* We want to define a metric M which assigns a score to a computed plan, based on the ground truth for the file targeted by the plan.
* This metric should reward correct schema matches and transformation instructions while penalising incorrect ones.
* The metric is defined as follows :
  + Let
    - P be the plan to evaluate, the plan acts on file F
    - GT be the ground truth/correct plan for F
    - X and Y be the source and target schemas respectively
    - m1,m2, …, mN be the mappings from X to Y defined in GT
    - t1, t2, …, tN be the transformation steps defined in GT
    - The metric M takes in P and GT and returns a normalised score. Namely, M(P, GT) = Score 🡪[0, 10] where 0 means that the plan is very poor and 10 means that the plan matches the GT completely.
* Score is calculated as follows
  + For all mappings “m” defined in GT
    - If m appears in P and it’s fully correct – total + 1.0
    - If m appears in P but partially correct – total +0.5
    - If m does not appear in P – total – 0.5
    - If P contains mappings not in GT, i.e. hallucinated, total – 1.0
  + For all transformations “t” defined in GT
    - If t appears in P and it’s fully correct – total + 1.0
    - If t appears in P but partially correct – total +0.5
    - If t does not appear in P – total – 0.5
    - If P contains transformations not in GT, i.e. hallucinated, total – 1.0
  + The total is then normalised to be in the range [0, 10]