Washington State Institute for Public Policy

110 Fifth Avenue SE, Suite 214 • PO Box 40999 • Olympia, WA 98504 • 360.586.2677 • www.wsipp.wa.gov

July 2016

Updated Inventory of Evidence- and Research-Based Practices: Washington's K–12 Learning Assistance Program

Washington State provides funding to school districts to help underachieving students through the Learning Assistance Program (LAP). The 2013 Washington State Legislature directed three efforts to identify effective practices for helping students served through LAP.

- ✓ The Washington State Institute for Public Policy (WSIPP) was directed to "prepare an inventory of evidence-based and research-based effective practices, activities, and programs for use by school districts in the learning assistance program."¹ The updated inventory is displayed on page 9 of this report and is also available online.²
- ✓ The Office of Superintendent of Public Instruction (OSPI) was directed to "convene a panel of experts, including the Washington state institute for public policy, to develop additional state menus of best practices and strategies for use in the learning assistance program to assist struggling students at all grade levels in English language arts and mathematics and reduce disruptive behaviors in the classroom."³
- ✓ OSPI was also directed to "convene a panel of experts, including the Washington state institute for public policy, to develop a state menu of best practices and strategies for intensive reading and literacy improvement designed to assist struggling students in reaching grade level in reading by the end of fourth grade."⁴

Legislative assignment	Due date	Follow-up
WSIPP to develop a LAP inventory of evidence- and research- based practices, activities, and programs	August 1, 2014	Update every two years thereafter*
OSPI to convene a panel of LAP experts to develop a menu of best practices and strategies	July 1, 2015	Update each July 1 st thereafter
OSPI to convene a panel of English language arts (ELA) experts to develop a menu of best practices and strategies to help students reach grade level in reading by the end of 4 th grade	July 1, 2014	Update each July 1 st thereafter

The table below summarizes the implementation timeline.

*WSIPP updated the LAP inventory in July 2015 to align with OSPI's menu timeline.

¹ Third Engrossed Substitute Senate Bill 5034, Chapter 4, Sec. 610(5), Laws of 2013.

² http://wsipp.wa.gov/ReportFile/1636

³ Engrossed Substitute Senate Bill 5946, Chapter 18, Sec. 206(3), Laws of 2013.

⁴ Ibid.

These legislative assignments are similar but distinct. Staff members from WSIPP and OSPI are coordinating the inventory and menu processes to ensure that the results are consistent. WSIPP staff served as non-voting members on the expert panels and provided research findings that informed the panels' deliberations. The expert panels helped WSIPP identify topics for analysis and reviewed topics for inclusion on the inventory when no evaluation research was available. The final classifications on the LAP inventory and expert panel menus, however, reflect each group's independent judgment.

The expert panel's menu of best practices is presented in a separate OSPI report, which can be found on OSPI's website.⁵

Creating the LAP Inventory

Washington State's Learning Assistance Program was created by the legislature in 1987 to assist underachieving students. In the 2015-17 biennium, the legislature appropriated \$453 million for LAP.⁶ The funds can be used for a variety of practices, strategies, and activities in K–12 schools, including:

- Tutoring support;
- Extended learning time;
- Professional development;
- Consultant teachers;
- Parent outreach;
- Community-based partnerships;
- Addressing disruptive behavior in the classroom; and
- Services for 8th, 11th, and 12th graders.⁷

WSIPP consulted with legislative staff, OSPI, and members of the expert panel to develop a list of the highest priority topics to investigate for this inventory. To date, we have reviewed 54 interventions.

Our approach to creating the inventory is the same approach we have used for legislatively directed inventories in other policy areas.⁸ The first step is to estimate the probability that various public policies and programs can achieve desired outcomes, such as improving high

⁵ http://www.k12.wa.us/LAP/default.aspx

⁶ Second Engrossed Substitute House Bill 2376, Chapter 36, Sec. 513, Laws of 2016.

⁷ RCW 28A.165.035.

⁸ Miller, M., Fumia, D., & Kay, N. (2015). Updated inventory of evidence-based, research-based, and promising practices prevention and intervention services for adult behavioral health. (Doc. No. 15-01-4101). Olympia: Washington State Institute for Public Policy.

school graduation rates or student test scores.⁹ For each topic, we carefully analyze all highquality studies from the United States and elsewhere to identify interventions that have been tried, tested, and found to either achieve or not achieve improvements in outcomes. We look for research studies with strong evaluation designs and exclude studies with weak research methods. Using all credible evaluations we can locate on a given topic, we then conduct a metaanalysis to determine the average effect of the program and a margin of error for that effect.¹⁰ These research standards are outlined in the box below.

The second step is to use the results from our analysis of program effects to determine whether the lifetime benefits of the program exceed the costs to Washington's taxpayers. That is, we conduct a formal benefit-cost analysis.

The third analytical step involves testing the robustness of our results. Any tabulation of benefits and costs involves some degree of uncertainty about future performance. This uncertainty is expected in any investment analysis, whether in the private or public sector. To assess the riskiness of our conclusions, we perform a "Monte Carlo simulation" in which we vary the key factors in our calculations. The purpose of the risk analysis is to determine the odds that the benefits of a particular policy option will exceed the costs. This type of analysis is used by many businesses in investment decision making.

Thus, for each option, we produce two "big picture" findings: expected benefit-cost results (net present values and benefit-cost ratios) and, given our understanding of the risks involved, the odds that the policy will at least have benefits greater than costs.

Standards of Research Rigor for Meta-Analysis

When WSIPP is asked by the legislature to conduct an evidence-based review, we follow a number of steps to ensure a rigorous and consistent analysis. These procedures include the following:

- ✓ We consider all available studies we can locate on a topic rather than selecting only a few; that is, we do not "cherry pick" studies to include in our reviews.
- To be included in our reviews, we require that an evaluation's research design include treatment and comparison groups from intent-to-treat samples. Random assignment studies are preferred, but we include quasi-experimental studies when the study uses appropriate statistical techniques. Natural experimental designs, including regression discontinuity and instrumental variables, are also considered.
- ✓ We then use a formal statistical procedure, meta-analysis, to calculate an average "effect size" that indicates the expected magnitude of the relationship between the treatment and the outcome of interest. This is how we determine whether the weight of the evidence indicates outcomes are, on average, achieved.

⁹ For the inventory, we look for studies measuring outcomes related to the goals of the Learning Assistance Program (to assist underachieving students and reduce disruptive behaviors in the classroom—RCW 28A.165.005). For example, we include studies that measure changes in test scores, graduation rates, grade point average, attendance, and suspensions/expulsions. We do not include studies that measure outcomes that may or may not be related to the change in students' educational outcomes (such as teacher or student satisfaction), if the studies did not also measure the outcomes of interest.

¹⁰ All methods are described in detail in

http://www.wsipp.wa.gov/TechnicalDocumentation/WsippBenefitCostTechnicalDocumentation.pdf

Classifying Practices as Evidence-Based, Research-Based, or Promising

The legislative assignment directs WSIPP to identify evidence- and research-based practices for LAP. Washington's K–12 laws do not define these terms. The adult behavioral health statutes, however, do provide definitions, and WSIPP recently published an adult behavioral health inventory using these definitions.¹¹ For the LAP inventory, we use the statutory definitions applicable to adult behavioral health to maintain consistency across policy areas.

We also include "promising practices" on the inventory when the OSPI-convened expert panel and/or the research evidence suggests the practice might improve student outcomes, but the topics did not meet the criteria for evidence- or research-based.

The following definitions are taken directly from the law.

Legislative Definitions of Evidence-Based, Research-Based, and Promising Practices

Evidence-based

A program or practice that has been tested in heterogeneous or intended populations with multiple randomized, or statistically controlled evaluations, or both; or one large multiple site randomized, or statistically controlled evaluation, or both, where the weight of the evidence from a systemic review demonstrates sustained improvements in at least one outcome. "Evidence-based" also means a program or practice that can be implemented with a set of procedures to allow successful replication in Washington and, when possible, is determined to be cost-beneficial.

Research-based practice

A program or practice that has been tested with a single randomized, or statistically controlled evaluation, or both, demonstrating sustained desirable outcomes; or where the weight of the evidence from a systemic review supports sustained outcomes [...] but does not meet the full criteria for evidence-based.

Promising practice

A practice that, based on statistical analyses or a well-established theory of change, shows potential for meeting the evidence-based or research-based criteria [...].

RCW 71.24.025

¹¹ RCW 71.24.025. WSIPP's adult behavioral health inventory can be found on our website; Miller, M., Fumia, D., & Kay, N. (2015). Updated Inventory of evidence-based, research-based, and promising practices prevention and intervention services for adult behavioral health. (Doc. No. 15-01-4101). Olympia: Washington State Institute for Public Policy.

For each program where research is available, we conduct meta-analysis and benefit-cost analysis to classify practices as evidence- or research-based according to the above definitions. If outcome evaluations exist but the evidence indicates a non-significant effect (p-value > 0.20)¹² on desired outcomes in the expected direction, then the program is designated as promising. When we locate no rigorous outcome evaluations for a program, we rely on the panel of experts assembled by OSPI to determine whether the program meets the criteria for promising.

To assemble the inventory, we operationalize each criterion in the statutory definitions. These are the same criteria WSIPP has used in assembling inventories in children's mental health, child welfare, juvenile justice, and adult behavioral health. The criteria are as follows:

 <u>Heterogeneity.</u> To be designated as evidence-based, the state statute requires that a program has been tested on a "heterogeneous" population. We operationalize heterogeneity in two ways. First, the proportion of program participants belonging to ethnic/racial minority groups must be greater than or equal to the proportion of minority children aged 0 to 17 in Washington. From the 2010 Census, for children aged 0 through 17 in Washington, 68% were white and 32% were minorities.¹³ Thus, if the weighted average of program participants in the outcome evaluations of the program is at least 32% ethnic/racial minority, then the program is considered to have been tested in a heterogeneous population.

Second, the heterogeneity criterion can also be achieved if at least one of a program's outcome evaluations was conducted with K–12 students in Washington and a subgroup analysis demonstrates the program is effective for ethnic/racial minorities.

Programs whose evaluations do not meet either of these two criteria do not meet the heterogeneity definition.

- <u>Weight of evidence</u>. To meet the evidence-based definition, results from a random effects meta-analysis (p-value < 0.20) of multiple evaluations or one large multiple-site evaluation must indicate the practice achieves the desired outcome(s). To meet the research-based definition, one single-site evaluation must indicate the practice achieves the desired outcomes (p-value < 0.20).
- 3) <u>Benefit-cost.</u> The statute defining 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.¹⁴ 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.

¹² In order to operationalize the benefit-cost criterion for a program to be classified as evidence-based, net benefits must exceed costs at least 75% of the time. After considerable analysis, we found that a typical program that WSIPP has analyzed may produce benefits that exceed costs roughly 75% of the time with a p-value cut-off of up to 0.20. Thus, we determined that programs with p-values less than or equal to 0.20 on desired outcomes should be considered research-based in order to avoid classifying programs with desirable benefit-cost results as promising.

¹³ United States Census Bureau, 2010. Retrieved from http://factfinder2.census.gov/.

¹⁴ For information about WSIPP's benefit-cost model, see

http://www.wsipp.wa.gov/TechnicalDocumentation/WsippBenefitCostTechnicalDocumentation.pdf

If a program is not listed on the inventory, we have not yet had the opportunity to review it. If a program is listed on the inventory but does not meet any of the criteria for evidence-based, research-based, or promising, then the program is ineffective or has adverse effects and should not be used if the goal is to achieve one of the desired outcomes, such as increasing student achievement or reducing disruptive behavior.

The LAP inventory is displayed at the end of this report and is also available on our website.¹⁵ Further information on the individual programs contained in the inventory can also be found on our website.¹⁶

Updates to the Inventory as of July 2016

Since the last inventory update in July 2015, WSIPP modified the statistical calculations applied to some types of studies and adjusted its benefit-cost methodology. These calculations affected the detailed statistical results for each program. Due to these changes, WSIPP reclassified five programs.

Two programs moved from research-based to evidence-based.

- ✓ Online Coaching¹⁷
- ✓ Other school-wide positive behavior programs¹⁸

One program moved from evidence-based to research-based.

✓ Good Behavior Game¹⁹

Two programs moved from promising to research-based.

- ✓ Becoming a Man with high-dosage tutoring²⁰
- ✓ Second Step²¹

WSIPP added seven programs that provide academic or behavioral services to identified students.

Two programs were classified as research-based.

- ✓ Coping Power Program
- ✓ PROSPER

¹⁵ http://wsipp.wa.gov/ReportFile/1636

¹⁶ http://www.wsipp.wa.gov/BenefitCost

¹⁷ This program was reclassified due to the addition of a new study.

¹⁸ This program was reclassified due to a reanalysis of the included study.

¹⁹ This program was reclassified due to an adjustment to our benefit-cost methodology.

²⁰ This program was reclassified due to a reanalysis of the included study.

²¹ This program was reclassified due to the addition of new studies and a reanalysis of the results.

Five programs were classified as promising.

- ✓ Academic vocabulary instruction
- ✓ Curriculum-Based Support Groups (CBSG)
- ✓ Responsive Classroom
- ✓ Transition program for incoming kindergarteners
- ✓ Tutoring: Supplemental computer-assisted instruction for struggling readers

Limitations

The benefit-cost analyses in this report reflect only those outcomes that were measured in the studies we reviewed. We focus primarily on outcomes that are "monetizable" with the current WSIPP benefit-cost model. "Monetizable" means that we can link the outcome to future economic consequences, such as labor market earnings, criminal justice involvement, or health care expenditures. At this time we are unable to monetize some relevant outcomes including suspensions/expulsions and attendance.

Future Updates

The next update to this inventory will be published by August 1, 2018.

Acknowledgments

We would like to thank staff at the Office of Superintendent of Public Instruction and the Panels of Experts convened by OSPI for their guidance in identifying interventions to review and assistance in describing specific programs. We are especially grateful to Paula Moore, Amy Thierry, Joshua Lynch, and Porsche Everson for their input and collaboration throughout this process.

Program/intervention	Level of evidence	Benefit-cost percentage	Reason program does not meet evidence-based criteria (see full definitions below)	Percent minority	
Tutoring support					
Tutoring: By certificated teachers, small-group, structured		96%		63%	
Tutoring: By adults, one-on-one, structured	•	94%		72%	
Tutoring: By non-certificated adults, small-group, structured		77%		69%	
Tutoring: By peers, cross-age [#]	۲	81%	Heterogeneity	NR	
Tutoring: By peers, same-age and classwide $^{\#}$	۲	74%	Benefit-cost	62%	
Tutoring: By adults, one-on-one, non-structured	۲	70%	Benefit-cost	75%	
Tutoring: Supplemental Educational Services (under Title I)	۲	58%	Benefit-cost	95%	
Tutoring: By adults, for English language learner students	Р	68%	Weight of evidence	91%	
Tutoring: Supplemental computer-assisted instruction for struggling readers	Р	58%	Weight of evidence	91%	
Extended learning time					
"Double dose" classes	•	98%		91%	
Out-of-school-time tutoring by adults	•	92%		75%	
Summer learning programs: Academically focused	•	86%		85%	
Summer book programs: One-year, with additional support	Р	58%	Weight of evidence	77%	
Summer book programs: One-year intervention	Р	56%	Weight of evidence	86%	
Summer book programs: Multi-year intervention	Р		Single evaluation	95%	
Professional development					
Teacher professional development: Use of data to guide instruction	•	98%		54%	
Teacher professional development: Targeted		78%		96%	
Teacher professional development: Online, targeted	۲	60%	Benefit-cost/heterogeneity	31%	
Teacher professional development: Induction/mentoring	Р	63%	Weight of evidence	92%	
Professional Learning Communities	Р		Research on outcomes of interest not yet available		
Teacher professional development: Not targeted	0	37%	Weight of evidence	51%	
Educator professional development: Use of data to guide instruction ("train the trainers")	0	31%	Weight of evidence	46%	
Consultant teachers					
Online coaching	•	86%		53%	
Coaching	•	82%		53%	
Coaching: Literacy Collaborative	۲	100%	Heterogeneity	29%	
Coaching: Content-Focused Coaching	۲		Single evaluation	96%	

• Evidence-based • Research-based • P Promising • Produces null or poor outcomes **NR** Not reported See definitions and notes on page 4.

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

Program/intervention	Level of evidence	Benefit-cost percentage	Reason program does not meet evidence-based criteria (see full definitions below)	Percent minority				
Parent outreach								
Parents as tutors with teacher oversight	۲	56%	Benefit-cost	58%				
Families and Schools Together (FAST)	۲	49%	Benefit-cost/mixed results	90%				
Conjoint Behavioral Consultation	Р		Single evaluation	25%				
Parent and family engagement coordinators	Р		Research on outcomes of interest not yet available					
Community partnerships								
Case management in schools	۲	96%	Mixed results	61%				
Mentoring for students: School-based (taxpayer costs only)	۲	74%	Benefit-cost	73%				
Mentoring for students: School-based (with volunteer costs)	۲	72%	Benefit-cost	73%				
Mentoring for students: Community-based (taxpayer costs only)	۲	71%	Benefit-cost	78%				
Mentoring for students: Community-based (with volunteer costs)	۲	66%	Benefit-cost	78%				
PROSPER	•	55%	Benefit-cost/heterogeneity	15%				
Behavior support								
Positive Action	•	88%		63%				
Other school-wide positive behavior programs	•	77%		72%				
Good Behavior Game	۲	71%	Benefit-cost	56%				
Behavioral Monitoring and Reinforcement Program (BMRP)	۲	63%	Benefit-cost	41%				
First Step to Success	۲	53%	Benefit-cost	59%				
Coping Power Program	۲	50%	Benefit-cost	80%				
Second Step	۲	47%	Benefit-cost	55%				
"Check-in" behavior interventions	۲	45%	Benefit-cost	72%				
Fast Track prevention program	۲	0%	Benefit-cost	53%				
Becoming a Man (BAM)	۲		Single evaluation	99%				
Becoming a Man (BAM) with high-dosage tutoring	۲		Single evaluation	99%				
Daily Behavior Report Cards	۲		Single evaluation	13%				
Responsive Classroom	Р	48%	Weight of evidence	57%				
Curriculum-based Support Group (CBSG)	Р		Single evaluation	90%				

• Evidence-based • Research-based • Promising • Produces null or poor outcomes **NR** Not reported See definitions and notes on page 4.

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

Program/intervention	Level of evidence	Benefit-cost percentage	Reason program does not meet evidence-based criteria (see full definitions below)	Percent minority
Services for 8th, 11th & 12th grades				
Credit retrieval	Р		Research on outcomes of interest not yet available	
Other				
Special literacy instruction for English language learner students	•	80%		98%
Academic Vocabulary Instruction	Р		Single evaluation	NR
Transition Programs for Incoming Kindergarteners	Р		Single evaluation	45%

• Evidence-based • Research-based • Promising • Produces null or poor outcomes • NR Not reported See definitions and notes on page 4.

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

Definitions and Notes:

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, the state statute requires that a program has been tested on a "heterogeneous" population. We operationalize heterogeneity in two ways. First, the proportion of program participants belonging to ethnic/racial minority groups must be greater than or equal to the proportion of minority children aged 0 to 17 in Washington. From the 2010 Census, for children aged 0 through 17 in Washington, 68% were white and 32% were minorities. Thus, if the weighted average of program participants in the outcome evaluations of the program is at least 32% ethnic/racial minority, then the program is considered to have been tested in 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 minorities (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.

Programs whose evaluations do not meet either of these two criteria do not meet the heterogeneity definition.

Mixed results: 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.

Research on outcomes of interest not yet available: The program has not yet been tested with a rigorous outcome evaluation.

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.

Weight of evidence: To meet the evidence-based definition, results from a random effects meta-analysis (p-value < 0.20) of multiple evaluations or one large multiple-site evaluation must indicate the practice achieves the desired outcome(s). To meet the research-based definition, one single-site evaluation must indicate the practice achieves the desired outcomes (p-value < 0.20).

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."
- Promising practice: 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:

Benefit-cost percentage: The percent of the time where the monetary benefits exceed costs.

Suggested citation: Lemon, M. (2016). *Updated inventory of evidence- and research-based practices: Washington's K– 12 Learning Assistance Program*. (Doc. No. 16-07-2201). Olympia: Washington State Institute for Public Policy.

For further information, contact: Matt Lemon at 360.586.2744, matt.lemon@wsipp.wa.gov

Document No. 16-07-2201

Washington State Institute for Public Policy

The Washington State Legislature created the Washington State Institute for Public Policy in 1983. A Board of Directors representing the legislature, the governor, and public universities—governs WSIPP and guides the development of all activities. WSIPP's mission is to carry out practical research, at legislative direction, on issues of importance to Washington State.