Question 2)

Part 1)

Courses:

There are 10,000 entries in the Courses table.

There will be 1 rid per entry.

The size of the data entry will be 10 + 10 = 20 bytes.

Entries in the leaf page = $\frac{2600}{20} = 130\ entries\ per\ leaf\ page.$

$\frac{total\ entries}{entries\ in\ leaf\ page} = \frac{10000}{130} = 76.9 => 76\ leaves$

Intermediate note entry size = cid + ptr = 10+ 6 = 16 bytes per entry.

Entries in intermediate node = $\frac{3000}{16} = 187\ entries\ per\ index\ page$

$\frac{\#\ leaf\ pages}{\#\ entries\ per\ index\ page} = 1 = height.$

Enrollment:

Rid: 10 bytes

Intermediate pointer: 6 bytes

Leaf pages: 65% filled.

Intermediate page: 75% filled

Both have 4k size

Enrolled tree: index : cid

5000 cid for every term

Average number of rid per data entry:

$\frac{250,000}{5000} = 50\ rid's\ on\ average\ per\ cid\ (i.e.\ data\ entry)$

10 * 50 + 10 = 510 bytes, size of every data entry

4000 * 0.65 = 2600

2600 / 510 = 5.098 = 5 data entries per leaf page

$\frac{5000\ entries}{5\ leaves\ per\ page} = 1000\ leaf\ pages$

6 bytes + 10 bytes = 16 bytes for each index node entry size

75% of 4000 bytes are filled, that is 3000 bytes

3000/16 = 187 entries per intermediate node

$$\frac{\#\ leaf\ pages}{\#\ entries\ per\ intermediate\ node} = \left(\frac{1000}{187}\right) = 6\ intermediate\ nodes$$

Height = 2

Part 2)

(a) The average number of rids per data entry will not change as part of the index includes the primary key of the tuples and that means that there will be no possible tuples that can include the similar cid.
(b) The size of the data entry will increase by 10 bytes
(c) The total number of data entries will not change as there are the same number of rids as before.
(d) The number of leaves will increase because the size of each data entry increased while the capacity of a leaf page does not.
(e) The average fanout will decrease because since the number of leaves increased, the number of intermediate nodes must increase as well. However, for each increase in the number of intermediate nodes, there will be capacity for way more leaves and that is why it will most likely outweigh the increase in the number of leaves by a factor that causes the fanout to decrease.