SCENARIO I

Q1.

Option A is that we will need to maintain only one fact table, by augmenting the instructor dimension in a way that it can contain instructor teams separated by "/", which can later be split using an SQL function. In that way, the fact table can have less no. of rows as compared to other options. The weakness of this option is that in case the course is taught by 4 or 5 instructors, then the the team_name column in the instructor Dimension might need to be declared a long length.

Option B is that without augmenting the instructor dimension, maintaining it's grain as per one row per instructor only, the fact table can still point to correct information. Whenever the DWH users want to check which or courdes or instructors are more popular, they can do so by proper aggregation over the enrollmentCount value. The weakness however, is that, if majority of the courses have multiple intructors, the size of the fact table may become abnormally large, due to multiple rows for a single enrollment.

Option C is that facts regarding course popularity and Instructor Popularity have been segregated here, but the shortcoming is that, both the facts will be having some amount of duplicate information (such as student, course etc)

Q2.

Option A, as it would serve the purpose of answering the business queries with minimum cost.

Q3.

No, even if majority of the classes or few classes had multiple instructors, we would choose option A only as it would help in implementing the DWH with minimum changes to the dimensions or facts.

SCENARIO II

Q1.

There's a shortcoming of option A, as if the business users want to know the scores of a customer for a day in past, the customer dimension won't be able to provide it, as it is being overwritten under SCD Type1.

The advantage of option B is that if the business users want to know the scores of a customer for a day in past, the customer dimension will be able to provide it, as it is being maintained under SCD Type2.

The purpose of option C can be implemented using Snowflake schema (as used in option D).

Q2.

As the business users might want to know the historical data regarding a particular customer also, I would choose option B for maintaining the customer Dimension as per SCD Type2.

Q3.

Yes, if the recalculation would be done more frequently, then I would choose option C as in option B, the size of the Customer dimension would increase too much.