Risk of Bias Assessment for Included Quantitative Intervention Studies

	Selection bias and confounding	Performance bias: Assessment, spillovers, and contamination	Outcome and analysis reporting biases	Other biases
Adrogue & Orlicki (2013)	**High risk of selection bias and confounding** This study uses a nonexperimental difference-in-difference design to determine the effect of the program on school-level early grade reading outcomes. We rated the study as high risk of selection bias because the study considers outcome variables 4 months after the start of the study as baseline values. However, these values may well have been affected by the program at that time. In addition, it is unclear whether the comparison group was similar to the beneficiary schools at the time the intervention began.	**Medium risk of performance bias** The authors do not account for the possibility of crossover effects. Students in the comparison schools may have switched to treatment schools because of the school feeding program. This behavioral change may result in spillovers to the comparison group.	**High risk of outcome and analysis reporting bias** The authors use an unusual difference-in-difference approach in which outcome measures after the start of the intervention are used as baseline values. This approach can result in considerable bias.	**Low risk of other biases** There is no evidence for other significant risks of bias.
Arbour et al. (2016)	**High risk of selection bias and confounding** The students self-select into the program, which significantly increases the risk of selection bias.	**Low risk of performance bias** The analysis compares students in beneficiary schools with students in control schools that appear to be sufficiently isolated from the beneficiary schools to prevent spillovers and contamination.	**Low risk of outcome and analysis reporting bias** There are no significant outcome and analysis reporting biases.	**Low risk of other biases** There is no evidence for other significant risks of bias.

	Selection bias and confounding	Performance bias: Assessment, spillovers, and contamination	Outcome and analysis reporting biases	Other biases
Bando (2010)	**Medium risk of selection bias and confounding** This study uses a regression analysis that includes school and state-year fixed effects to determine the effect of a school governance program on early grade reading outcomes. Although this method does not fully account for the risk of selection bias, the risk of selection bias is only medium.	**Low risk of performance bias** The analysis compares beneficiary schools with comparison schools that appear to be sufficiently isolated from the beneficiary schools.	**Low risk of outcome and analysis reporting bias.** There are no significant outcome and analysis reporting biases. The study uses a number of robustness checks to assess the validity of the results.	**Low risk of other biases.** There is no evidence for other significant risks of bias.
Barrera-Osorio & Linden (2009)	**Low risk of selection bias and confounding** This study uses a cluster-randomized controlled trial (RCT) to determine the impact of the distribution of computers on early grade reading outcomes. Although attrition was high, the authors were able to credibly account for this in the analysis. Thus, the risk of selection bias and confounding was low.	**Low risk of performance bias** The analysis compares beneficiary schools with comparison schools that appear to be sufficiently isolated from the beneficiary schools to prevent performance bias.	**Low risk of outcome and analysis reporting bias** There are no significant outcome and analysis reporting biases.	**Low risk of other biases** There is no evidence for other significant risks of bias.

	Selection bias and confounding	Performance bias: Assessment, spillovers, and contamination	Outcome and analysis reporting biases	Other biases
Beuermann et al. (2015)	**Low risk of selection bias** The study uses a cluster-RCT to determine the impact of the distribution of laptops to children on early grade reading outcomes. There are no major concerns about selection bias.	**Low risk of performance bias** The study uses a credible social network analysis to determine the spillover effects of the program.	**Low risk of outcome and analysis reporting bias** There are no significant outcome and analysis reporting biases.	**Low risk of other biases** There is no evidence for other significant risks of bias.
Bowne (2014)	**High risk of selection bias and confounding** The study uses a cluster-RCT to determine the impact of the program. However, the main research question focuses on the impact of teachers' language and vocabulary instruction and children's vocabulary growth. The researchers do not account for endogeneity in this relationship.	**Low risk of performance bias** The analysis compares children in beneficiary schools with children in control schools that appear to be sufficiently isolated from the beneficiary schools to prevent performance bias.	**High risk of outcome and analysis reporting bias** The researchers do not use instrumental variable regression analysis to account for endogeneity. Instead, they rely on unusual methods in which they use intermediate outcomes as explanatory variables.	**Low risk of other biases** There is no evidence for significant other risks of bias.
Bowne et al. (2016)	**Medium risk of selection bias and confounding** The study uses a clustered RCT to determine the impact of the program. However, the sample size is too low to accurately measure impact on low-frequency behaviors being analyzed.	**Low risk of performance bias** The analysis compares beneficiary schools with comparison schools that appear to be sufficiently isolated from the beneficiary schools to prevent performance bias.	**Low risk of outcome and analysis reporting bias** There are no significant outcome and analysis reporting biases.	**Low risk of other biases** There is no evidence for other significant risks of bias.

	Selection bias and confounding	Performance bias: Assessment, spillovers, and contamination	Outcome and analysis reporting biases	Other biases
Campos et al. (2011)	**High risk of selection bias and confounding** The study uses hierarchical regression analysis to determine the impact of participation in preschool on early grade reading outcomes. This methodology is not a well-established method to account for selection bias.	**Low risk of performance bias** The analysis compares beneficiary schools with comparison schools that appear to be sufficiently isolated from the beneficiary schools.	**High risk of outcome and analysis reporting bias** The study reports only statistically significant effects in the tables. However, the narrative suggests that not all results were statistically significant. This is an indication of outcome and analysis reporting bias.	**Medium risk of other biases** The study does not account for clustering in the estimation of standard errors.
Cardoso-Martins et al. (2011) Experiment 1	**Medium risk of selection bias** The study uses an RCT to determine the impact of the program on early grade reading outcomes. However, the sample consisted of only 32 students. This sample size is insufficient to ensure equivalence in observable and unobservable characteristics.	**High risk of performance bias** The study used randomization at the student level within the same school. there is a lot of interaction between beneficiary and control students. This interaction creates a major risk of performance bias.	**Low risk of outcome and analysis reporting bias** There are no significant outcome and analysis reporting biases.	**Low risk of other biases** There is no evidence for other significant risks of bias.
Cardoso-Martins et al. (2011) Experiment 2	**Medium risk of selection bias** The study uses an RCT to determine the impact of the program on early grade reading outcomes. However, the sample consisted of only 20 students. This sample size is insufficient to ensure equivalence in observable and unobservable characteristics.	**High risk of performance bias** The study used randomization at the student-level within the same school. Thus, there is a lot of interaction between beneficiary and control students. This interaction creates a major risk of performance bias.	**Low risk of outcome and analysis reporting bias** There are no significant outcome and analysis reporting biases.	**Low risk of other biases** There is no evidence for other significant risks of bias.

	Selection bias and confounding	Performance bias: Assessment, spillovers, and contamination	Outcome and analysis reporting biases	Other biases
Cristia et al. (2012)	**Low risk of selection bias** The study uses a cluster-RCT to determine the impact of the distribution of laptops to children on early grade reading outcomes. There are no major concerns about selection bias.	**Low risk of performance bias** The analysis compares beneficiary schools with control schools that appear to be sufficiently isolated from the beneficiary schools.	**Low risk of outcome and analysis reporting bias** There are no significant outcome and analysis reporting biases.	**Low risk of other biases** There is no evidence for other significant risks of bias.
De Felicio, Terra, & Zoghbi (2012)	**Medium risk of selection bias** The study uses a propensity score matching design to assess the effects of participation in preschool on early grade reading outcomes. This design enables the researchers to correct for selection bias from observable characteristics. However, the selection bias is still medium because the methodology does not allow the researchers to account for unobservable characteristics from self-selection into preschool.	**Medium risk of performance bias** The study compares beneficiaries with nonbeneficiaries in the same municipality. Thus, there is a risk of interaction between beneficiary and comparison students, which we interpret as a medium risk of performance bias.	**High risk of outcome and analysis reporting bias** The authors fail to report statistically insignificant effects. However, the narrative indicates that the results are not statistically significant in all specifications. This discrepancy in reporting indicates a high risk of outcome and analysis reporting bias.	**Low risk of other biases** There is no evidence for other significant risks of bias.

	Selection bias and confounding	Performance bias: Assessment, spillovers, and contamination	Outcome and analysis reporting biases	Other biases
Ferrando et al. (2011)	**Medium risk of selection bias** The study uses a propensity score matching design to assess the effects of the distribution of laptops to children on early grade reading outcomes. This design enables the researchers to correct for selection bias from observable characteristics. However, the selection bias is still medium because the methodology does not allow the researchers to account for unobservable characteristics.	**Medium risk of performance bias** The analysis compares beneficiary schools with comparison schools that appear to be sufficiently isolated from the beneficiary schools.	**Medium risk of outcome and analysis reporting bias** The study uses only a subset of available control variables for the propensity score matching. It is unclear why the authors do not include the other potential control variables. This approach may be an indication of outcome and analysis reporting bias.	**Medium risk of other biases** The study does not account for clustering in the estimation of standard errors.
Gomez Franco (2014)	**High risk of selection bias** The study uses a cluster-RCT to determine the impact of the program on early grade reading outcomes. However, the study analyzes data for beneficiaries that comply with the instructions during the training. This nonrandom sample significantly increases the risk of selection bias. In addition, the authors use several potentially endogenous characteristics as control variables.	**Low risk of performance bias** The analysis compares beneficiary schools with control schools that appear to be sufficiently isolated from the beneficiary schools.	**High risk of outcome and analysis reporting bias** The study uses several potentially endogenous characteristics in the estimation of the impact of the program. This approach is an indication of outcome and analysis reporting bias.	**High risk of other biases** The study does not account for clustering in the estimation of standard errors.

	Selection bias and confounding	Performance bias: Assessment, spillovers, and contamination	Outcome and analysis reporting biases	Other biases
Ismail et al. (2014)	**Medium risk of selection bias** The study uses a propensity score matching design to assess the effects of a school feeding program on early grade reading outcomes. This design enables the researchers to correct for selection bias from observable characteristics. However, the selection bias is still medium because the methodology does not allow the researchers to account for unobservable characteristics.	**Medium risk of performance bias** The analysis suggests that comparison schools may not be sufficiently isolated from the beneficiary schools. Thus, there is a medium risk of performance bias.	**Low risk of outcome and analysis reporting bias** There are no significant outcome and analysis reporting biases.	**Low risk of other biases** There is no evidence for other significant risks of bias.
Larrain et al. (2012) Study 1	**High risk of selection bias** The study randomly assigns two classrooms to the treatment group and two classrooms to the control group. This sample size is too small to ensure equivalence in observable and unobservable characteristics. In addition, the authors do not present evidence for equivalence in observable characteristics. Balance tables are not reported.	**Medium risk of performance bias** The program is randomly assigned as the classroom level within the same school. Thus, there may be interaction between beneficiary students and control students, which may result in spillovers.	**Low risk of outcome and analysis reporting bias** There are no significant outcome and analysis reporting biases.	**Low risk of other biases** There is no evidence for other significant risks of bias.

	Selection bias and confounding	Performance bias: Assessment, spillovers, and contamination	Outcome and analysis reporting biases	Other biases
Larrain et al. (2012) Study 2	**High risk of selection bias** The study randomly assigns two classrooms to the treatment group and two classrooms to the control group. This sample size is too small to ensure equivalence in observable and unobservable characteristics. In addition, the authors do not present evidence for equivalence in observable characteristics. Balance tables are not reported.	**Medium risk of performance bias** The program is randomly assigned as the classroom level within the same school. Thus, there may be interaction between beneficiary students and control students, which may result in spillovers.	**Low risk of outcome and analysis reporting bias** There are no significant outcome and analysis reporting biases.	**Low risk of other biases** There is no evidence for other significant risks of bias.
Lockheed, Harris, & Jayasundera (2010)	**Medium risk of selection bias** The study uses a propensity score matching design to determine the impact of a school governance program on early grade reading outcomes. This methodology enables the researchers to control for observable characteristics. However, the risk of selection bias remains medium because the design does not allow for controlling for unobservable characteristics.	**High risk of performance bias** The study reports that the comparison schools also often received the program but does not account for this in the analysis.	**Low risk of outcome and analysis reporting bias** There are no significant outcome and analysis reporting biases	**Medium risk of other biases** The study does not account for clustering in the estimation of the standard errors.

	Selection bias and confounding	Performance bias: Assessment, spillovers, and contamination	Outcome and analysis reporting biases	Other biases
Maluccio et al. (2009)	**Medium risk of selection bias** The study uses a cluster-RCT to determine the impact of a nutrition program on early grade reading outcomes. However, the sample size is very small and does not ensure equivalence in observable characteristics between treatment and control villages. The authors account for this concern by showing descriptive statistics, but there is nonetheless a medium risk of selection bias.	**Low risk of performance bias** The study uses village-level randomization to determine the impact of the program on early grade reading outcomes. The villages appear to be sufficiently isolated, which limits the potential for bias from spillovers or contamination.	**Low risk of outcome and analysis reporting bias** There are no significant outcome and analysis reporting biases.	**Low risk of other biases** There is no evidence for other significant risks of bias.
Mascareno et al. (2016)	**High risk of selection bias** The study uses regression analysis to determine the impact of the program. This methodology is not a well-established method to account for selection bias.	**Medium risk of performance bias** The treatment and control students are not sufficiently isolated from each other, which increases the risk of performance bias.	**Low risk of outcome and analysis reporting bias** There are no significant outcome and analysis reporting biases.	**Low risk of other biases** There is no evidence for other significant risks of bias.

	Selection bias and confounding	Performance bias: Assessment, spillovers, and contamination	Outcome and analysis reporting biases	Other biases
Mendive et al. (2016)	**High risk of selection bias** The study uses hierarchical least squares regression analysis to determine the effect of compliance with a teacher training program on early grade reading outcomes. Compliance is determined by self-selection. The use of regression analysis does not enable controlling for this self-selection. Thus, the risk of selection bias is high.	**Low risk of performance bias** The program is randomly assigned as the school level. This approach limits the interaction between beneficiary and control students, which reduces the risk of bias from spillovers or contamination.	**High risk of outcome and analysis reporting bias** The authors use arbitrary thresholds for determining whether the program was implemented with sufficient adherence and dosage. In addition, the authors use ordinary least squares (OLS) regression analysis as opposed to instrumental variable regression analysis. The authors should have used the randomization as an instrument for compliance in order to appropriately estimate the impact of compliance with the program.	**Medium risk of other Biases** The use of videos to measure teacher behavior could have resulted in Hawthorne effects, which could bias the impact of the program.
Neugebauer & Currie-Rubin (2009)	**Medium risk of selection bias** The study uses random assignment, but the sample size is insufficient to ensure equivalence in observable characteristics between treatment and control students.	**High risk of performance bias** The program uses random assignment at the individual level, which increases the risk of spillovers and contamination considerably.	**Medium risk of outcome and analysis reporting bias** The authors report results for only one outcome variable but collected data for several other outcome variables. This is an indication of outcome and analysis reporting bias.	**Medium risk of other biases** The reporting of the results suggests that the study may be biased due to Hawthorne effects.

	Selection bias and confounding	Performance bias: Assessment, spillovers, and contamination	Outcome and analysis reporting biases	Other biases
Pallante & Kim (2013)	**Low risk of selection bias** The study uses a cluster-RCT trial to determine the impact of the program on early grade reading outcomes. The authors also present evidence for balance in observable characteristics across treatment and control conditions. Thus, the risk of selection bias is low.	**Medium risk of performance bias** The study used random assignment at the classroom level. Thus, there may have been interaction between beneficiary and control students. This interaction results in a medium risk of bias from spillovers or contamination.	**Low risk of outcome and analysis reporting bias** There are no significant outcome and analysis reporting biases.	**Low Risk of Other Biases** There is no evidence for other significant risks of bias.
Powell et al. (1998)	**Low risk of selection bias** The study uses student-level randomization to determine the impact of the program on early grade reading outcomes. The study also shows evidence for balance in observable characteristics. Thus, we consider this study at low risk for selection bias.	**High risk of performance bias** The study compares beneficiary and control students within the same classroom. This approach significantly increases the risk of spillovers and contamination.	**High risk of outcome and analysis reporting bias** The study reports statistically significant effects. However, our effect size calculations suggest that the results are not statistically significant.	**Low Risk of Other Biases** There is no evidence for other significant risks of bias.
Simeon, Grantham-McGregor, & Wong (1995)	**Low risk of selection bias** The study uses student-level randomization to determine the impact of the program on early grade reading outcomes. The study also shows evidence for balance in observable characteristics. Thus, we consider this study at low risk for selection bias.	**High risk of performance bias** The study compares beneficiary and control students in the same classroom. This approach significantly increases the risk of spillovers and contamination.	**Low risk of outcome and analysis reporting bias** There are no significant outcome and analysis reporting biases.	**Low risk of other biases** There is no evidence for other significant risks of bias.

	Selection bias and confounding	Performance bias: Assessment, spillovers, and contamination	Outcome and analysis reporting biases	Other biases
Tapia & Benitez (2013)	**High risk of selection bias** The study uses random assignment with a sample of 10 treatment and 10 control students. This sample size is far too low to ensure equivalence in observable characteristics between treatment and control students.	**High risk of performance bias** The study compares beneficiary and control students in the same school. This approach significantly increases the risk of spillovers and contamination.	**Medium risk of outcome and analysis reporting bias** The study reports statistically significant effects with a very small sample size However, the authors do not report the results of the outcome equation. Instead, the results are presented in graphs. Thus, we consider this study as medium risk of outcome and analysis reporting bias.	**Medium risk of other biases** The study suggests that researchers were heavily involved in data collection, which may have resulted in Hawthorne effects.
Murad & Topping (2000)	**High risk of selection bias** The study uses random assignment but the sample size is too small to ensure equivalence in observable characteristics. In addition, treatment students were switched to the control group because they could not comply with the intervention. Together, these constraints result in a high risk of selection bias.	**High risk of performance bias** The study compares beneficiary and control students in the same school. This approach significantly increases the risk of spillovers and contamination.	**Medium risk of outcome and analysis reporting bias** The authors exclude outliers from their analysis for unclear reasons. The exclusion of these outliers may have affected the statistical significance of the impact estimates.	**Medium risk of other biases** The study suggests that researchers were heavily involved in data collection, which may have resulted in Hawthorne effects.

	Selection bias and confounding	Performance bias: Assessment, spillovers, and contamination	Outcome and analysis reporting biases	Other biases
Viramontes et al. (2016)	**High risk of selection bias** The authors claim to use random assignment, but based on the article it is unclear whether the random assignment was successful. In addition, the sample size of the study was too small to ensure equivalence in observable and unobservable characteristics.	**High risk of performance bias** It appears as if the study compared beneficiary and control students in the same school. This approach could have significantly increased the risk of spillovers and contamination.	**High risk of outcome and analysis reporting bias** The analysis is mostly based on graphical analysis, which is unusual.	**Medium risk of other biases** The intervention is poorly described, which makes it hard to determine whether the study suffers from other biases.
Vivas (1996) Experiment 1	**Medium risk of selection bias** The study uses random assignment but the sample size is too small to ensure equivalence in observable characteristics.	**Low risk of performance bias** The program is randomly assigned at the school level. This approach limits the interaction between beneficiary and control students, which reduces the risk of bias from spillovers or contamination.	**Medium risk of outcome and analysis reporting bias** The study estimates the impact of the program by comparing the median value of early grade reading outcomes between beneficiary and control students. This approach is unusual and may be an indication of outcome and analysis reporting bias.	**Low Risk of Other Biases** There is no evidence for other significant risks of bias.
Vivas (1996) Experiment 2	**Medium risk of selection bias** The study uses random assignment but the sample size is too small to ensure equivalence in observable characteristics.	**Low risk of performance bias** The program is randomly assigned at the school level. This approach limits the interaction between beneficiary and control students, which reduces the risk of bias from spillovers or contamination.	**Medium risk of outcome and analysis reporting bias** The study estimates the impact of the program by comparing the median value of early grade reading outcomes between beneficiary and control students. This approach is unusual and may be an indication of outcome and analysis reporting bias.	**Low Risk of Other Biases** There is no evidence for other significant risks of bias.

	Selection bias and confounding	Performance bias: Assessment, spillovers, and contamination	Outcome and analysis reporting biases	Other biases
Yoshikawa et al. (2015)	**Low risk of selection bias** The study uses a cluster-RCT to determine the impact of the program on early grade reading outcomes. The authors also present evidence for balance in observable characteristics across treatment and control conditions. Thus, the risk of selection bias is low.	**Low risk of performance bias** The program is randomly assigned at the school level. This approach limits the interaction between beneficiary and control students, which reduces the risk of bias from spillovers or contamination.	**Low risk of outcome and analysis reporting bias** There are no significant outcome and analysis reporting biases.	**Medium risk of other biases** The use of videos to measure teacher behavior could have resulted in Hawthorne effects, which could bias the impact of the program.