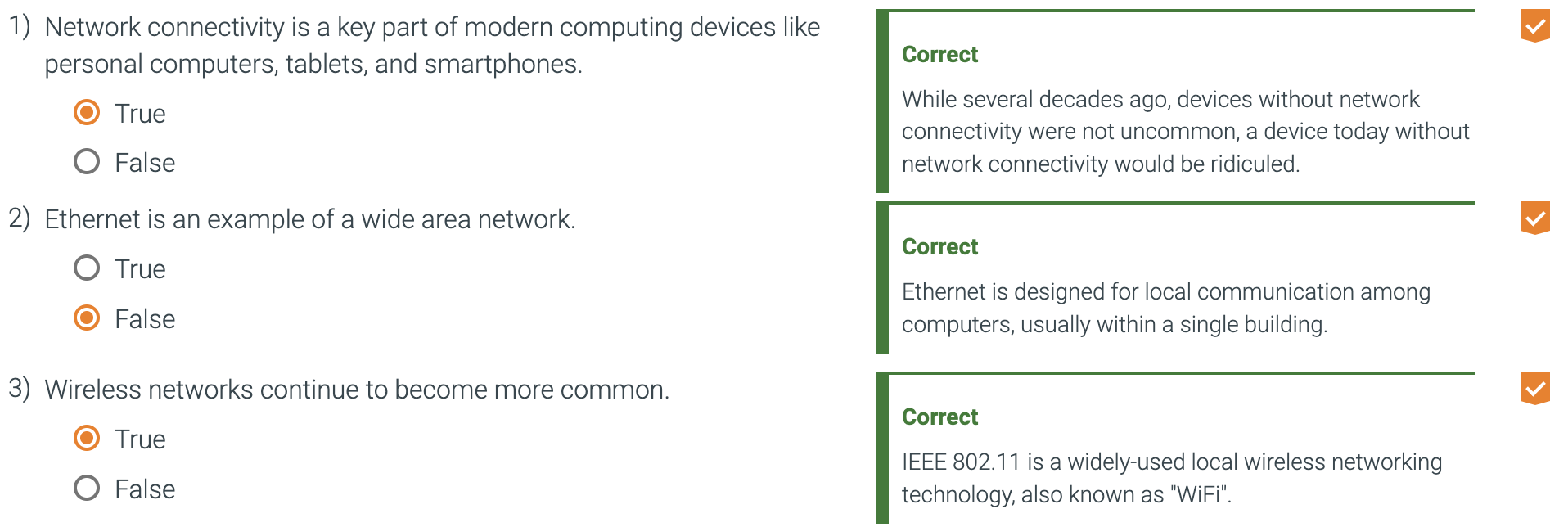
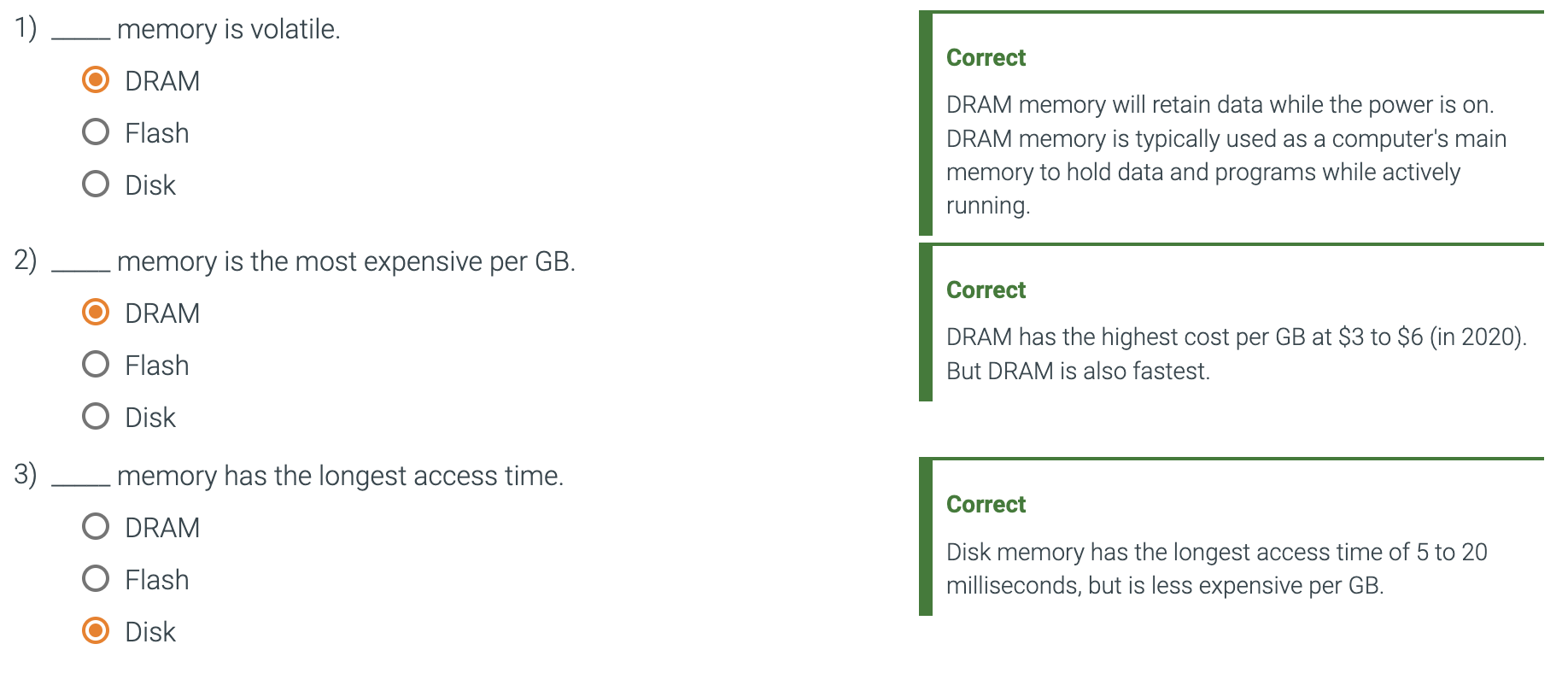
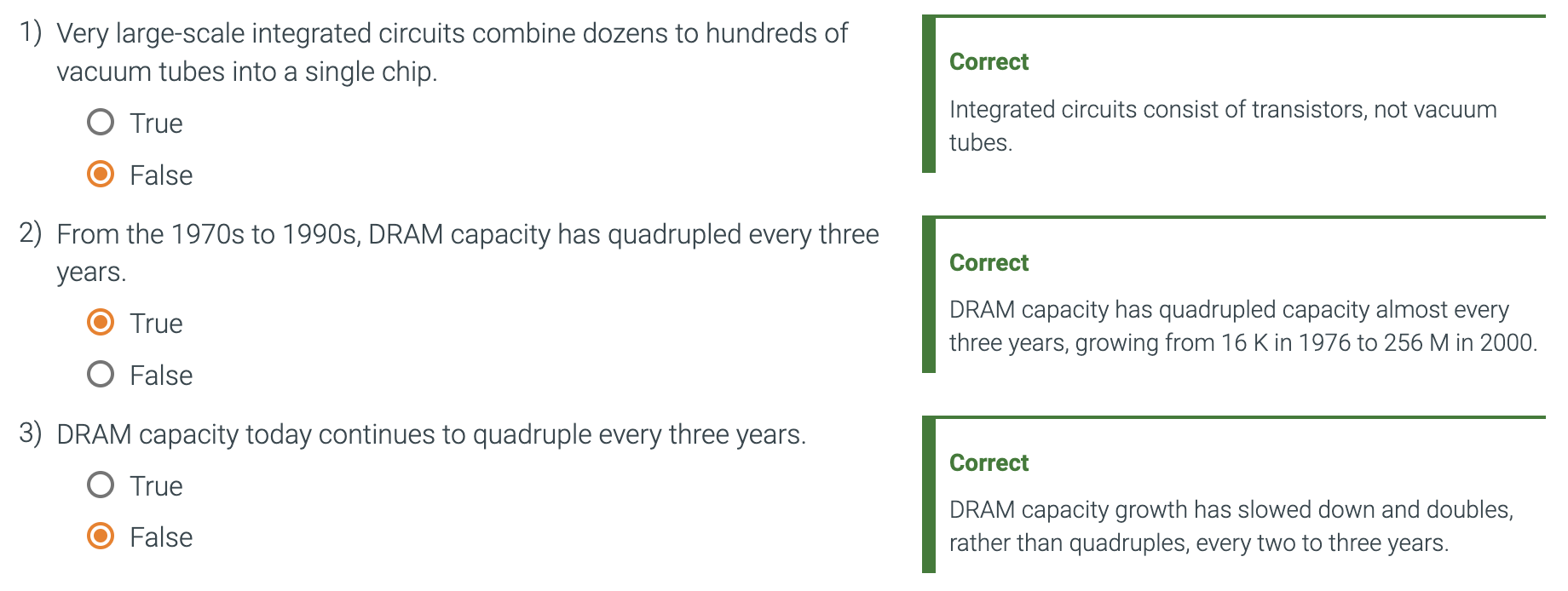
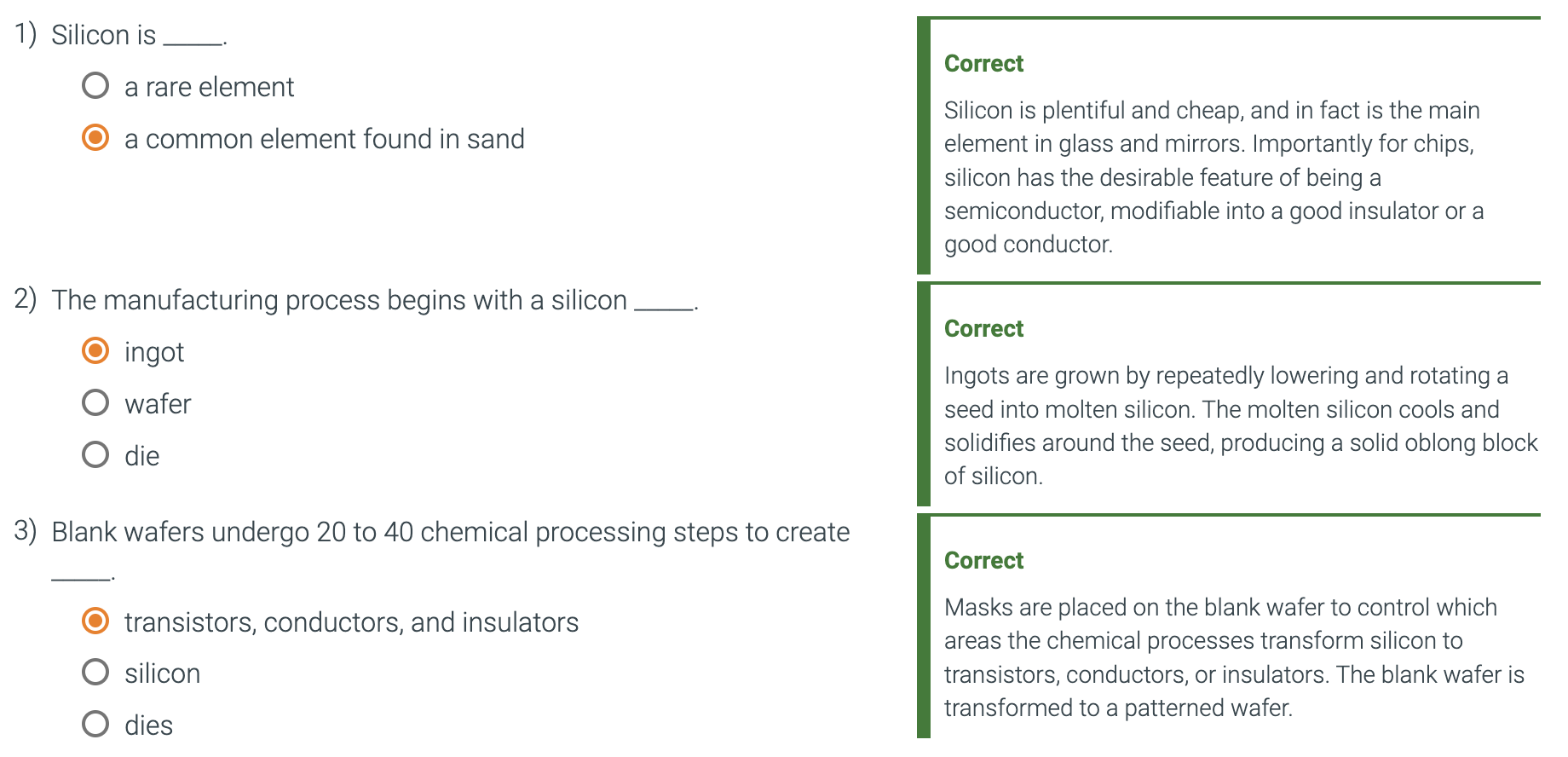
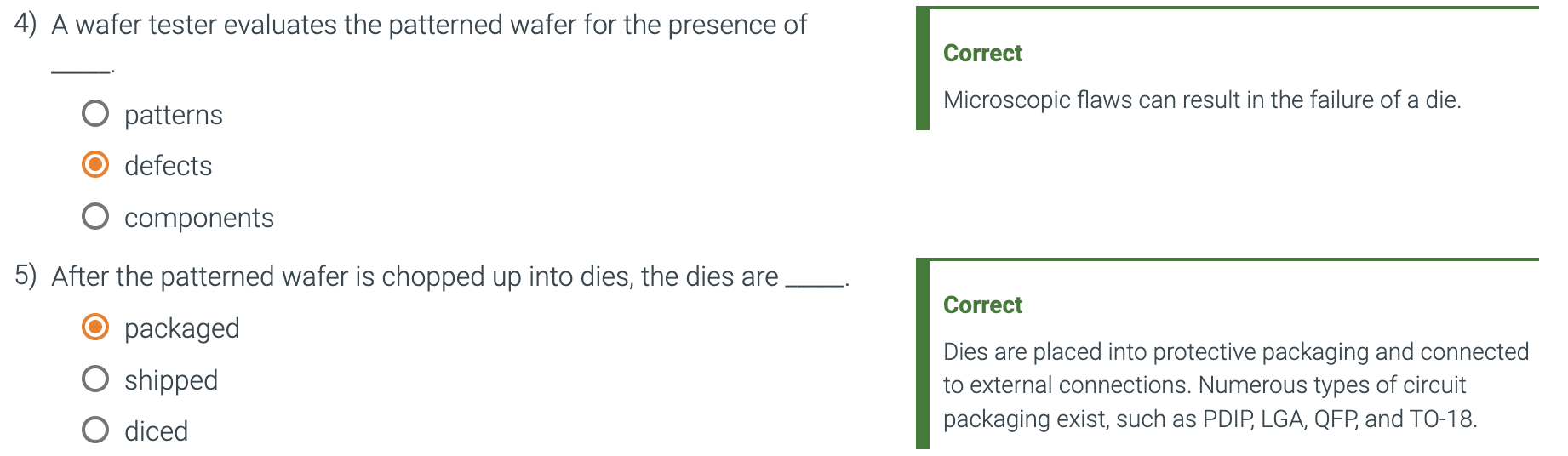
**Chapter 1.1 through 1.5 Notes**

* Introduction 1.1
  + Computers have led to a third revolution for civilization, with the **information revolution** taking its place alongside the agricultural and the industrial revolutions.
  + Each time the cost of computing improves by another factor of 10, the opportunities for computers multiply. Applications that were economically infeasible suddenly become practical.
  + In the recent past, the following applications were "computer science fiction."
    - Computers in automobiles
    - Cell Phones
    - Human Genome Project
    - World Wide Web
    - Search Engines
  + Broadly speaking, computers are used in three different classes of applications: personal computers, servers, and embedded computers.
    - A p**ersonal computer (PC)** is a computer designed for use by an individual, usually incorporating a graphics display, a keyboard, and a mouse.
    - An **embedded computer** is a computer inside another device used for running one predetermined application or collection of software.
      * The largest class of computers and they span the widest range of applications and performance.
      * Range from microprocessors in your car to computers in a TV to the networks of processors that control a modern airplane or cargo ship.
      * A popular term that relates to this would be the IOT or internet of things, which suggests many small devices that all communicate wirelessly over the internet.
      * Often have unique application requirements that combine a minimum performance with stringent limitations on cost or power.
      * Often have lower tolerance for failure, since the results can vary from upsetting to devastating.
      * Dependability is achieved primarily through simplicity—the emphasis is on doing one function as perfectly as possible.
    - A **server** is a computer used for running larger programs for multiple users, often simultaneously, and typically accessed only via a network.
      * Servers are the modern form of what were once much larger computers, and are usually accessed only via a network.
      * Oriented to carry large workloads or many small jobs.
      * Servers span the widest range in cost and capability.
      * They can range from a desktop computer with no screen or keyboard to a supercomputer.
      * Supercomputers are usually used for high-end scientific and engineering calculations.
        + They represent a small fraction of the overall computer market in terms of revenue.
  + **Elaborations** are short features used throughout the text to provide more detail on a particular subject that may be of interest.
  + Personal mobile devices (PMDs) are battery operated with wireless connectivity to the Internet and typically cost hundreds of dollars, and, like PCs, users can download software ("apps") to run on them.
    - No keyboard or mouse, usually have a touch screen or voice recognition.
    - Smart phones and PMD have shown massive growth compared to normal computers.
  + **Cloud Computing** is taking over the traditional server.
    - Relies upon giant datacenters that are now known as **Warehouse Scale Computers (WSCs)**.
    - Today, developers often have a portion of their application that runs on the PMD and a portion that runs on the cloud.
  + **Software as a Service (SaaS)** delivers software and data as a service over the Internet, usually via a thin program such as a browser that runs on local client devices, instead of binary code that must be installed, and runs wholly on that device.
    - Examples include web search and social networking.
  + In the last two decades, advances in computer design and memory technology have greatly reduced the importance of small memory size in most applications other than those in embedded computing systems.
    - We no longer minimize memory to make programs faster.
    - Today’s programmers worry about the energy efficiency of their programs either on the PMD or in the Cloud.
  + **Multicore microprocessor:** A microprocessor containing multiple processors ("cores") in a single integrated circuit.
  + How does a hardware or software component affect performance:
    - Algorithm
      * Determines both the number of source-level statements and the number of I/O operations executed.
    - Programming language, compiler, and architecture
      * Determines the number of computer instructions for each source-level statement.
    - Processor and memory system
      * Determines how fast instructions can be executed.
    - I/O System (Hardware and operating system)
      * Determines how fast I/O operations may be executed.
  + Terabyte (TB)
    - Originally 1,099,511,627,776 (2^40) bytes, but developers went on to use 1,000,000,000,000 (10^12) bytes.
    - To reduce confusion we now use **tebibyte (TiB)** for 2^40 bytes and **terabyte (TB)** to mean 10^12 bytes.



* Seven great ideas in computer architecture 1.2
  + Use Abstraction to Simplify design
    - A major productivity technique for hardware and software is to use **abstractions** to represent the design at different levels of representation.
    - Lower-level details are hidden to offer a simpler model at higher levels.
  + Make the Common Case Fast
    - Making the **common case** fast will enhance performance better than optimizing the rare case.
    - Since the common case is usually simpler, it is easier to enhance.
    - This implies you know what the common case is.
  + Performance Via Parallelism
    - Designs that offer more performance are operations performed in parallel.
  + Performance Via Pipelining
    - Pipelining is a pattern of parallelism.
    - It moves multiple operations through hardware units that each do a piece of an operation
  + Performance Via Prediction
    - The idea around prediction is that in some cases it can be faster on average to guess and start working rather than wait until you know for sure.
  + Hierarchy of Memories
    - Programmers can address conflicting demands of fast, large, and cheap memory with a **hierarchy of memories**.
    - The fastest, smallest, and most expensive memory per bit at the top of the hierarchy.
    - The slowest, largest, and cheapest per bit at the bottom.
  + Dependability Via Redundancy
    - Computers need to be dependable.
    - We can make systems dependable by including redundant components that can take over when a failure occurs and to help detect failures.
    - ‘**Designing for Moore’s Law’ by Gordon Moore**, one of the founders of intel, made a prediction in 1965 that integrated circuit resources would double every year.
      * This shaped computer architecture.
      * Since computer designs can take years, the resources available per chip can easily double or triple between the start and finish of the project.
      * Since exponential growth cannot last forever, Moore’s law is not accurate anymore.
    - Semiconductor technology will continue to improve, but much more slowly than in the past.
* Below Your Program 1.3
  + **Systems Software** is software that provides services that are commonly useful, including operating systems, compilers, loaders, and assemblers.
    - Two types of systems software are central to every computer system today: an operating system and a compiler.
    - An **operating system** interfaces between a user’s program and the hardware and provides a variety of services and supervisory functions.
      * Supervising program that manages the resources of a computer for the benefit of the programs that run on that computer.
    - The most important functions are:
      * Handling basic input and output
      * Allocating storage and memory
      * Providing for protected sharing of the computer among multiple applications using it simultaneously.
    - Hardware is the physical machine that runs software
    - Systems Software, like an operating system, runs on the hardware and provides useful services.
    - Application software runs on top of systems software.
    - A **compiler** is a program that translates high-level language statements into assembly language statements.
  + Computers are slaves to our commands, which are called instructions.
    - **Instructions** are collections of bits that the computer understands and obeys and can be thought of as numbers.
  + **Binary Digit**, also called a bit, is one of the two numbers in base 2 (0 or 1) that are the components of information.
  + An **Assembler** is a program that translates a symbolic version of instructions into the binary version.
  + **Assembly Language** is a symbolic representation of machine instructions.
  + **Machine Language** is a binary representation of machine instructions.
  + **High-Level Programming Language** is a portable language such as C, C++, or Visual Basic that is composed of words and algebraic notation that can be translated by a compiler into assembly language.
    - There are several important benefits of High level programming languages.
      * First, allow the program to think in a more natural language, resulting in programs that look much more like text than symbols.
      * Second, is improved programmer productivity.
      * Third would be that the programming languages allow programs to be independent of the computer on which they were developed, since compilers and assemblers can translate high-level language programs to the binary instructions of any computer.
* Under the Covers 1.4
  + Input device
    - A mechanism through which the computer is fed information, such as a keyboard.
  + Output device
    - A mechanism that conveys the result of a computation to a user, such as a display for another computer.
  + The five classic components of a computer are:
    - Input
      * Input write to memory.
    - Output
      * Output read data from memory
    - Memory
      * Processor gets its instructions and data from memory.
    - Datapath
      * Operates on data.
    - Control
      * Control sends signals that determine operations of datapath, memory, input, and output.
  + Liquid Crystal Display (LCD)
    - A display technology using a thin layer of liquid polymers that can be used to transmit or block light according to whether a charge is applied.
    - Controls transmission of light, does not control the light.
    - Includes rod-shaped molecules in a liquid that forms a twisting helix that bends light entering the display.
      * Either from behind the display or less often from reflected light.
  + Active Matrix Display
    - A liquid crystal display using a transistor to control the transmission of light at each individual pixel.
  + Bit Map
    - An image composed of matrix picture elements or pixels which can be represented as a matrix of bits.
  + Pixel
    - The smallest individual picture element
    - Screens are composed of hundreds of thousands to millions of pixels organized in a matrix.
    - Every pixel has a corresponding (x,y) coordinate.
    - Typically involves 24 bits with 8 bits for each of red, blue, and green.
  + Computer hardware support for graphics consists mainly of a raster refresh buffer, or frame buffer, to store the bit map.
    - Image to be displayed is stored in the frame buffer.
    - The bit pattern per pixel is read out the the graphics display at the refresh rate.
  + Touchscreen displays are replacing keyboard and mouse.



* + The list of I/O devices includes a capacitive multitouch LCD display, front-facing camera, rear-facing camera, microphone, headphone jack, speakers, accelerometer, gyroscope, Wi-FI network, and Bluetooth network. 
    - The datapath, control, and memory are a tiny portion of the components
  + Integrated Circuit
    - Also called a chip.
    - A device combining dozens to millions of transistors.
  + Central Processor Unit (CPU)
    - Also called processor.
    - The active part of the computer, which contains the datapath and control and which adds numbers, tests numbers, signals I/O devices to activate, and so on.
  + The processor usually contains two main components, the datapath and the control.
    - Datapath performs the arithmetic instructions on the processor.
    - The control tells the datapath, memory, and I/O devices what to do according to the instructions of the program.
  + Datapath (processor)
    - The component of the processor that performs arithmetic operations.
  + Control (processor)
    - The component of the processor that commands the datapath, memory, and I/O devices according to the instructions of the program.
  + Memory
    - Storage area in which programs are kept when they are running and that contains data needed by the running programs.
  + Dynamic Random Access Memory (DRAM )
    - Memory built as an integrated circuit; it provides random access to any location. Access times are 50 nanoseconds and cost per gigabyte in 2020 was $3 to $6.
    - DRAMs are used together to contain the instructions and data of a program
    - In contrast to sequential access memories, such as magnetic tapes, the RAM portion of the term DRAM means that memory accesses take basically the same amount of time no matter what portion of the memory is read.
  + Cache Memory
    - A small, fast memory that acts as a buffer for a slower, larger memory.
  + Static Random Access Memory (SRAM)
    - Also memory built as an integrated circuit, but faster and less dense than DRAM.
  + One of the most important abstractions is the interface between the hardware and the lowest-level software.
    - Software communicates to hardware via a vocabulary.
    - Words of the vocabulary are called instructions.
    - The vocabulary itself is called the instruction set architecture, or the architecture of a computer.
  + Instruction Set Architecture
    - Also called architecture.
    - An abstract interface between the hardware and the lowest-level software that encompasses all the information necessary to write a machine language program that will run correctly, including instructions, registers, memory access, I/O, and so on.
    - Allows computer designers to talk about functions independently from the hardware that performs them.
  + Application binary interface (ABI)
    - The user portion of the instruction set plus the operating system interfaces used by application programmers.
    - Defines a standard for binary portability across computers.
  + Implementation
    - Hardware that obeys the architecture abstraction.
  + **Both hardware and software consist of hierarchical layers using abstraction, with each lower layer hiding details from the level above.**
    - **One key interface between the levels of abstraction is the instruction set architecture—the interface between the hardware and low-level software.**
    - **This abstract interface enables many implementations of varying cost and performance to run identical software.**
  + Volatile Memory
    - Storage, such as DRAM, that retains data only if it is receiving power.
  + Nonvolatile Memory
    - Form of memory that retains data even in the absence of a power source and that is used to store programs between runs.
    - A dvd disk is nonvolatile.
  + Main Memory
    - Also called primary memory.
    - Memory used to hold programs while they are running
    - Typically consists of DRAM in today's computers.
  + Secondary Memory
    - Nonvolatile memory used to store programs and data between runs.
    - Typically consists of flash memory in PMDs and magnetic disks in servers.
  + Magnetic Disk
    - Also called hard disk.
    - A form of nonvolatile secondary memory composed of rotating platters coated with a magnetic recording material.
    - Because they are rotating mechanical devices, access times are about 5 to 20 milliseconds and cost per gigabyte in 2020 was $0.01 to $0.02.
  + Communication
    - Information is exchanged between computers at high speeds
  + Resource sharing
    - Rather than each computer having its own i/o devices, computers on a network can share them.
  + Nonlocal Access
    - By connecting computers over long distances.
  + Networks vary in length and performance, with the cost of communication increasing according to both the speed of communication and the distance that information travels.
    - Most popular type of network is ethernet
      * It can be up to a kilometer long and transfer at up to 100 gigabits per second.
      * Its length and speed make Ethernet useful to connect computers on the same floor of a building.
      * It is an example of what is generically called a local area network.
    - Wide area networks cross continents are the backbone of the web/
      * Typically based on optical fibers and are leased from telecommunication companies.
  + Local Area Network (LAN)
    - Network designed to carry data within a geographically confined area like a building.
  + Wide area Network (WAN)
    - Network extended over hundreds of kilometers that can span a continent.
  + Today, local area network technology offers a capacity of from 1 to 100 gigabits per second, usually shared by at most a few computers.
* Technologies for building processors and memory 1.5
  + Transistor
    - An on/off switch controlled by an electric signal
  + Very Large-Scale Integrated (VLSI) Circuit
    - A device containing hundreds of thousands to millions of transistors.
  + For decades, the industry has consistently quadrupled capacity every 3 years, resulting in an increase in excess of 16,000 times!
    - In recent years, the rate has slowed down and is somewhat closer to doubling every three years.
  + The manufacturing of a chip begins with silicon.
    - Silicon
      * A natural element that is a semiconductor
      * Materials can be added to silicon to allow areas to transform into one of three devices.
        + Excellent conductors of electricity (Microscopic copper or aluminum wire)
        + Excellent insulators from electricity (plastic sheathing or glass)
        + Areas that can conduct or insulate under special conditions (as a switch, transistors are this!)
    - Semiconductor
      * A substance that does not conduct electricity well.
    - Silicon crystal ingot
      * A rod composed of a silicon crystal that is between 8 and 12 inches in diameter and about 12 to 24 inches long.
    - Wafer
      * A slice from a silicon ingot no more than 0.1 inches thick, used to create chips.
  + Chip manufacturing processL
    - After being sliced from the silicon ingot, blank wafers are put through 20 to 40 steps to create patterned wafers.
    - These patterned wafers are then tested with a wafer tester, and a map of the good parts is made.
    - Then, the wafers are diced into dies.
    - These good dies are then bonded into packages and tested one more time before shipping the packaged parts to customers. One bad packaged part was found in this final test.
  + Defect
    - A microscopic flaw in a wafer or in patterning steps that can result in the failure of the die containing that defect.
  + Die
    - The individual rectangular sections that are cut from a wafer, more informally known as chips.
  + Dicing enables you to discard only those dies that were unlucky enough to contain the flaws, rather than the whole wafer.
    - This concept is quantified by the yield of a process, which is defined as the percentage of good dies from the total number of dies on the wafer.
  + Yield
    - The percentage of good dies from the total number of dies on the wafer.
  + Once you've found good dies, they are connected to the input/output pins of a package, using a process called bonding.
  + These packaged parts are tested a final time, since mistakes can occur in packaging, and then they are shipped to customers.