2.1) For this, we assume that it is consistent with the Markov Property, because there is no dependency between the tasks that the student is trying to solve. We define possible states as tuples:

$S_{x1, x2, x3, x4, x5, x6, x7}$ = <x1, x2, x3, x4, x5, x6, x7> where:

xi = { 1 if student solved task 'i'
0 otherwise; 1 <= 'i' <= 7

So, the number of states is $2^7 = 128$. Now we would have the following seven actions: **a1(s), a2(s), …, a7(s); s Є S,** where ai means to try to solve task 'i', so supossing that the student just started with state $S_{0,0,0,0,0,0,0} = <0, 0, 0, 0, 0, 0, 0>$ and he performed action a1 successfully, his new state would be: $S_{1,0,0,0,0,0,0} = <1, 0, 0, 0, 0, 0, 0>$.

Then, transition probabilities would be:

$\mathbf{P^{ai}_{ss'} = p_i}$; where p_i is defined in the question statement, s has task xi as 0 and s' as 1.

And finally, our reward expectations:

$\mathbf{R^{ai}_{ss'} = r_i}$; where r_i is the points of task 'i' defined in the question statement as well.

2.2) First we have to consider that for the student to pass the exam we need a final score of 23, which is 50% of the sum of all the points in the exam, the we need to calculate the probability that the expected return is greater or equal 23, following a certain Policy P. Therefore:

$Pr(V^P(S_{0,0,0,0,0,0,0}) \geq 23)$

Where $V^P(s) = \sum_a p(s,a) \sum_{s'} P^a_{ss2}[R^a_{ss2} + V^p(s2)]$

Note that in this case the gamma factor is one because during this process a final state is always reached, whether it success on all tasks or the number of possible attempts is reached.

2.3) After running the program (because the tree of all possible states and transitions is considerably big to do the calculation by hand) we can see that the expected returns are the following: (Pa: policy 1; Pb: policy B)

$V^{Pa}(S_{0,0,0,0,0,0,0}) = 19,975$
$V^{Pb}(S_{0,0,0,0,0,0,0}) = 22,862$

2.4) As a new policy, we could use "solve the tasks with more points first" an as show in the program this approach has a better expected return, which actually would make the student pass the exam.

$V^{Pc}(S_{0,0,0,0,0,0,0}) = 23,538$

2.5) Suppose that we have a certain number of balls n in a sack, and there are different colors for the balls, let's say red and blue balls. Then, if we pick 1 ball randomly we will have a certain probability of picking up a red ball and another probability of picking up a blue ball, each time we are going to pick up a new ball from the sack, the probability will be affected for the previous ball that we have picked (actions taken). Therefore, the Markov assumption is not justified here. One way to augment the state and justify the Markov assumption is enable replacement of the balls after

every action taken, in this case, the probability of taking a ball of a certain color would be unaltered after every action, so the probability of taking a certain color is not dependent of the past actions and here the Markov Assumption is justified.