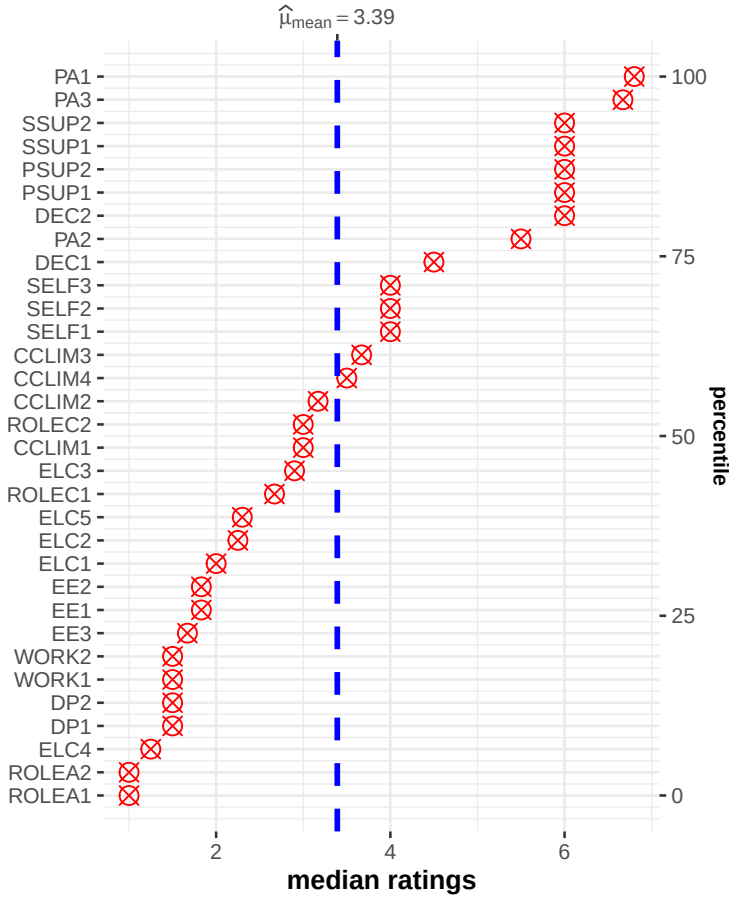
Figure 2 Distributions of median rating for each item

(a) Calibration dataset

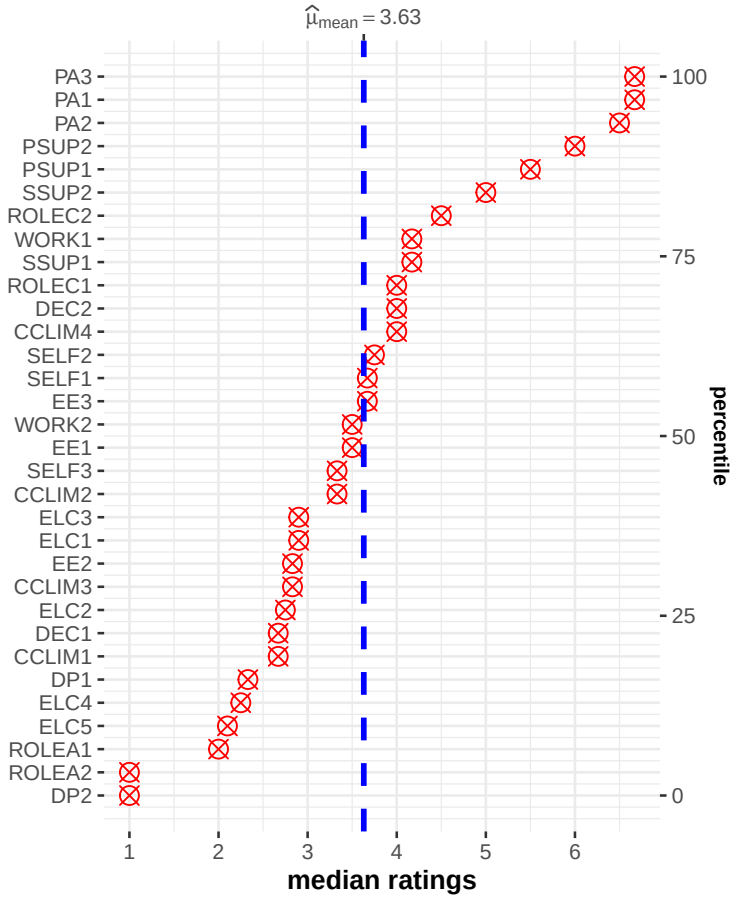
$t_{\text{Student}}(31) = 10.33$, $p = 1.48\text{e}{-11}$, $\widehat{g}_{\text{Hedges}} = 1.78$, $\text{CI}_{95\%}$



$\log_e(\text{BF}_{01}) = -20.24$, $\widehat{\delta}^{\text{posterior}}_{\text{difference}} = 3.32$, $\text{CI}^{\text{ETI}}_{95\%}$ [2.63, 4.04], $r^{\text{JZS}}_{\text{Cauchy}} = 0.71$

(b) Validation dataset

$t_{\text{Student}}(31) = 13.96$, $p = 6.52\text{e}{-15}$, $\widehat{g}_{\text{Hedges}} = 2.41$, $\text{CI}_{95\%}$ [1



$\log_e(\text{BF}_{01}) = -27.62$, $\widehat{\delta}^{\text{posterior}}_{\text{difference}} = 3.59$, $\text{CI}^{\text{ETI}}_{95\%}$ [3.05, 4.14], $r^{\text{JZS}}_{\text{Cauchy}} = 0.71$