

## Top Tier Evidence Initiative:

### **A Validated Resource For Distinguishing Research-Proven Social Programs from Everything Else**

The nonprofit, nonpartisan Coalition for Evidence-Based Policy launched the Top Tier Evidence initiative in 2008, in consultation with senior federal officials, to identify social programs meeting the top tier evidence standard set out in recent legislative provisions: “*well-designed randomized controlled trials [showing] sizable, sustained effects on important... outcomes*” [e.g., Public Laws 110-161 and 111-8]. This paper briefly summarizes the purpose of the initiative and how it works.

#### **Policy question this effort addresses:**

**Across social policy, which program models/strategies (“interventions”) are supported by definitive evidence of sizable, sustained effects?** This question might be asked, for example, by policy officials who wish to focus their efforts on replicating or scaling up the few interventions in their area for which research provides strong confidence of a sizable effect on people’s lives.

- A. Consistent with a recent [National Academy of Sciences report](#), the Top Tier initiative recognizes well-conducted randomized controlled trials as needed to answer this question.** Per the Academy report, evidence of effectiveness generally “cannot be considered definitive” without ultimate confirmation in well-conducted randomized controlled trials, “even if based on the next strongest designs.”<sup>1</sup> This concept, and the evidence supporting it, are discussed more fully below.
- B. The goal is not to identify *all* evidence-based interventions – just those whose evidence provides strong confidence that a faithful replication would produce important life improvement.** We recognize that, for many social problems, no interventions currently meet the Top Tier because of gaps in research or other reasons. Thus, public officials seeking to address these problems may need to rely on evidence that falls below the Top Tier, including nonrandomized studies as well as randomized studies with small samples or other limitations. The Top Tier initiative does not review such evidence; however, we understand its value and refer users to other high-quality resources that do.

#### **Why this initiative is needed:**

- A. Government programs set up to address important social problems often fall short by funding specific interventions that are not effective.** When evaluated in scientifically-rigorous studies, social interventions in K-12 education, job training, crime prevention, and other areas are frequently found ineffective or marginally effective. Interventions that produce sizable, sustained effects on important life outcomes – such as high school graduation, teen pregnancy, criminal arrests, and workforce earnings – tend to be the exception. This pattern occurs in diverse areas of social policy, as well as other fields where rigorous studies have been conducted, such as medicine.
- B. Improving social programs is critically needed. The United States has failed to make significant progress in key areas such as –**
- **Poverty:** The U.S. poverty rate now stands at 14.5%, and has shown little overall change (whether by official or alternative National Academy measures) since the late 1970s.<sup>2</sup>
  - **K-12 education:** Reading and math achievement of 17-year-olds – the end product of our K-12 education system – is virtually unchanged over the past 40 years, according to official measures,<sup>3</sup> despite a 90% increase in public spending per student (adjusted for inflation).<sup>4</sup>

- **Well-being of low to moderate income Americans:** The average yearly income of the lowest 40% of U.S. households, now at \$21,100, has changed little since the 1970s.<sup>5</sup>

**C. A few interventions meeting the Top Tier do exist and, if implemented more widely, could help spark rapid progress against major national problems.** The following are examples of interventions that the initiative has already identified as meeting the Top Tier:

- [Nurse-Family Partnership](#) – a nurse visitation program for low-income, first-time mothers during pregnancy and children’s infancy (reduced child abuse/neglect and injuries by 20-50% over 2-15 years, compared to the control group).
- [Carrera Adolescent Pregnancy Prevention Program](#) – a youth development program for low-income teens (at age 17, reduced girls’ pregnancies and births by 40-50%, compared to the control group).
- [H&R Block College Financial Aid Application Assistance](#) – streamlined personal assistance for low and moderate income families with a dependent child near college age (over a 3½-4 year period, increased college enrollment and persistence by 29%, compared to the control group).
- [Career Academies](#) – small learning communities in low-income high schools, offering academic and technical/career courses as well as workplace opportunities (8 years after high school, increased average earnings by \$2,200 per year, compared to the control group).

**D. Currently, there is no efficient way for public officials to distinguish the few interventions backed by Top Tier evidence from the many that *claim to be*.** There are currently about 15 widely-cited federal, state, and private websites and related resources profiling evidence-based interventions in various areas of social policy. The Coalition has carefully examined these sites, in projects with the Office of Management and Budget and Justice Department, and found the following:

- **Most sites list interventions backed by strong evidence along with many backed by more preliminary evidence that is often not confirmed in later, more definitive evaluations.** Such evidence includes, for example, nonrandomized comparison-group studies (“quasi-experiments”), or randomized controlled trials with only short-term follow-up, assessment of intermediate rather than final outcomes (e.g., condom use versus actual teen pregnancies), or other key limitations in study design or implementation. As noted above, these studies can be valuable for decision-making, and these websites can therefore be useful, in the absence of stronger evidence.

Too often, however, findings from such initial studies are not confirmed in larger, more definitive randomized controlled trials. Reviews in medicine, for example, have found that 50-80% of promising results from phase II studies (mostly quasi-experiments or preliminary “efficacy” trials) are overturned in subsequent phase III randomized controlled trials.<sup>6</sup> Similarly, in education, the U.S. Education Department’s Institute of Education Sciences has sponsored large randomized controlled trials of more than 75 educational interventions since the Institute’s establishment in 2002; the large majority of these have found weak or no positive effects – including for a number of interventions widely believed to be effective based on quasi-experiments or preliminary trials (e.g., the LETRS teacher professional development program for reading instruction,<sup>7</sup> the Cognitive Tutor for teaching math<sup>8</sup>).

Systematic “design replication” studies comparing large, well-conducted randomized controlled trials with quasi-experiments in welfare, employment, and education policy also have found that many widely-used quasi-experimental methods produce unreliable estimates of program impact.<sup>9</sup>

- **Public officials seeking the few Top Tier interventions – backed by strong evidence of sizable effects – often cannot distinguish them from the many others on these sites.** Program providers frequently cite a listing of their intervention on one of these sites as proof that it is supported by strong evidence, when in fact the evidence is usually more preliminary. Public officials, most of whom are not researchers, often have no efficient way to assess the providers’ claims.

**How this initiative addresses the need:**

**It provides policy officials with a clear, validated resource to distinguish interventions meeting the Congressional Top Tier standard from everything else.**

- A. **The initiative’s expert Panel includes nationally-recognized, evidence-based researchers and former public officials.** They are: Jonathan Crane, Deborah Gorman-Smith, Denise Gottfredson, Ron Haskins, Lynn Karoly, Dan Levy, Larry Orr, Sean Reardon, and Howard Rolston (their titles/affiliations are [shown here](#)).
- B. **Through a systematic review process, the Panel identifies interventions meeting the “Top Tier” or “Near Top Tier” evidence standards.**
  - **“Top Tier” includes:** *Interventions shown in well-designed and implemented randomized controlled trials, preferably conducted in typical community settings, to produce sizable, sustained benefits to participants and/or society.* This standard includes a requirement for replication – i.e., demonstration of effectiveness in at least two well-conducted trials or, alternatively, one large multi-site trial.
  - **“Near Top Tier” includes:** *Interventions shown to meet almost all elements of the Top Tier standard, and which only need one additional step to qualify. This category includes, for example, interventions that meet all elements of the standard in a single site, and just need a replication trial to confirm the initial findings and establish that they generalize to other sites.* The purpose of this category is to help increase the number of Top Tier interventions, by enabling policy officials and others to identify particularly strong candidates for replication trials whose results, if positive, would provide the final element needed for Top Tier.
  - **The solicitation process, review criteria, and procedure for reporting results, are summarized in the [solicitation/review process](#).** This link also provides guidance on how to nominate an intervention for the panel’s review.
- C. **We report the Panel’s decisions on which interventions to identify as Top Tier and Near Top Tier twice per year, and briefly summarize Panel findings ([see results here](#)).**
- D. **Impartiality: The initiative is administered by the nonprofit, nonpartisan Coalition for Evidence-Based Policy, a national leader in evidence-based reform with no affiliation to any program.** The Coalition’s work, including the Top Tier Evidence initiative, has had an important impact on federal policy and enacted legislation (as [summarized here](#)), and been cited in the national [press](#) and numerous policy publications.

**Conclusion:**

**Currently, there are a few social interventions backed by strong, replicated evidence of effectiveness in preventing educational and workforce failure, child abuse, teen pregnancy, crime, and other problems that damage millions of American lives each year. The Top Tier Evidence initiative gives public officials a clear, validated tool to distinguish these Top Tier interventions from everything else, so that they can be put into widespread use.**

## References

- 
- <sup>1</sup> National Research Council and Institute of Medicine. (2009). *Preventing Mental, Emotional, and Behavioral Disorders Among Young People: Progress and Possibilities*. Committee on Prevention of Mental Disorders and Substance Abuse Among Children, Youth and Young Adults: Research Advances and Promising Interventions. Mary Ellen O’Connell, Thomas Boat, and Kenneth E. Warner, Editors. Board on Children, Youth, and Families, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press. Recommendation 12-4, page 371.
- <sup>2</sup> DeNavas-Walt, Carmen and Bernadette D. Proctor, U.S. Census Bureau, Current Population Reports, P60-249, *Income and Poverty in the United States: 2013*, U.S. Government Printing Office, Washington, DC, 2014. U.S. Census Bureau, [Official and National Academy of Sciences \(NAS\) Based Poverty Rates: 1999 to 2011](#), 2012. Kathleen Short, U.S. Census Bureau, HHES Division, *Estimating Resources for Poverty Measurement, 1993 – 2003*, 2005. Panel on Poverty and Family Assistance, National Academy of Sciences, *Measuring Poverty: A New Approach*, 1995, pp. 31-36. Christopher Wimer, Liana Fox, Irv Garfinkel, Neeraj Kaushal, and Jane Waldfogel, [Trends in Poverty with an Anchored Supplemental Poverty Measure](#), December 2013.
- <sup>3</sup> *The Nation’s Report Card: Trends in Academic Progress 2012*, NCES 2013-456, National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, 2013.
- <sup>4</sup> Cornman, S.Q., and A.M. Noel, *Revenues and Expenditures for Public Elementary and Secondary School Districts: School Year 2008–09 (Fiscal Year 2009)* (NCES 2012-313). U.S. Department of Education. Washington, DC: National Center for Education Statistics, 2011. Richard H. Barr, *Revenues and Expenditures for Public Elementary and Secondary Education, 1973-74* (NCES-76-140). U.S. Department of Health, Education & Welfare, National Institute of Education. Washington, DC: National Center for Education Statistics, 1976.
- <sup>5</sup> U.S. Census Bureau, Current Population Reports, 2014, op. cit., no. 2. This refers to inflation-adjusted income. It includes income from the economy (such as earnings) but not government transfers (such as Food Stamps). However, the evidence suggests that the overall story of income stagnation for the bottom 40% of households changes little even when one adjusts income for government transfers and other items that affect household living standards. Specifically, the Census Bureau’s alternative, National Academy of Sciences-based poverty measures make adjustments for government transfers, as well as factors such as state and local taxes, work expenses such as child care, out-of-pocket medical expenses, and geographic differences in housing costs. These adjustments change the poverty rate in any given year, as well as the composition of those in poverty, but do not change the overall trend in the poverty rate over time – i.e., little overall progress since the late 1970s. (The relevant citations are in endnote 1.) Although the National Academy-based poverty measures only apply to a subset of the bottom 40% of U.S. households, their corroboration of no meaningful improvement for that key subset suggest that similar findings would be obtained for the larger group.
- <sup>6</sup> John P. A. Ioannidis, “Contradicted and Initially Stronger Effects in Highly Cited Clinical Research,” *Journal of the American Medical Association*, vol. 294, no. 2, July 13, 2005, pp. 218-228. Mohammad I. Zia, Lillian L. Siu, Greg R. Pond, and Eric X. Chen, “Comparison of Outcomes of Phase II Studies and Subsequent Randomized Control Studies Using Identical Chemotherapeutic Regimens,” *Journal of Clinical Oncology*, vol. 23, no. 28, October 1, 2005, pp. 6982-6991. John K. Chan et. al., “Analysis of Phase II Studies on Targeted Agents and Subsequent Phase III Trials: What Are the Predictors for Success,” *Journal of Clinical Oncology*, vol. 26, no. 9, March 20, 2008.
- <sup>7</sup> Garet, Michael S. et. al., *The Impact of Two Professional Development Interventions on Early Reading Instruction and Achievement* (NCEE 2008-4030). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education
- <sup>8</sup> Campuzano, L., Dynarski, M., Agodini, R., and Rall, K. (2009). *Effectiveness of Reading and Mathematics Software Products: Findings From Two Student Cohorts—Executive Summary* (NCEE 2009-4042). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

---

<sup>9</sup> Howard S. Bloom, Charles Michalopoulos, and Carolyn J. Hill, "Using Experiments to Assess Nonexperimental Comparison-Groups Methods for Measuring Program Effects," in *Learning More From Social Experiments: Evolving Analytic Approaches*, Russell Sage Foundation, 2005, pp. 173-235. Thomas D. Cook, William R. Shadish, and Vivian C. Wong, "Three Conditions Under Which Experiments and Observational Studies Produce Comparable Causal Estimates: New Findings from Within-Study Comparisons," *Journal of Policy Analysis and Management*, vol. 27, no. 4, 2008, pp. 724-50. Steve Glazerman, Dan M. Levy, and David Myers, "Nonexperimental versus Experimental Estimates of Earnings Impact," *The American Annals of Political and Social Science*, vol. 589, September 2003, pp. 63-93.