✓ **Congratulations! You passed!**
TO PASS 80% or higher

Keep Learning

Techniques: Extract-Transform-Load and Terminology Mapping

LATEST SUBMISSION GRADE

100%

1. Which of the following is an advantage of ELT (extract-load-transform) over ETL (extract-transform-load)? **1 / 1 point**

   ○ The extraction step is easier to program with ELT than with ETL

   ○ ELT scripts are easier to understand and debug than are ETL scripts

   ● ELT is more secure than ETL because data never leaves the source or target database management system

   ○ ELT processes can more easily automated than can ETL processes

   > ✓ **Correct**
   > Correct!

2. Which of the following are challenges in performing structural mapping (all that apply): **1 / 1 point**

   ☑ There is no acceptable table/column in the source database that can be used to obtain values for a table/column in the target database

   > ✓ **Correct**
   > Correct! Happens all the time. The reverse also happens all the time -- there is no table/column in the target for storing a value in the source database

   ☐ The JOINs required to restructure the target database are too complex for the source database to execute

   ☐ The target database does not support the required tables the CDM specifications

   ☑ The source database stores data in a format that cannot be converted to the required format in the target database

   > ✓ **Correct**
   > Correct! Especially for terms that are stored as strings in the source and must be converted to numeric, date or categorical values

3. Which of the following are challenges in performing terminology mapping (all that apply): **1 / 1 point**

   ☐ Writing the SQL statements that perform complex mappings

   ☑ Finding the right terms in the target vocabulary that matches, or closely matches, a term in the source vocabulary

   > ✓ **Correct**
   > Correct! This is the hardest, most tedious part of terminology mapping -- determining when a less-than-perfect match is "good enough"

   ☑ Understanding the meaning of terms in the source database

   > ✓ **Correct**
   > Correct! Often there is no data dictionary that gives the intended meaning of the terms stored in a table/column

   ☐ Understanding the allowed values in the target database

4. One of the most difficult ETL tasks is finding a target term that matches a source term with sufficient precision. Which of the following are challenges with this step: **1 / 1 point**

   ☑ Determining if a target term is close enough to a source term depends on the types of users/queries that will use the CDM

   > ✓ **Correct**
   > Correct!

☑ Selecting a specific target term when there are more than one target term tham may be used

✓ **Correct**
Correct!

☐ Finding the best location in the target CDM to put a mapped term since the mapped term may not be an exact match

☑ Finding the best target term in a large set of terms since the equivalent target term may not use the same text description

✓ **Correct**
Correct!

5. **ETL is an iterative process because (all that apply):**   `1 / 1 point`

☑ End users change the ETL specifications.

✓ **Correct**

☑ Programmers do not fully understand the data, ETL specifications and intended target data model.

✓ **Correct**

☑ Testing reveals cases that are not programmed correctly.

✓ **Correct**

☑ New data anomalies are discovered during the coding process that requires special handling.

✓ **Correct**

6. **White Rabbit lists values found in a table/column and the number of times that value is used in a source database.**   `1 / 1 point`

Which of the following does not use profiling information to inform the ETL process?

○ A data value used in high frequency may indicate a default value.

⦿ The presence of two values which have opposite meanings in roughly the same frequency may indicate confusion of how to use the values.

○ The absence of any missing values may indicate a field that is auto-populated by the application.

○ A high frequency for NULL or empty may indicate a data quality issue.

✓ **Correct**