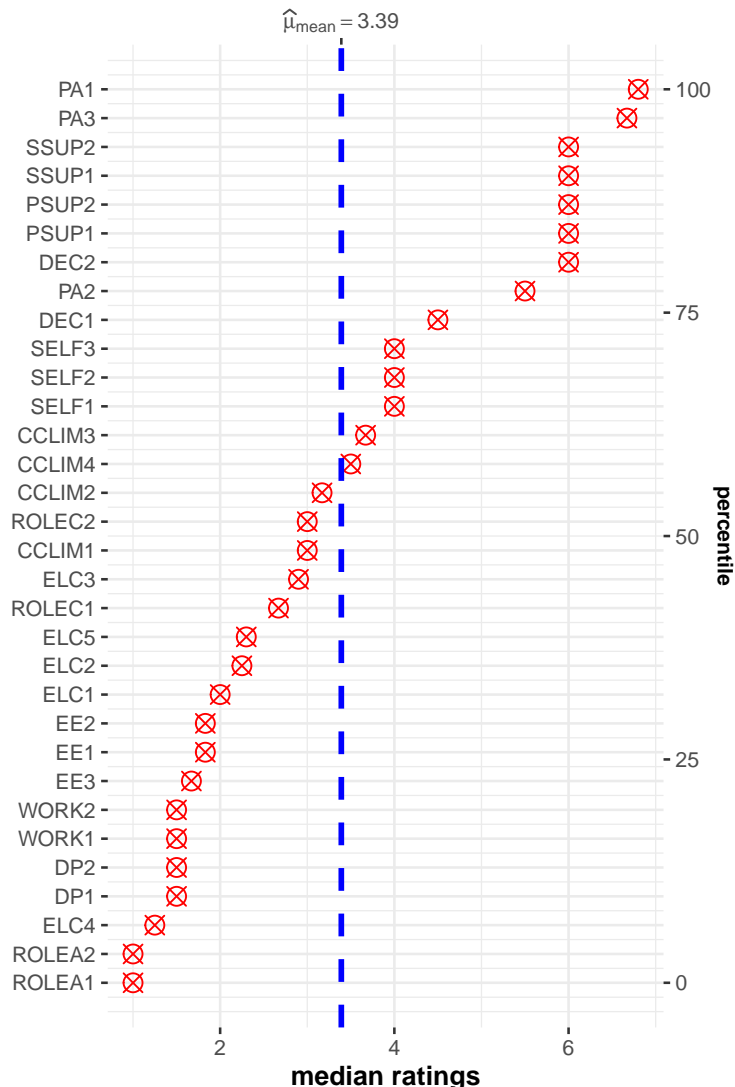


Figure 2 Distributions of median rating for each item

(a) Calibration dataset

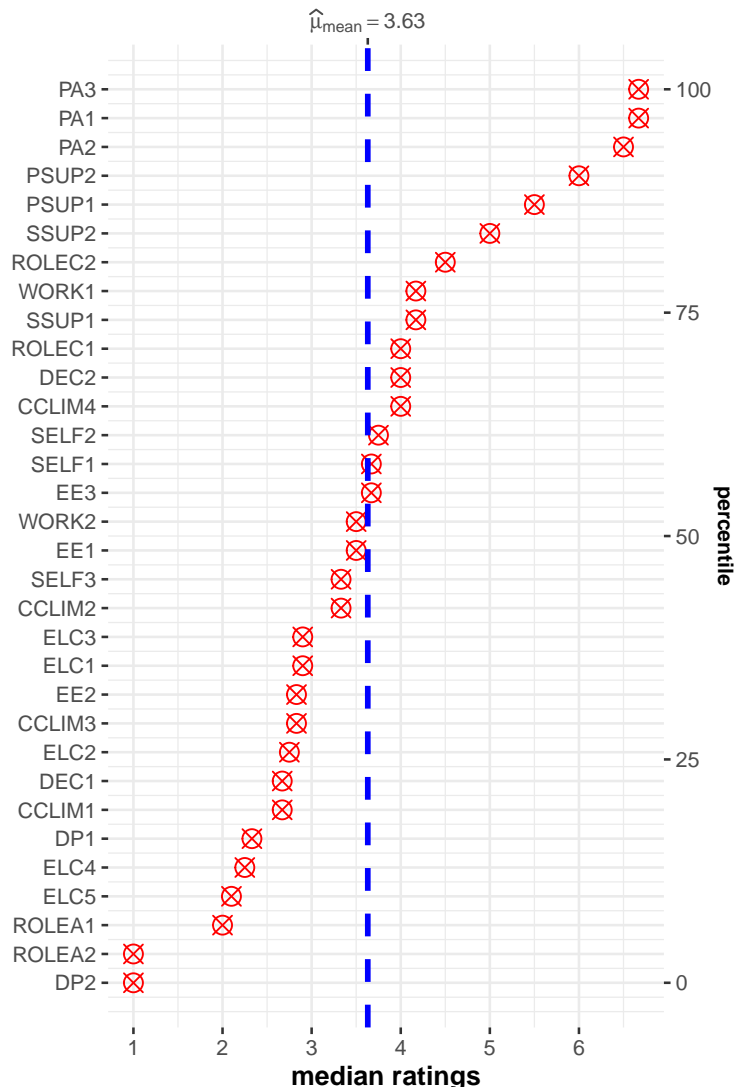
$t_{\text{Student}}(31) = 10.33, p = 1.48\text{e-}11, \hat{g}_{\text{Hedges}} = 1.78, \text{CI}_{95\%}$



$\log_e(\text{BF}_{01}) = -20.24, \hat{\theta}_{\text{difference}}^{\text{posterior}} = 3.33, \text{CI}_{95\%}^{\text{ETI}} [2.64, 4.02], r_{\text{Cauchy}}^{\text{JZS}} = 0.71$

(b) Validation dataset

$t_{\text{Student}}(31) = 13.96, p = 6.52\text{e-}15, \hat{g}_{\text{Hedges}} = 2.41, \text{CI}_{95\%}$



$\log_e(\text{BF}_{01}) = -27.62, \hat{\theta}_{\text{difference}}^{\text{posterior}} = 3.59, \text{CI}_{95\%}^{\text{ETI}} [3.08, 4.14], r_{\text{Cauchy}}^{\text{JZS}} = 0.71$