**Appendix**

**Effect Sizes**

To measure the effects based on differences between means, Cohen’s *D* was calculated for all studies if possible. The formula to do so was (Mean1-Mean2)/ spooled. spooled was calculated using the formula √[(s12+ s22) / 2].

**Celume et al.**

No post-test *SD* was provided, so Cohen’s *D* was calculated using only pre-test *SD*

Theory of Mind

Pre-test mean 17.344, post-test mean 20.98, *SD* based on Pre-test 3.799; Cohen’s *D* 0.957 (Large)

Collaborative behavior

This effect size is not fully accurate, since the pre-test on collaborative behavior was done in the second week after the training started, which might explain the extremely low *SD*.

Pre-test mean 3.12, post-test mean 4.52, *SD* based on Pre-test 0.13; Cohen’s *D* 10.77 (Large)

**Centigöz et al.**

Pre-test mean 190, post-test mean 268.2, *SD* based on pre-test 21.75 and post-test 18.76; Cohen’s *D* 3.85 (Large)

**Cruz et al.**

Scale A, showing emotions with facial expressions

Pre-test mean 3.04, post-test mean 3.55, *SD* based on pre-test 0.40 and post-test 0.45; Cohen’s *D* 1.197 (Large)

Scale B, recognizing emotions based on video

Pre-test mean 15.83, post-test mean 14.65, *SD* based on pre-test 2.88 and post-test 2.42; Cohen’s *D* -.444 (Small)

Scale C, recognizing emotions based on pictures

Pre-test mean 12.82, post-test mean 16.82, *SD* based on pre-test 2.64 and post-test 3.81; Cohen’s *D* 1.22 (Large)

**Freeman et al.**

Self-concept

Pre-test mean 95.8625, post-test mean 103.1525, *SD* based on pre-test 16.4125 and post-test 14.05; Cohen’s *D* .477 (Small)

Problem behavior

Pre-test mean 102.47, post-test mean 100.37, *SD* based on pre-test 13.83 and post-test 11.56; Cohen’s *D* -.165 (Small)

Social skills

Pre-test mean 96.11, post-test mean 100.76, *SD* based on pre-test 17.47 and post-test 17.39; Cohen’s *D* .267 (Small)

**Joronen et al.**

Social relationships

Pre-test mean 72.26, post-test mean 81.57, *SD* based on pre-test 11.86 and post-test 12.49; Cohen’s *D* .764 (Medium)

No *SD* was given for bullying, therefore no Cohen’s *D* could be calculated

**Moore et al.**

Since the pre-test results were not published, these calculations are based on comparing the mean of one of the experiment groups, either the imagination group or the affect group, to the mean of the control group.

Imagination

Control mean 4.374, imagination mean 7.107, *SD* based on control 3.741 and imagination 4.291; Cohen’s *D* .679 (Medium)

Affect

Control mean 4.374, affect mean 5.836, *SD* based on control 3.741 and affect 3.543; Cohen’s *D* .401 (Small)

**Oztug & Ciner**

Pre-test mean 218,64, post-test mean 267.54, *SD* based on pre-test 50 and post-test 32.32; Cohen’s *D* 1.162 (Large)

**Pekdoğan**

Pre-test mean 144, post-test mean 225.96, *SD* based on pre-test 58 and post-test 59; Cohen’s *D* 1.401 (Large)

**Tsiarias**

Popular

Pre-test mean 2.914, post-test mean 3.377, *SD* based on pre-test .549 and post-test .537; Cohen’s *D* .853 (Large)

Rejected

Pre-test mean 2.844, post-test mean 2.333, *SD* based on pre-test .564 and post-test .477; Cohen’s *D* -.979 (Large)

Neglected

Pre-test mean 3.243, post-test mean 2.779, *SD* based on pre-test .521 and post-test .587; Cohen’s *D* -.836 (Large)

Controversial

Pre-test mean 1.446, post-test mean 1.620, *SD* based on pre-test .507 and post-test .611; Cohen’s *D* .310 (Small)

Average

Pre-test mean 9.777, post-test mean 9.998, *SD* based on pre-test .823 and post-test .947; Cohen’s *D* .249 (Smal)

**Wright**

No means or *SD*’s were given, therefore Cohen’s *D* could not be calculated.