b) Your friend tells you their objective is better than yours. Can you justify their claim.

Their method has the potential to generalize to unseen data better, as ε can prevent you from over fitting.

Suppose you want to train a linear model $\hat{y} = ax$ and have the following data.

Train	Test
(1, 1.15)	(1,1)
(2, 2.30)	(2,2)

Training with MSE yields the fitted model : $\hat{y}_{(1)} = 1,15x$

with test error : $(0,15)^2 + (0,3)^2 = 0,1125$

Hawever, if you train with

$\min \tilde{R}(g) = \min |MSE - \varepsilon| + \varepsilon$

and set $\varepsilon = 0,1125$, you may yield the fitted model $y = x$

with test error : $(1-1)^2 + (2-2)^2 = 0$

c) Your friend claims their objective can be even more effective if ε is chosen more carefully. Explain how.

This is the bias-variance tradeoff. Lowering ε increases variance and reduces bias. You can tune epsilom to achieve better model generelization.