APPENDIX D

This Appendix contains three figures for the article "Multidimensional Test Assembly Using Mixed-Integer Linear Programming: An Application of Kullback-Leibler Information".

Figure D.1 presents the distribution of the $\overline{\mathrm{MRD}_V}$ and the $\overline{\mathrm{RMSD}_T}$ for the 18 Minimax models in the 20 replications of simulation study 1. The maximum relative error variance gap and the maximum score gap, averaged over the 20 replications are added to the plot as gray bullets connected with a dashed line. Note that for some Minimax models these average gaps fall outside the plotting area. We considered a difference of 0.5 points acceptable on a 30-item test. Likewise, we considered a relative difference in measurement error between the forms below five percent acceptable. Black dashed lines that represent these thresholds were added to the plots as reference lines.

Figure D.2 and Figure D.3 summarize the results of the second and third simulation study, respectively. Simulation 2 applied the strategies to two-dimensional item pools, Simulation 3 applied the strategies to empirically based item pools with items following a three-dimensional bi-factor model.
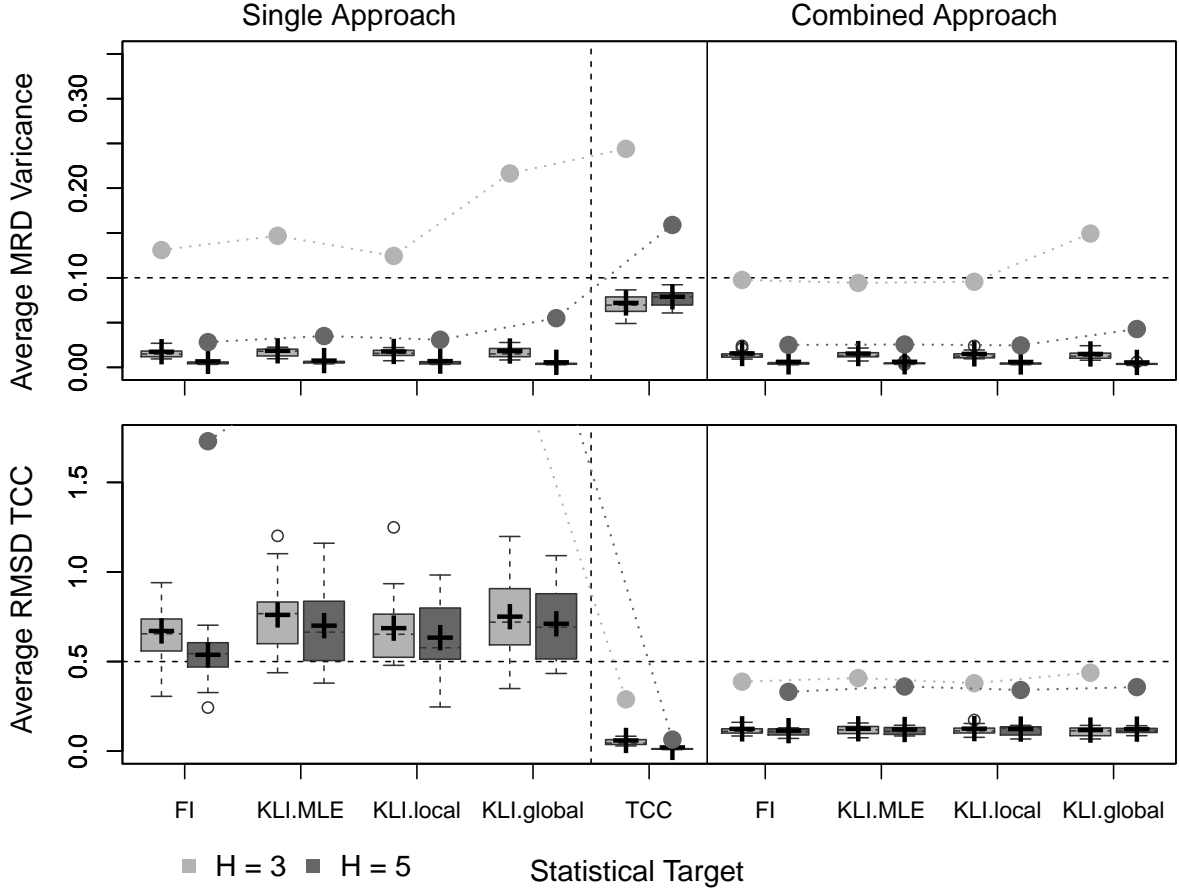
Figure D.1: Distribution of the 20 $\overline{\mathrm{MRD}_V}$ (upper panels) and the $\overline{\mathrm{RMSD}_T}$ (lower panels) for each of the $2 \times 5 \times 2 - 2 = 18$ Minimax models for the simulated unidimensional item pools. The bullets connected with dashed lines represent the average maximum relative error variance gap and the maximum score gap in respectively the upper and the lower panels. The horizontal dashed lines represent a relative measurement error difference of about five percent in the upper panels, and a score difference of 0.5 points in the lower panels.
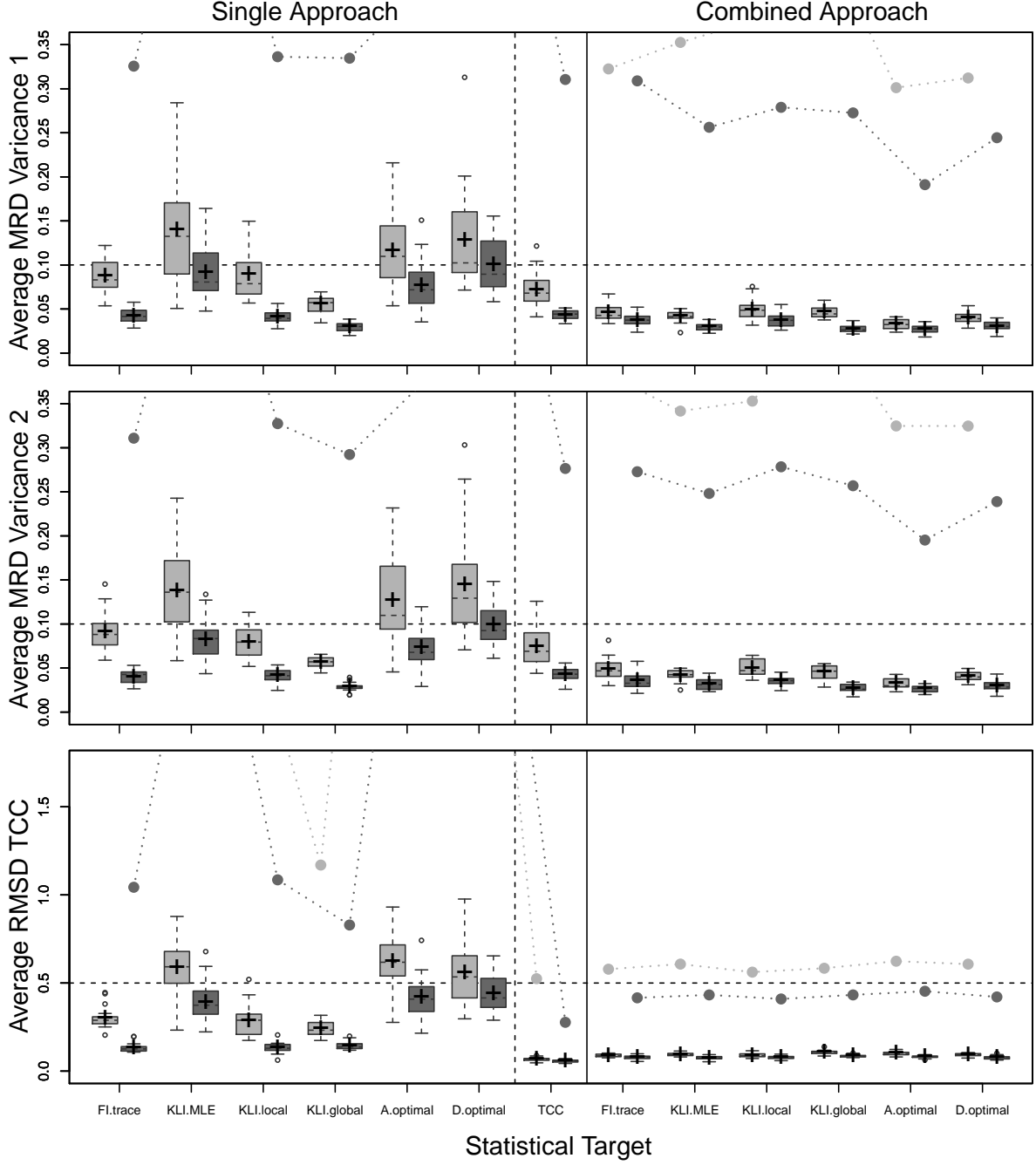
Figure D.2: Distribution of the 20 $\overline{\mathrm{MRD}_{V_1}}$ (upper panels), the $\overline{\mathrm{MRD}_{V_2}}$ (middle panels) and the $\overline{\mathrm{RMSD}_T}$ (lower panels) for each of the $2 \times 7 \times 2 - 2 = 26$ Minimax models for the simulated two-dimensional item pools. The bullets connected with dashed lines represent the average maximum relative error variance gap upper & middle panels, and the maximum score gap in the lower panels. The horizontal dashed lines represent a relative measurement error difference of about five percent in the upper panels, and a score difference of 0.5 points in the lower panels.
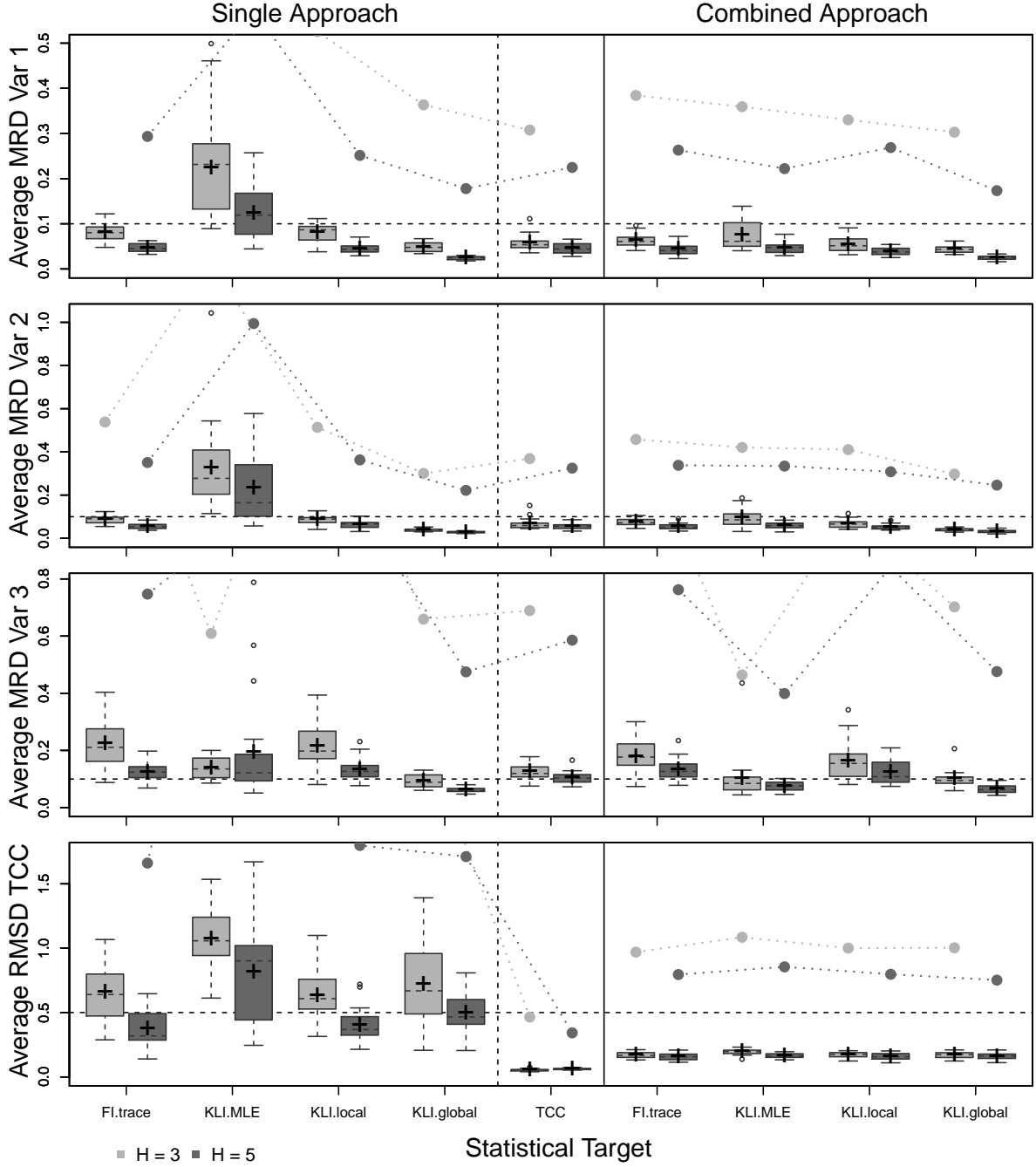
3

Figure D.3: Distribution of the 20 $\overline{\mathrm{MRD}_{V_1}}$ (upper panels, cf. the common dimension), the $\overline{\mathrm{MRD}_{V_2}}$ and $\overline{\mathrm{MRD}_{V_3}}$ (middle panels) and the $\overline{\mathrm{RMSD}_T}$ (lower panels) for each of the $2 \times 5 \times 2 - 2 = 18$ Minimax models for the simulated three-dimensional bifactor item pools. The bullets connected with dashed lines represent the average maximum relative error variance gap upper & middle panels, and the maximum score gap in the lower panels. The horizontal dashed lines represent a relative measurement error difference of about five percent in the upper panels, and a score difference of 0.5 points in the lower panels.

4