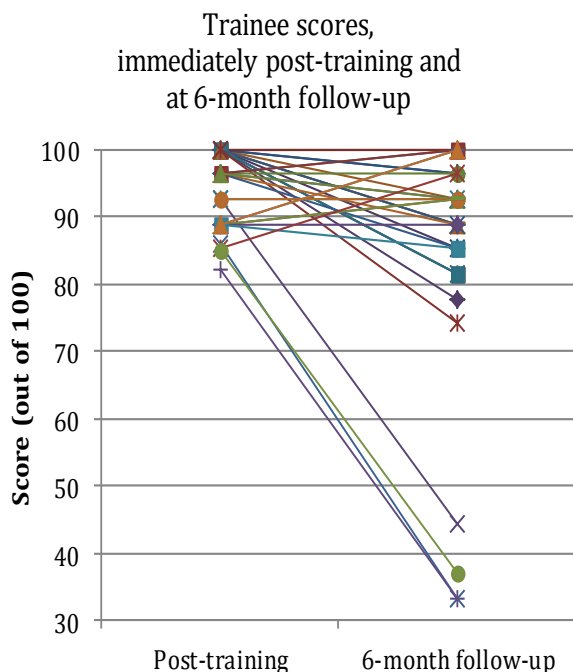Supplemental Digital Figure 1

Item	Residents who correctly performed the item, No. (%)	Residents who omitted or incorrectly performed the item, No. (%)
Checked that all necessary equipment is available and ready to use	181 (76)	48 (24)
Opened lumbar puncture kit carefully to maintain sterility	170 (71)	69 (29)
Prepped insertion point with iodine swabs in concentric circles	147 (61)	92 (39)
[Other checklist items, not shown here]	—	—

Trainee scores, immediately post-training and at 6-month follow-up

Correlation between post-training and 6-month follow-up: $r = 0.44$

If post-training mastery standard is set at...	Then at 6-month follow up...	
	Minimum score	Mean score
100	74.07	88.10
95	74.07	89.95
90	44.44	88.74
85	33.33	86.20
80	33.33	86.20

Supplemental Digital Figure 1 Performance data examples for standard-setting judges in non-mastery and mastery contexts. The *top panel* presents a hypothetical past-performance data

display for a traditional item-based standard setting exercise for lumbar puncture, showing the numbers and percentages of residents who did or did not accomplish each item; data based on

performance of 239 residents. The *bottom panel* presents a hypothetical predictive performance data display for a mastery item-based standard-setting exercise for lumbar puncture; data based on performance of 34 residents.