

Average numerical rating for each test item

| item | rating | item | rating |
|------|--------|------|--------|
| X1 | 4.7 | X11 | 4.9 |
| X2 | 4.4 | X12 | 4.7 |
| X3 | 4.5 | X13 | 4.3 |
| X4 | 4.9 | X14 | 4.9 |
| X5 | 0.8 | X15 | 0.4 |
| X6 | 0.1 | X16 | 0.5 |
| X7 | 0.1 | X17 | 0.2 |
| X8 | 0.1 | X18 | 0.1 |
| X9 | 4.7 | X19 | 0.3 |
| X10 | 4.4 | X20 | 0.1 |
| | | X68 | 3.0 |

Table 6: Average rating of the baseline items.

| item | rating | item | rating | item | rating |
|------|--------|------|--------|------|--------|
| X21 | 4.5 | X36 | -4.2 | X52 | 0.7 |
| X22 | 3.6 | X37 | 4.4 | X53 | 1.0 |
| X23 | 1.8 | X38 | -1.3 | X54 | 0.4 |
| X24 | 2.8 | X39 | 4.5 | X55 | 0.6 |
| X25 | 2.8 | X40 | -2.9 | X56 | 4.1 |
| X26 | 4.3 | X42 | -2.6 | X57 | 2.2 |
| X27 | 3.9 | X43 | -4.8 | X58 | 4.1 |
| X28 | 2.5 | X44 | -0.7 | X59 | 4.5 |
| X29 | 1.6 | X45 | -3.5 | X60 | 4.6 |
| X30 | -2.4 | X46 | 1.5 | X61 | 5.0 |
| X31 | -3.9 | X47 | 1.5 | X62 | 4.7 |
| X32 | 0.6 | X48 | 2.2 | X63 | 3.2 |
| X33 | -0.1 | X49 | 4.2 | X64 | 4.8 |
| X34 | 0.7 | X50 | 1.0 | X65 | 3.9 |
| X35 | -4.0 | X51 | 0.8 | X67 | 2.6 |

Table 7: Average rating of the critical items.