I. Для определения среднего балла мальчиков и девочек подставим в линейную регрессию соответствующие значения переменной пола:

Мальчики (переменная is_male = 1): grade = 4.16026667 - 0.18740000*1 = 3.97286667
Девочки (переменная is_female = 0): grade = 4.16026667 - 0.18740000*0 = 4.16026667
Определить средний балл по всем ученикам мы не можем, так как нам неизвестны данные о количествах мальчиков и девочек или их соотношении.

II. Grade = a +b*is_female
3.97286667 + b = 4.16026667
Нам нужно подобрать такие коэффициенты a и b, так что grade = a + b*is_female было правильно для среднего балла мальчиков и девочек, рассчитанного в предыдущем пункте (банальная система из двух уравнений). В данном случае в регрессии будет сложение, а не вычитание, так как для девочек, у которых средний балл больше, переменная is_female принимает большее значение.
Заметим, что для мальчиков переменная is_female принимает значение 0, значит для них a = grade = 3.97286667 - свободный член регрессии. Подставляем его в уравнение для девочек: 4.16026667 = 3.97286667 + b*1, откуда находим b = 0.1874. Таким образом, наша регрессия будет иметь вид: grade = 3.97286667 + 0.1874*is_female.

III. Добавление обоих переменных в регрессию бесполезно и не приведет к улучшению объяснения оценок учеников. Это связано с тем, что одна эти переменные зависимы друг от друга (когда одна имеет значение 0, другая имеет значение 1 и наоборот) и введение двух переменных в данном случае не увеличивает количество информации, которой мы располагаем, и соответственно не улучшит уровень предсказания оценки.

IV. В итоге данного действия регрессия также не изменится, ведь как и в прошлом случае, мы не получаем дополнительной информации, а лишь повторяем предыдущую. Таким образом, ни коэффициенты регрессии, ни ее качество не изменятся