Program/intervention	Level of evidence	Benefit-cost percentage	Reason program does not meet suggested evidence-based criteria	Percent children of color
Tutoring support				
Tutoring: By certificated teachers, small-group, structured	٠	97%		63%
Tutoring: By adults, one-on-one, structured	•	92%		66%
Tutoring: By peers	٠	81%		46%
Tutoring: By non-certificated adults, small-group, structured	۲	69%	Benefit-cost	65%
Tutoring: By adults, one-on-one, non-structured	۲	69%	Benefit-cost	77%
Tutoring: Supplemental Educational Services (under Title I)	۲	62%	Benefit-cost	95%
Tutoring: By adults, for English language learner students	Null	60%	Weight of evidence	91%
Tutoring: Supplemental computer-assisted instruction for struggling readers (vs. regular classroom instruction)	Null	64%	Weight of evidence	91%
Extended learning time				
Double dose classes	٠	98%		91%
Out-of-school-time tutoring by adults	•	93%		75%
Summer learning programs: Academically focused	•	87%		85%
Summer book programs: One-year, with additional support	Null	58%	Weight of evidence	77%
Summer book programs: One-year intervention	Null	57%	Weight of evidence	86%
Summer book programs: Multi-year intervention	Р	71%	Weight of evidence	95%
Professional development				
Teacher professional development: Use of data to guide instruction	•	98%		54%
Teacher professional development: Targeted	•	79%		96%
Teacher professional development: Online, targeted	۲	60%	Benefit-cost/heterogeneity	31%
Teacher professional development: Induction/mentoring	Null	64%	Weight of evidence	92%
Teacher professional development: Not targeted	Null	38%	Weight of evidence	51%
Educator professional development: Use of data to guide instruction ("train the trainers")	Null	29%	Weight of evidence	46%
Professional learning communities	Р		No rigorous evaluation with outcome of interest	
Consultant teachers				
Consultant teachers: Online coaching	•	93%		53%
Consultant teachers: Coaching	•	81%		53%
Consultant teachers: Literacy Collaborative	۲	100%	Heterogeneity	29%
Consultant teachers: Content-focused coaching	۲		Single evaluation	96%

• Evidence-based • Research-based • Promising • Poor outcomes Null Null outcomes NR Not reported See definitions and notes on pages 4 and 5.

Note:

[#] This program is a special analysis for this inventory and does not have a program-specific webpage on WSIPP's website.

The classifications in this document are current as of July 2020.

For the most up-to-date results, please visit the program's page on our website http://www.wsipp.wa.gov/BenefitCost

Program/intervention	Level of evidence	Benefit-cost percentage	Reason program does not meet suggested evidence-based criteria	Percent children of color
Parent outreach				
Parents as tutors with teacher oversight	۲	56%	Benefit-cost	58%
Families and Schools Together (FAST)	Null	50%	Weight of evidence	83%
Conjoint behavioral consultation	Null	23%	Weight of evidence	21%
Parent and family engagement coordinators	Р		No rigorous evaluation with outcome of interest	
Community partnerships				
Case management in schools [#]	۲	68%	Mixed results/benefit-cost	61%
Mentoring: School-based (taxpayer costs only) [#]	۲	19%	Benefit-cost	74%
Mentoring: School-based (with volunteer costs) $^{\#}$	۲	16%	Benefit-cost	74%
Mentoring: Community-based (taxpayer costs only) [#]	۲	66%	Benefit-cost	68%
Mentoring: Community-based (including volunteer costs) [#]	۲	57%	Benefit-cost	68%
PROSPER (PROmoting School-community-university Partnerships to Enhance Resilience)	۲	39%	Benefit-cost/heterogeneity	15%
Behavior support				
Positive Action	•	94%		63%
Good Behavior Game	•	76%		50%
Second Step	•	85%		53%
Becoming a Man (BAM)	۲	74%	Benefit-cost	98%
Becoming a Man (BAM) with high-dosage tutoring	۲		Single evaluation	99%
Mentoring: Community-based for children with disruptive behavior disorders	۲	67%	Benefit-cost/heterogeneity	7%
Mentoring: School-based by teachers or staff	۲	71%	Benefit-cost	86%
School-Wide Positive Behavioral Interventions and Supports	۲	74%	Benefit-cost	50%
Behavioral Monitoring and Reinforcement Program (BMRP)	۲	64%	Benefit-cost	81%
Coping Power Program	۲	57%	Benefit-cost	80%
"Check-in" behavior interventions	۲	57%	Benefit-cost	72%
Restorative justice in schools	۲	11%	Mixed results/benefit-cost	65%
Fast Track prevention program	۲	0%	Benefit-cost	53%
Daily Behavior Report Cards	۲		Single evaluation	13%
First Step to Success	Null	47%	Weight of evidence	59%
Caring School Community (formerly Child Development Project)	Null	60%	Weight of evidence	47%
Responsive classroom	Null	4%	Weight of evidence	57%
Curriculum-based Support Group (CBSG)	Р		Single evaluation	90%

• Evidence-based • Research-based • Promising • Poor outcomes Null Null outcomes NR Not reported See definitions and notes on pages 4 and 5.

Note:

[#] This program is a special analysis for this inventory and does not have a program-specific webpage on WSIPP's website.

The classifications in this document are current as of July 2020.

For the most up-to-date results, please visit the program's page on our website http://www.wsipp.wa.gov/BenefitCost

Program/intervention	Level of evidence	Benefit-cost percentage	Reason program does not meet suggested evidence-based criteria	Percent children of color
Services for 8 th , 11 th , & 12 th grades				
Credit retrieval	Р		No rigorous evaluation with outcome of interest	
Other				
Special literacy instruction for English language learner students	٠	81%		98%
Growth mindset interventions	۲	56%	Benefit-cost	71%
Academic vocabulary instruction	Р		Weight of evidence	NR
Transition programs for incoming kindergarteners	Р		Single evaluation	45%

• Evidence-based • Research-based • Promising • Poor outcomes Null Null outcomes NR Not reported See definitions and notes on pages 4 and 5.

Note:

[#] This program is a special analysis for this inventory and does not have a program-specific webpage on WSIPP's website.

Definitions and Notes

Level of Evidence:

- Evidence-based: A program or practice that has been tested in heterogeneous or intended populations with multiple randomized and/or statistically-controlled evaluations, or one large multiple-site randomized and/or statistically-controlled evaluation, where the weight of the evidence from a systematic review demonstrates sustained improvements in at least one outcome. Further, "evidence-based" means a program or practice that can be implemented with a set of procedures to allow successful replication in Washington and, when possible, has been determined to be cost-beneficial.
- <u>Research-based</u>: A program or practice that has been tested with a single randomized and/or statistically-controlled evaluation demonstrating sustained desirable outcomes; or where the weight of the evidence from a systematic review supports sustained outcomes as identified in the term "evidence-based" in RCW (the above definition) but does not meet the full criteria for "evidence-based."
- <u>Promising practice</u>: A program or practice that, based on statistical analyses or a well-established theory of change, shows potential for meeting the "evidence-based" or "research-based" criteria, which could include the use of a program that is evidence-based for outcomes other than the alternative use.

Other Definitions:

<u>Cost-beneficial</u>: Benefit-cost estimation is repeated many times to account for uncertainty in the model. This represents the percentage of repetitions producing overall benefits that exceed costs. Programs with a benefit-cost percentage of at least 75% are considered to meet the "cost-beneficial" criterion in the "evidence-based" definition above.

Reasons Programs May Not Meet Suggested Evidence-Based Criteria:

- <u>Benefit-cost</u>: The proposed definition of evidence-based practices requires that, when possible, a benefit-cost analysis be conducted. We use WSIPP's benefit-cost model to determine whether a program meets this criterion. Programs that do not have at least a 75% chance of a positive net present value do not meet the benefit-cost test. The WSIPP model uses Monte Carlo simulation to test the probability that benefits exceed costs. The 75% standard was deemed an appropriate measure of risk aversion.
- Heterogeneity: To be designated as evidence-based, a program must have been tested on a "heterogeneous" population. We operationalized heterogeneity in two ways. First, the proportion of program participants must be greater than or equal to the proportion of children of color in Washington State aged 0 to 17. From the 2010 Census, among children aged 0 through 17 in Washington, 68% were White and 32% were children of color. Thus, if the weighted average of program participants were at least 32% children of color then the program was considered to have been tested on a heterogeneous population. Second, the heterogeneity criterion can also be achieved if at least one of the studies has been conducted on youth in Washington and a subgroup analysis demonstrates the program is effective for children of color (p < 0.20). Programs passing the second test are marked with a ^. Programs that do not meet either of these two criteria do not meet the heterogeneity definition.
- <u>Mixed results</u>: If findings are mixed from different measures (e.g., undesirable outcomes for behavior measures and desirable outcomes for test scores), the program does not meet evidence-based criteria.

No rigorous evaluation measuring outcome of interest: The program has not yet been tested with a rigorous outcome evaluation.

- Null outcomes: If results from multiple evaluations or one large multiple-site evaluation indicate that a program has no significant effect on outcomes of interest (p > 0.20), a program is classified as producing "null outcomes."
- <u>Poor outcome(s)</u>: If results from multiple evaluations or one large multiple-site evaluation indicate that a program produces undesirable effects (p < 0.20), a program is classified as producing "poor outcomes."
- Single evaluation: The program does not meet the minimum standard of multiple evaluations or one large multiple-site evaluation contained in the current or proposed definitions.
- <u>Weight of evidence</u>: Results from a random-effects meta-analysis (p > 0.20) indicate that the weight of the evidence does not support desired outcomes, or results from a single large study indicate the program is not effective.