Appendix

Appendix A1.1 Study characteristics: Sallows & Graupner, 2005

Characteristic	Description
Study citation	Sallows, G. O., & Graupner, T. D. (2005). Intensive behavioral treatment for children with autism: Four-year outcome and predictors. *American Journal on Mental Retardation, 110*(6), 417–438.
Participants	Children were recruited for the study through local birth to age 3 special education programs. Eligible children were (1) age 24 to 42 months at intake, (2) had a mental age divided by chronological age of the Mental Development Index equal to or greater than 35, (3) neurologically within "normal" limits, and (4) diagnosed with autism by an independent child psychiatrist. The parents of all eligible children agreed to participate. Thirteen children entered the study in 1996 and 11 in 1997. Twenty-four children were matched on pretest measures of IQ and randomly assigned to the treatment or comparison condition. One child dropped out of the study,[1] resulting in an analytic sample of 23 children (13 in the treatment group and 10 in the comparison group).
Setting	Madison, Wisconsin
Intervention	The program was based on the UCLA *Lovaas Model*, which initially focuses on one-on-one discrete trials and progresses from simpler to more complex skills. The therapist would engage the child in favorite activities, providing brief task instruction, such as "sit down" or making a request. Reinforcements, such as edibles or physical play, were given after each trial; in between, children were encouraged to generalize the lessons into more natural settings and develop social responsiveness. The program was intended to be 40 hours a week of direct treatment, although the averages for years 1 and 2 were 39 and 37 hours a week, respectively. The hours of treatment declined in subsequent years as the children began school. Children received 6 to 10 hours per week of in-home supervision from a senior therapist and weekly consultations with the senior author or clinic supervisor, during which the senior author/clinic supervisor observed the child and recommended appropriate changes to the program.
Comparison	Children in the comparison group received a parent-directed intervention consistent with the UCLA *Lovaas Model*. The parents in this group selected how many weekly treatment hours their children received from the therapist, averaging 32 hours in year 1 and 31 hours in year 2, with the exception of one family that chose to have 14 hours of treatment. The children received 6 hours per month of in-home supervision from a senior therapist and consultations every two months with the senior author or clinic supervisor, during which the senior author/clinic supervisor observed the child and recommended appropriate changes to the program.
Primary outcomes and measurement[2]	Communication/language competencies, social-emotional development and behavior, and functional abilities were assessed with the Vineland Adaptive Behavior Scales and the Autism Diagnostic Interview–Revised. Functional abilities also were assessed with the Vineland Adaptive Behavior Scales. For a more detailed description of these outcome measures, see Appendices A2.2–A2.4.
Staff/teacher training	Therapists had completed at least one year of college and attended 30 hours of training, at least 10 of which were one-on-one training and feedback while working with their assigned child. Therapists attended weekly or biweekly team meetings. Senior therapists had a minimum of a four-year degree, one year of experience as a therapist with two or more children, and a 16-week internship at the UCLA facility.

1. The study does not report the child's treatment condition.
2. Sallows and Graupner (2005) reported additional outcomes on IQ and language not included in this report. These outcomes combined different measures and were excluded because they did not meet the topic area's requirements for combined outcome measures, as outlined in the protocol. A composite measure of the Vineland Adaptive Behavior Scales also was not reported because the measure included results from different domains. The separate subtests are included in the report.

Appendix A1.2 Study Characteristics: Smith, Groen, & Wynn, 2000

Characteristic	Description
Study citation	Smith, T., Groen, A. D., & Wynn, J. W. (2000). Randomized trial of intensive early intervention for children with pervasive developmental disorder. *American Journal on Mental Retardation, 105*(4), 269–285.
Participants	For this study, all referrals to the ULCA Young Autism Project were eligible if the following criteria were met: (1) referred to the treatment center between 1989 and 1992, (2) chronological age between 18 and 42 months, (3) residence within a one-hour drive of treatment center, (4) IQ ratio between 35 and 75, (5) diagnosis of autism or pervasive developmental disorder not otherwise specified, and (6) no other major medical problems. Of those determined to meet the criteria, one family declined, and eight others were determined not to be eligible. Children were divided into two strata: those with autism and those with other diagnoses. Within the strata, 32 children were divided into pairs based on IQ and randomly assigned to the treatment or comparison condition. After random assignment, four children with mental retardation and no pervasive developmental disorder were excluded. The final sample included 28 children (15 in the intervention group and 13 in the comparison group).
Setting	Participants resided within a one-hour drive to the research site, located at the University of California–Los Angeles (UCLA) Young Autism Project.
Intervention	The intervention for the treatment group was based on Lovaas's behavior treatment model. The program was implemented with four to six student therapists, supervised by the study authors. Treatment occurred in the children's homes for approximately 30 hours a week for two to three years. Participants' primary caregivers were asked to conduct five hours per week of treatment alongside a student therapist for the first three months of treatment. In the beginning of the treatment, student therapists largely used one-to-one discrete trials, which involve a cue for behavior, guidance on the appropriate response, and reinforcement. As children mastered skills, the treatment progressed from simple tasks to more complex skills. Once the children entered school, student therapists served as aides to help the children adjust to classrooms, and the one-to-one treatment decreased as children progressed.
Comparison	Parents of children in the comparison group received training to use treatment approaches described in the Lovaas et al. (1981)[1] manual. Parents received in-home training in two weekly sessions, totaling five hours per week, for three to nine months. The first author of the study met with parents at the first and last training sessions, and every three months in between. In the first session, parents were asked to identify three goals for the child, and the first author and parent trainer created a plan using the Lovaas approach to reach those goals. In subsequent sessions, the trainer would demonstrate the techniques, observe the parent, and provide feedback.
Primary outcomes and measurement[2]	Cognitive development was assessed with a measure of intellectual functioning, or IQ. For a more detailed description of this outcome measure, see Appendix A2.1.
Staff/teacher training	Training for student therapists for the intervention group was not explicitly stated. Trainers for the comparison group had a minimum of one year of experience at the UCLA Young Autism Project, which included at least six months instructing novice therapists. The trainers for the comparison group received one hour per week of individual supervision from the first author, with additional supervision as needed. The study authors had a combined experience of 10 years at the UCLA Young Autism Project under Lovaas's supervision.

1. Lovaas, O. I., Ackerman, A. B., Alexander, D., Firestone, P., Perkins, J., & Young, D. (1981). *Teaching developmentally disabled children: The ME book*. Austin, TX: Pro-Ed, Inc.
2. Smith et al. (2000) reported additional outcomes not included in this report. These outcomes were excluded because they did not meet the WWC requirements for baseline equivalence (that is, within 0.25 standard deviations at baseline) or required a statistical adjustment, which was not included by the authors.

Appendix A2.1 Outcome measures for the cognitive development domain

Outcome measure	Description
IQ	In the Smith et al. (2000) study, the Stanford-Binet Intelligence Scale was administered to all children. If a basal could not be established, children were given the Bayley Scales of Infant Development–Mental Development Index, which is appropriate for children with developmental levels of 0 to 30 months. The Bayley was scored as a ratio IQ and the Stanford-Binet as a deviation IQ (Smith et al., 2000). At baseline, no children established a basal on the Stanford-Binet. At follow-up, four children in the treatment group and five in the comparison group did not establish a basal on the Stanford-Binet.

Appendix A2.2 Outcome measures for the communication/language competencies domain

Outcome measure	Description
Vineland: Communication subscale	A subscale of the Vineland Adaptive Behavior Scales, which has psychometric data for preschool age children (Sallows & Graupner, 2005). No other information is provided.
Autism Diagnostic Interview–Revised: Communication subscale	Subscale of the Autism Diagnostic Interview–Revised. No other information is provided.

Appendix A2.3 Outcome measures for the social-emotional development and behavior domain

Outcome measure	Description
Vineland: Social Skills subscale	A subscale of the Vineland Adaptive Behavior Scales, which has psychometric data for preschool age children (Sallows & Graupner, 2005). No other information is provided.
Autism Diagnostic Interview–Revised: Social Skills subscale	Subscale of the Autism Diagnostic Interview–Revised. No other information is provided.
Autism Diagnostic Interview–Revised: Ritual subscale	Subscale of the Autism Diagnostic Interview–Revised. No other information is provided.

Appendix A2.4 Outcome measures for the functional abilities domain

Outcome measure	Description
Vineland: Daily Living Skills subscale	A subscale of the Vineland Adaptive Behavior Scales, which has psychometric data for preschool age children (Sallows & Graupner, 2005). No other information is provided.

Appendix A3.1 Summary of study findings included in the rating for the cognitive development domain[1]

Outcome measure	Study sample	Sample size (children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Lovaas Model* group	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[3] (*Lovaas Model* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Smith, Groen, & Wynn, 2000[7]								
IQ	Full sample	28	66.65 (24.08)	49.67 (19.74)	16.98	+0.74	ns	+27
Domain average for cognitive development[8]						**+0.74**	**na**	**+27**

ns = not statistically significant
na = not applicable

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the cognitive development domain.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. For Smith, Groen, and Wynn (2000), the treatment mean is the sum of the comparison mean and the adjusted mean difference, which accounts for pretest differences.
4. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results for the intervention group.
7. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Smith, Groen, and Wynn (2000), no corrections for clustering or multiple comparisons were needed. However, the authors used a one-tailed test, whereas the WWC uses two-tailed tests, so the significance levels may differ from those reported in the original study.
8. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A3.2 Summary of study findings included in the rating for the communication/language competencies domain[1]

Outcome measure	Study sample	Sample size (children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Lovaas Model* group	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[3] (*Lovaas Model* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Sallows & Graupner, 2005[7]								
Vineland: Communication	Full sample	23	79.43 (32.32)	81.40 (24.33)	–1.97	–0.07	ns	–3
ADI-R: Communication	Full sample	23	8.13 (6.91)	8.80 (7.43)	–0.67	–0.09	ns	–4
Domain average for communication/language competencies[8]						**–0.08**	**na**	**–3**

ADI-R = Autism Diagnostic Interview–Revised
ns = not statistically significant
na = not applicable

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the communication/language competencies domain.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. For Sallows and Graupner (2005), the treatment mean is the sum of the comparison mean and the adjusted mean difference, which accounts for pretest differences.
4. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results for the intervention group.
7. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Sallows and Graupner (2005), no corrections for clustering or multiple comparisons were needed; the latter because no impacts were statistically significant.
8. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A3.3 Summary of study findings included in the rating for the social-emotional development/behavior domain[1]

Outcome measure	Study sample	Sample size (children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Lovaas Model* group	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[3] (*Lovaas Model* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Sallows & Graupner, 2005[7]								
Vineland: Social Skills	Full sample	23	75.84 (23.49)	68.90 (10.11)	6.94	0.35	ns	+14
ADI-R: Social Skills	Full sample	23	13.69 (10.58)	13.10 (9.42)	0.59	0.06	ns	+2
ADI-R: Ritual	Full sample	23	6.10 (3.75)	5.60 (3.50)	0.50	0.13	ns	+5
Domain average for social-emotional development/behavior[8]						**0.18**	**na**	**+7**

ADI-R = Autism Diagnostic Interview–Revised
ns = not statistically significant
na = not applicable

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the social-emotional development/behavior domain.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. For Sallows and Graupner (2005), the treatment mean is the sum of the comparison mean and the adjusted mean difference, which accounts for pretest differences.
4. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results for the intervention group.
7. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Sallows and Graupner (2005), no corrections for clustering or multiple comparisons were needed; the latter because no impacts were statistically significant.
8. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A3.4 Summary of study findings included in the rating for the functional abilities domain[1]

Outcome measure	Study sample	Sample size (children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *Lovaas Model* group	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[3] (*Lovaas Model* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Sallows & Graupner, 2005								
Vineland: Daily Living Skills	Full sample	23	66.51 (25.95)	64.20 (12.42)	2.31	0.10	ns	+4
Domain average for functional abilities[8]						**0.10**	**na**	**+4**

ns = not statistically significant

na = not applicable

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the functional abilities domain.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. For Sallows and Graupner (2005), the treatment mean is the sum of the comparison mean and the adjusted mean difference, which accounts for pretest differences.
4. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results for the intervention group.
7. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Sallows and Graupner (2005), no corrections for clustering or multiple comparisons were needed.
8. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A4.1 *Lovaas Model* rating for the cognitive development domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of cognitive development, the WWC rated the *Lovaas Model* as having potentially positive effects for preschool children with disabilities. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, negative effects) were not considered, as the *Lovaas Model* was assigned the highest applicable rating.

Rating received

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

  Met. The one study that met WWC evidence standards with reservations showed a substantively important positive effect.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

  Met. No studies showed a statistically significant or substantively important negative effect.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a *strong* design.

  Not met. Only one study met WWC evidence standards with reservations.

AND

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

  Met. The one study that met WWC evidence standards with reservations did not show statistically significant or substantively important negative effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

Appendix A4.2 *Lovaas Model* rating for the communication/language competencies domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of communication/language competencies, the WWC rated the *Lovaas Model* as having no discernible effects for children with disabilities.

Rating received

No discernible effects: No affirmative evidence of effects.

- Criterion 1: No studies showing a statistically significant or substantively important effect, either *positive* or *negative*.

  Met. The one study that measured communication/language competencies did not show a statistically significant or substantively important effect.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a *strong* design.

  Not met. No studies showed statistically significant positive effects.

AND

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

  Met. No studies showed statistically significant or substantively important negative effects.

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

  Not met. No studies showed a statistically significant or substantively important positive effect.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

  Met. No studies showed a statistically significant or substantively important negative effect.

Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

  Not met. No studies showed a statistically significant or substantively important positive effect.

OR

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

  Not met. No studies showed a statistically significant or substantively important effect.

(continued)

Appendix A4.2 *Lovaas Model* rating for the communication/language competencies domain *(continued)*

Potentially negative effects: Evidence of a negative effect with no overriding contrary evidence.

- Criterion 1: One study showing a statistically significant or substantively important *negative* effect and no studies showing a statistically significant or substantively important *positive* effect.

  Not met. No studies showed a statistically significant or substantively important negative effect.

OR

- Criterion 2: Two or more studies showing statistically significant or substantively important *negative* effects, at least one study showing a statistically significant or substantively important *positive* effect, and more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

  Not met. No studies showed a statistically significant or substantively important positive effect.

Negative effects: Strong evidence of a negative effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *negative* effects, at least one of which met WWC evidence standards for a *strong* design.

  Not met. No studies showed a statistically significant negative effect.

AND

- Criterion 2: No studies showing statistically significant or substantively important *positive* effects.

  Met. No studies showed a statistically significant positive effect.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

Appendix A4.3 *Lovaas Model* rating for the social-emotional development/behavior domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of social-emotional development/behavior, the WWC rated the *Lovaas Model* as having no discernible effects for children with disabilities.

Rating received

No discernible effects: No affirmative evidence of effects.

- Criterion 1: No studies showing a statistically significant or substantively important effect, either *positive* or *negative*.

  Met. The one study that measured social-emotional development/behavior did not show a statistically significant or substantively important effect.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a *strong* design.

  Not met. No studies showed statistically significant positive effects.

AND

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

  Met. No studies showed statistically significant or substantively important negative effects.

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

  Not met. No studies showed a statistically significant or substantively important positive effect.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

  Met. No studies showed a statistically significant or substantively important negative effect.

Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

  Not met. No studies showed a statistically significant or substantively important positive effect.

OR

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

  Not met. No studies showed a statistically significant or substantively important effect.

(continued)

Appendix A4.3 *Lovaas Model* rating for the social-emotional development/behavior domain *(continued)*

Potentially negative effects: Evidence of a negative effect with no overriding contrary evidence.

- Criterion 1: One study showing a statistically significant or substantively important *negative* effect and no studies showing a statistically significant or substantively important *positive* effect.

  Not met. No studies showed a statistically significant or substantively important negative effect.

OR

- Criterion 2: Two or more studies showing statistically significant or substantively important *negative* effects, at least one study showing a statistically significant or substantively important *positive* effect, and more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

  Not met. No studies showed a statistically significant or substantively important positive effect.

Negative effects: Strong evidence of a negative effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *negative* effects, at least one of which met WWC evidence standards for a *strong* design.

  Not met. No studies showed a statistically significant negative effect.

AND

- Criterion 2: No studies showing statistically significant or substantively important *positive* effects.

  Met. No studies showed a statistically significant positive effect.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

Appendix A4.4 *Lovaas Model* rating for the functional abilities domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of functional abilities, the WWC rated *Lovaas Model* as having no discernible effects for children with disabilities.

Rating received
No discernible effects: No affirmative evidence of effects.
• Criterion 1: No studies showing a statistically significant or substantively important effect, either *positive* or *negative*.
Met. The one study that measured functional abilities did not show a statistically significant or substantively important effect.
Other ratings considered
Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.
• Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a *strong* design.
Not met. No studies showed statistically significant positive effects.
AND
• Criterion 2: No studies showing statistically significant or substantively important *negative* effects.
Met. No studies showed statistically significant or substantively important negative effects.
Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.
• Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.
Not met. No studies showed a statistically significant or substantively important positive effect.
AND
• Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.
Met. No studies showed a statistically significant or substantively important negative effect.
Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.
• Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.
Not met. Only one study examined functional abilities.
OR
• Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.
Not met. Only one study examined functional abilities, and it did not show a statistically significant or substantively important effect.

(continued)

Appendix A4.4 *Lovaas Model* rating for the functional abilities domain *(continued)*

Potentially negative effects: Evidence of a negative effect with no overriding contrary evidence.

- Criterion 1: One study showing a statistically significant or substantively important *negative* effect and no studies showing a statistically significant or substantively important *positive* effect.

  Not met. No studies showed a statistically significant or substantively important negative effect.

OR

- Criterion 2: Two or more studies showing statistically significant or substantively important *negative* effects, at least one study showing a statistically significant or substantively important *positive* effect, and more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

  Not met. No studies showed a statistically significant or substantively important positive effect.

Negative effects: Strong evidence of a negative effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *negative* effects, at least one of which met WWC evidence standards for a *strong* design.

  Not met. No studies showed a statistically significant negative effect.

AND

- Criterion 2: No studies showing statistically significant or substantively important *positive* effects.

  Met. No studies showed statistically significant or substantively important positive effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

Appendix A5 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence[1]
		Centers	Children	
Cognitive development	1	1	28	Small
Communication/language competencies	1	1	23	Small
Literacy	0	na	na	na
Math competencies	0	na	na	na
Social-emotional development/behavior	1	1	23	Small
Functional abilities	1	1	23	Small
Physical well-being	0	na	na	na

na = not applicable/not studied

1. A rating of "medium to large" requires at least two studies and two schools across studies in one domain and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is "small." For more details on the extent of evidence categorization, see the WWC Procedures and Standards Handbook, Appendix G.