THE UNEXAMINED MULTI-BILLION DOLLAR BUSINESS: CONGRESSIONAL AND STATE LEGISLATIVE PILOT PROJECTS

*William Wesley Patton**

Congress and state legislatures annually spend hundreds of billions of dollars on pilot projects to determine whether those small experiments warrant expansion into nation-wide or state-wide implementation. Unfortunately, there are currently few pre-approval methodologies and even fewer interim and final assessment instruments required before pilot project contracts are finalized. The result is that often the pilot project studies produce data and outcome measures that are not useful in determining whether to scale-up the programs. This study compares several federal and state pilot projects and suggests new pilot project requirements that will permit policy makers to better analyze the benefits of these experiments, more easily compare similar pilot projects, and determine longitudinal assessments of these legislative mandates in order to avoid economic waste.

* Professor and J. Alan Cook & Mary Schalling Cook Children's Law Scholar, Whittier Law School; Assistant Clinical Vol. Professor, UCLA David Geffen School of Medicine, Department of Psychiatry.

INTRODUCTION

Each year Congress and state legislatures approve hundreds of billions of dollars for experimental projects and programs.[1] These projects receive various labels, such as pilot programs, pilot projects, demonstration projects or programs, pilot experiments, and pilot studies.[2] The use of terminology to describe these legislative experiments is not consistent among different state jurisdictions or even within legislative committees in the same state.[3] For simplicity, this Article will use the term "pilot project" and "pilot program" to refer to all of these variously labeled statutory authorizations.[4]

---

[1] *See infra* text accompanying notes 14–20, for a discussion of the cumulative cost of pilot projects throughout the nation. Until recently, empirical research on the effectiveness of different types of funding sources for social services was not a primary focus of governmental grants. For instance, "[p]ublic health law research usually focuses on the substantive laws that effect health, not the accompanying funding law." Susan Mangold et al., *Public Health Law Research: Measuring the Impact of Sources and Types of Funding on Health Care Outcomes for Children in Foster Care in Ohio*, http://publichealthlawresearch.org/project/measuring-impact-sources-and-types-funding-health-care-outcomes-children-foster-care-ohio. Newer studies, however, are beginning to analyze whether "the source and/or type of funding, not just the amount of funding, impact[s] health outcomes." *Id.*

[2] Some use the term "pilot experiment" as a general category that includes other types of temporary or termed small-scale legislative trials such as "pilot programs" and smaller experiments termed "pilot projects." Raphaël Billé, *Action Without Change? On the Use and Usefulness of Pilot Experiments in Environmental Management*, SAPIENS, July 10, 2010, at 1, 2–3, *available at* http://sapiens.revues.org/index979.html. *See also* Edwin R. van Teijlingen & Vanora Hundley, *The Importance of Pilot Studies*, SOCIAL RESEARCH UPDATE, Winter 2001, *available at* http://sru.soc.surrey.ac.uk/SRU35.html; Lehana Thabane et al., *A Tutorial on Pilot Studies: The What, Why, and How*, BMC MED. RES. METHODOLOGY, vol. 10, 2010, *available at* http://www.biomedcentral.com/1471-2288/10/1.

[3] One of the main difficulties of studying pilot projects is the lack of consistent definition even within specific fields of study. For instance, one analysis of medical pilot projects found that there was "no formal methodological guidance as to what constitutes a pilot study," and that even journal editors that publish pilot studies "said that they had no publication policy." Gillian A. Lancaster et al., *Design and Analysis of Pilot Studies: Recommendations for Good Practices*, 10 J. EVALUATION CLINICAL PRAC. 307, 308–09 (2004).

[4] This Article will not analyze the largest funded research projects, often called "demonstration" projects, but rather will focus on the smaller and more numerous pilot projects. Demonstration projects also receive billions of federal dollars a year. For instance, the Department of Health and Human Services (HHS) alone has seen the number and budget for demonstration projects increase. In 2008, HHS provided 2,311 demonstration grants for a total of $1,052,778,855. U.S. DEP'T OF HEALTH AND HUMAN SERVS, HHS GRANT AWARDS FISCAL YEAR 2008, at 19 (2009), *available at* http://taggs.hhs.gov/AnnualReport/FY2008/documents/TAGGS_2008_Annual_Report.pdf. By 2010, the number of HHS demonstration grants increased to 2,755 for a total of $1,532,430. U.S. DEP'T OF HEALTH AND HUMAN SERVS.,

Pilot project[5] legislation can be characterized by the following variables: (1) the experiment has a limited temporal duration, usually between one to four years; (2) it usually has a limited geographical designation rather than being state-wide or nation-wide in implementation;[6] (3) it is designed to test a limited scope of new ideas, practices and/or policies;[7] and (4) it is often a political compromise that results from the failure to obtain sufficient legislative support for a permanent and/or more expansive legislative mandate. Reducing a fully implemented program to a pilot project is perceived in some cases as a way for a politician to maintain credibility in the face of opposition to the author's legislation. "[C]alling an initiative a pilot project has long been understood as giving a project permission to fail. The term lessens the accountability of the politician proposing the project since it is merely an experiment."[8]

Pilot projects potentially provide significant social benefits and methodologies that can increase the chances that new ideas can be successfully implemented.[9] Pilot projects can: (1) foster teamwork; (2) identify and train leaders and personnel for future projects; (3) provide realistic cost projections for expanding the pilot studies; (4) identify vari-

---

TAGGS 2010 ANNUAL REPORT 12 (2010), *available at* http://taggs.hhs.gov/AnnualReport/FY2010/documents/TAGGS_2010_Annual_Report.pdf. I will leave for a different article the important differences between accountability variables and concerns about demonstration projects.

5 "The term *pilot study* is used in two different ways in social science research. It can refer to so-called feasibility studies which are 'small scale version[s], or trial run[s], done in preparation for the major study' . . . . However, a pilot study can also be the pre-testing or 'trying out' of a particular research instrument . . . ." van Teijlingen & Hundley, *supra* note 2, at 2.

6 Billé, *supra* note 2, at 2. Although most pilot studies are used to determine whether full scale implementation should progress, sometimes so little is known about a topic that a pre-pilot study is required. For an excellent example of a small pre-pilot study see MICHAEL W. FINIGAN ET AL., CIVIL RIGHT TO COUNSEL, PHASE II PILOT STUDY: NEEDS ASSESSMENT AND COST ELEMENTS 1, *available at* http://www.npcresearch.com/Files/Civil_Right_to_Counsel_Phase_II_Pilot_Study_0610.pdf. For instance, that pre-large scale pilot project determined that the full-scale pilot project would take more time because "full court records are not publicly available" and they needed "more staff hours to obtain needed access to court records." *Id.* at 22.

7 Larry Hirschorn, *Campaigning for Change*, HARV. BUS. REV., July 2002, at 98, 103.

8 Karl Schulz, *Evaluating the Energy Independence and Security Act of 2007: Inclusions, Exclusions, and Problems with Implementation*, 38 ENVTL. L. REP. NEWS & ANAL. 10763, 10764 (2008).

9 Some have demonstrated the benefits of microsimulation analysis even prior to the design and implementation of small pilot projects. Microsimulation analysis "'involves modeling the impact of governmental programs at the level at which they are intended to operate' That is, microsimulation looks at the impact of government programs on the relevant individual decision units." David H. Greenberg et al., *Using Microsimulation to Help Design Pilot Demonstrations: An Illustration from the Canadian Self-Sufficiency Project*, 19 EVALUATION REV. 687, 692 (1995) (citations omitted). In effect, microsimulation analysis helps determine the best designs for pilot projects, and thus increases the chances a successful experiment.

ables necessary for expansion into statewide or nationwide programs;[10] (5) determine shared interests among disparate communities or constituents; (6) encourage innovation in stagnant systems in which status quo incentives thrive;[11] and (7) reduce economic waste by testing theory and implementation in a reduced environment rather than in a full-scale model.[12] Pilot projects can also expedite smaller governmental experiments "unencumbered by the red tape and budget line items" inherent in larger programs and the political difficulty of gaining consensus.[13]

Academics have provided almost no analysis of governmental pilot projects. Analyses and "[f]ull reports of pilot studies are rare in the research literature."[14] Thus, the various empirical variables inherent in pilot projects have been inadequately investigated, including: (1) their frequency and patterns of use over time; (2) whether the number and size of pilot projects cycle consistently with economic trends or whether they are relatively constant in number and/or the size of their budgets; (3) what is the success rate of pilot projects, and does their success rate justify that economic investment; (4) what quality controls have been established to judge and assess their success or failure; and (5) how have legislative term limits affected legislative sponsors' accountability re-

---

10 Of course, not all pilot projects seek data that is generalizable to larger scale implementation. There is a difference in purpose and design "between implementation science that aims to produce generalizable knowledge and quality improvement projects that aim to produce local knowledge with a focus on an organization's own delivery system and process." Ken Cheung & Naihua Duan, *Design of Implementation Studies for Quality Improvement Programs: An Effectiveness—Cost-Effectiveness Framework*, 104 AM. J. PUB. HEALTH 23, 23 (2014).

11 MARIA DAKOLIAS & JAVIER SAID, JUDICIAL REFORM: A PROCESS OF CHANGE THROUGH PILOT COURTS 2–4 (1999).

12 Pilot projects as investments permit a "limit [of] its losses in case of negative outcomes, while maintaining the profits resulting from more favorable scenarios." Eymen Errais & Jeffrey Sadowsky, *Valuing Pilot Projects in a Learning by Investing Framework: An Approximate Dynamic Programming Approach*, 35 COMPUTERS & OPERATIONS RESEARCH 90, 91–92 (2008), *available at* http://dl.acm.org/citation.cfm?id=1268209. The economic cost/benefit ratio of pilot projects can shift during the pilot timeline. "[I]ncreasing the learning coefficient while keeping the average technical volatility across pilot stages constant adds value and increases the incentive to invest in early pilot stages, while it does the opposite in later pilot stages." *Id.* at 110.

13 Amber Corrin, *Cleared for Takeoff: How Pilot Projects Are Fast-Tracking Mobile Tech*, FCW, Apr. 15, 2013, at 20, 21, *available at* http://fcw.com/articles/2013/04/11/dod-mobility-pilot-programs.aspx. In one Department of Defense small pilot project involving the investigation of hand held devices, Michael McCarthy, director of operations at the Army Brigade Modernization Command, stated that he "'never sat down and totaled how much we've spent on mobile projects.'" *Id.* at 22.

14 van Teijlingen & Hundley, *supra* note 2, at 3. "[T]he literature is somewhat limited in providing a well-defined structured methodology on how a pilot study should be used" since most studies focus on results/findings, not methodology. Raktim Pal et al., *Role of Pilot Study in Assessing Viability of New Technology Projects: The Case of RFID in Parking Operations*, 23 COMMC'NS ASS'N FOR INFO. SYS. 257, 259 (2008).

garding their authored pilot projects that may not conclude before their political terms expire?

This Article will begin to sketch the parameters of legislative pilot projects, determine the quantitative trends between the early 1990s and 2010, identify variables of pilot project success and failure, and isolate the most relevant variables in determining the assessment of pilot project results and factors which help predict pilot project success.

I. THE FREQUENCY, PATTERN, AND COST OF PILOT PROJECTS BETWEEN THE 1990S AND 2009–2010

One of the frustrations inherent in the study of pilot projects is the lack of databases that exist to track their frequency, cost, and results. Even if one attempts to analyze the etiology and evolution of a single bill giving rise to a pilot project, it may be almost impossible to determine basic data such as the actual pilot project budget. For example, the California Legislature passed the Performance and Results Act of 1993,[15] which created "pilot testing performance budgeting in four departments," and the pilot study was to be completed by 1996.[16] The full pilot project report, however, was not submitted in a timely manner, and the California Legislative Analyst's Office's report on the project merely noted that "Significant [r]esources [h]ave [b]een [i]nvested" in the pilot project, and that although it recommended that the Department of Finance estimate the costs of the pilot project, "no such estimate has been provided."[17] The Legislative Analyst's Office again recommended that "the department provide the Legislature a status report on the pilot program at the time of the budget hearings."[18] No report on the actual costs of the pilot project was ever produced,[19] but it was noted that many of the actual

---

15 1993 Cal. Stat. 3738 (Cal. 1993).

16 *1996-1997 Budget Analysis: Statewide Administrative Crosscutting Issues*, LEGIS. ANALYST'S OFFICE, http://www.lao.ca.gov/analysis_1996/a96hcc.html (last visited NOV. 15, 2014).

17 *Id.*

18 *Id.*

19 It is not unusual for pilot project budgets to be non-transparent. For instance, even though California Assembly Bill 1102 (February 25, 1999) created eight pilot projects to test environmental management systems, *see* A.B. 1102, 1999-2000 Reg. Sess. (Cal. 1999), *available at* http://www.leginfo.ca.gov/pub/99-00/bill/asm/ab_1101-1150/ab_1102_bill_19990706_chaptered.html, in none of the six annual reports issued between 1999 and 2001, nor in the final January 3, 2003 report, were the specific cumulative costs and savings of the pilot projects discussed. *See Archives, 1999–2003 Publications: CalEPA Agency Wide Environmental Management System*, CAL. ENVTL. PROT. AGENCY, http://www.calepa.ca.gov/EMS/Archives/ (last visited Nov. 15, 2014). "Only two pilots provided economic data on the costs and benefits of EMS implementation. Further, the data provided was limited especially when identifying costs of EMS implementation." ENVTL. MGMT. AND SUSTAINABILITY PROGRAM INNOVATION INITIATIVE, CAL. ENVTL. PROT. AGENCY, REPORT ON THE CAL/EPA ENVIRONMENTAL MANAGEMENT SYSTEM PROJECT 64 (2003), *available at* http://www.calepa.ca.gov/EMS/Archives/2003/LegReport/FinalReport.pdf [hereinafter CAL/EPA REPORT].

costs, like the costs that departments "absorb[ed]," such as staff time and consultants' costs, were beyond and in addition to the formal pilot project budget.[20]

In addition to the incomplete legislative reporting on pilot projects, neither the federal government nor any state legislature requires an annual compilation of the cumulative number, types, costs, or results of pilot projects. Research on pilot projects at this stage remains relatively ad hoc and methodologically challenged. However, the following analysis begins to sketch the scope, costs, strengths, and weaknesses of the current legislative pilot project process.

A. Trends in Federal Pilot Projects

The number of Congressional pilot projects increased substantially between the 1991–1992 Congressional term and the 2009–2010 term.[21] The results demonstrated a 68% increase in the number of funded pilot studies during that period:[22]

FEDERAL PILOT PROJECTS

	1991-1992	2009-2010	Percent Change
Pilot Project	81	132	+62.9%
Pilot Program	306	572	+86.9%
Total	387	704	+81.9%

TABLE 1

B. Trends in State Pilot Projects

Unlike federal pilot project legislation, there was not a consistent increase in the number of pilot projects in every state legislature during the same time period. In fact, although some jurisdictions realized large increases, other states approved fewer pilot projects in 2009–2010 than in the early 1990s.[23] Therefore, there was a 45% total increase in the

[20] CAL/EPA REPORT, *supra* note 19, at 50.

[21] The study consisted of a search of the terms "pilot project" and "pilot program" for two time periods of federal legislation, 1991–1992 and 2009–2010. *See*, THOMAS SEARCH BILL TEXT, LIBRARY OF CONG., http://thomas.loc.gov/home/LegislativeData.php?&n=BillText (last visited Nov. 15, 2014).

[22] In a subsequent paper, I intend to provide a more sophisticated analysis of pilot project trends that will track relevant year-to-year comparative variables such as: (1) which political party had a majority in Congress or in different state legislatures and (2) how various economic indicators such as inflation, the consumer price index, the stock exchange valuations, trade deficits, and the national debt affected willingness to fund pilot projects.

[23] This Article only compares the number of pilot projects and programs among six states: Oregon, Pennsylvania, Florida, Texas, Arizona, and California. In a future paper, I will provide a comparative analysis of all fifty states to determine whether there is a stronger connection with variables such as the geographical state regions, political party control, and

number of examined state pilot projects during that almost two-decade period.

STATE PILOT PROJECTS

	1990's	2009-10	Percent Change
Oregon	97	132	+36.0%
Arizona	32	28	-12.5%
Florida	124	113	-.088%
Pennsylvania	39	36	-.077%
Texas	39	128	+228%
California	88	172	+95.5%

TABLE 2

Although this is a limited sample of states' patterns regarding the constancy in the number of pilot projects during different decades, it is noteworthy that there were almost no significant reductions in the percentage of projects, but there were significant increases in some states in the number of pilot projects approved in 2009–2010. The total number of pilot projects among these seven states increased from 419 during the 1990s to 609 in 2009–2010, for a total increase of 48.8%. That increase was similar to the 45% increase in federal pilot projects for the same time period.

C. Estimated Cost of Pilot Projects

There are currently no estimates regarding the annual federal and state expenditures on pilot projects. The cost of Federal pilot programs obviously dwarfs expenditures by individual state pilot programs.[24] As

state economic conditions. It was not possible to conduct identical Westlaw searches for each of the six states because some states' electronic legislative records do not start until a few years after 1990. For the current study, I used the internet for the following relevant periods for each state: (1) Oregon: 1995–1996 vs. 2009–2010, *see Bills and Laws*, OR. STATE LEGISLATURE, https://www.oregonlegislature.gov/bills_laws (last visited Nov. 15, 2015); (2) Pennsylvania: 1991–1992 v. 2009–2010, *see Bills and Amendments*, PA. GEN. ASSEMBLY, http://www.legis.state.pa.us/cfdocs/legis/home/bills/ (last visited Nov. 15, 2014); (3) Florida: 1998–1999 v. 2009–2010, *see Bills for Regular Session*, FLA. HOUSE OF REPRESENTATIVES, http://www.myfloridahouse.gov/Sections/Bills/bills.aspx (last visited Nov. 15, 2014); (4) Texas: 1993 v. 2009, *see Direct Search*, LEGISLATIVE REFERENCE LIBRARY OF TEX., http://www.lrl.state.tx.us/legis/billsearch/lrlhome.cfm (last visited Nov. 15, 2014); (5) Arizona: 1997 v. 2009, *see Bill Search*, ARIZ. STATE LEGISLATURE, http://www.azleg.gov/bills.asp (last visited Nov. 15, 2014); and (6) California: 1991 v. 2009, *see Bill Search*, CAL. LEGISLATIVE INFO., http://leginfo.legislature.ca.gov/faces/billSearchClient.xhtml (last visited Nov. 15, 2014). Therefore, this research provides a thumbnail sketch of the quantitative changes in six different states between substantial time periods; however, it does not provide a complete empirical analysis of those trends.

[24] For instance, the federal Medicaid pilot projects were budgeted for $437 billion, and those pilot programs are expected to cost $32 billion more than expected. Rich Daly, *Reform*

indicated, *supra* Part I.A, there were 704 federal pilots in 2009–2010. These programs' costs ranged from tens of millions of dollars to billions of dollars per year for projects with durations between three and five years.[25]

The data, *supra* Part I.B., demonstrates that in 2009–2010 there were 609 pilot projects in the six states that were analyzed, or approximately 100 pilot projects per state. Assuming that those six states (located on the West Coast, East Coast, and in the South) are representative of the pilot project activity in other states throughout the country, one could estimate that in 2009–2010, states funded approximately 5,000 pilot projects and programs. Even if one assumes a very conservative cost of $1,000,000 per state pilot project, the estimated cost of state pilot programs is $5 billion annually. It is thus surprising that with so many billions of dollars at stake in pilot projects, this significant and continuing element of state and federal governmental fiscal activity has received so little attention.[26]

Update: Medicaid Pilot Projects Cost $32 Billion More than Expected, GAO Says, MODERN HEALTHCARE (July 23, 2013), http://www.modernhealthcare.com/article/20130723/NEWS/307239956; *see also* U.S. GOV'T ACCOUNTABILITY OFFICE, MEDICAID DEMONSTRATION WAIVERS: APPROVAL PROCESS RAISES COST CONCERNS AND LACKS TRANSPARENCY (2013), *available at* http://www.gao.gov/assets/660/655483.pdf.

25 For instance, expenditures for pilot programs for a single federal program, called the "Patient Protection and Affordable Care Act (PPACA) and Health Care and Education Reconciliation Act of 2010," included: (1) the Regionalized Systems for Emergency Care Pilots, Sec. 3504 was funded for "$24 million . . . annually for 2010–2014"; (2) the Healthy Aging, Living Well Pilot Program, Sec. 4202 was funded for $50 million for 2010–2014; (3) the Demonstration Projects to Address Health Professions, Sec. 5507 was funded for $85 million; (4) $25 million for Childhood Obesity, Sec. 4306; (5) $60 million for Alternative Dental Care pilots; (6) up to $100 million for Complex Diagnostic Laboratory demonstration; (7) $75 million for Emergency Psychiatric demonstration project; (8) $50 million for Tort Litigation project; (9) $25 million for At Home Demonstration Pilot; (10) $250 million for Nurse Education Demonstration Project; (11) $500 million for a Payment Adjustments For Home Health Care demonstration project. *See* CTR. FOR MEDICARE AND MEDICAID INNOVATION, PILOT PROGRAMS AND DEMONSTRATION PROJECTS: THE PATIENT PROTECTION AND AFFORDABLE CARE ACT (PPACA) AND HEALTH CARE AND EDUCATION RECONCILIATION ACT OF 2010 (2010). (12) $500 million for pilot projects in the Water Resources Development Act of 2013, S. 601. CONGRESSIONAL BUDGET OFFICE, COST ESTIMATE S. 601 WATER RESOURCES DEVELOPMENT ACT OF 2013, at 2 (2013), *available at* http://www.cbo.gov/sites/default/files/cbofiles/attachments/s601.pdf. For another discussion federal medical pilot programs, *see* ANDREW COHEN, CTR. FOR HEALTH LAW AND ECON., UNIV. OF MASS. MED. SCH., PATIENT PROTECTION AND AFFORDABLE CARE ACT (H.R. 3590) – PILOT PROGRAMS, DEMONSTRATION PROJECTS, AND GRANTS (2010). This cost estimate for pilot projects does not include the research budget for the Department of Defense, which is budgeted for $67.5 billion in 2014 alone. *See* U.S. Dep't. of Def., Fiscal Year 2014 Budget Request (2013), *available at* http://www.globalsecurity.org/military/library/budget/fy2014/sum/fy-2014-overview.pdf.

26 These cost estimates are only based on Congressional and state legislative pilot projects and do not contain the additional expenditures for federal and state executive and judicial pilot projects. For example, state courts through their administrative office of the courts fund hundreds of pilot projects annually. *See, e.g.*, *Request for Proposals Archive*, CAL. COURTS, http://www.courts.ca.gov/4834.htm#2012 (last visited Nov. 15. 2014). I have been

II. HIGHLY REGULATED FEDERAL PILOT PROJECTS

This Part compares Congress's administrative controls of U.S. Department of Transportation pilot projects with three state pilot projects regarding child welfare. This comparison will identify effective and ineffective procedures that help to: (1) control costs; (2) assure accountability and methodological soundness; and (3) increase success rates for pilot projects.[27]

A. United States Department of Transportation Pilot Projects

The U.S. Department of Transportation oversees some of the most expensive pilot programs in the nation. For instance, in 2011 the Federal Transit Administration (FTA) reported to Congress that it had a total of $43,181,353,804 in total capital costs for "New Starts and Small Starts Projects."[28] In its 2008–2009 report to Congress, the Department stated that $44,762,000 was appropriated, and in 2009, $45,700,000 was provided for FTA research projects.[29] Those projects ranged from small experiments, such as the Intelligent Transportation System Pilot Project funded for $452,212 a year for two years,[30] to much larger projects like

---

unable to locate financial literature directly discussing the relevance of project financing to governmental pilot projects. "Project finance involves a public or private sector sponsor investing in a single-purpose asset through a legally independent entity" and "a number of key characteristics of project finance . . . [are] high leverage and non-recourse debt . . . ." Marco Sorge, *The Nature of Credit Risk in Project Finance*, BIS Q. REV., Dec. 2004, at 91, 91–92. For a discussion of the benefits of project financing see *What Is Project Finance?*, EAGLE TRADERS, https://eagletraders.com/loans/loans_what_is_project_finance.php (last visited Nov. 15, 2014); *Project Finance Portal*, HARVARD BUS. SCH., http://www.people.hbs.edu/besty/projfinportal/index.htm (last updated June 23, 2010).

27 In 2010, Congress passed the Partnership Fund for Program Innovation pursuant to the Consolidated Appropriations Act of 2010, Pub. L. No. 111-117, 123 Stat. 3034. The Partnership Fund budgeted $37.5 million to help select and evaluate pilot projects. Memorandum from Jeffrey D. Zients, Deputy Director for Management, Exec. Office of the President, Office of Mgmt. and Budget, on Pilot Projects for the Partnership Fund for Program Integrity Innovation 1 (Oct. 19, 2010) *available at* http://www.whitehouse.gov/sites/default/files/omb/memoranda/2011/m11-01.pdf.

28 FED. TRANSIT ADMIN., ANNUAL REPORT ON FUNDING RECOMMENDATIONS: FISCAL YEAR 2011 NEW STARTS, SMALL STARTS, AND PAUL S. SARBANES TRANSIT IN PARKS PROGRAM: REPORT OF THE SECRETARY OF TRANSPORTATION TO THE UNITED STATES CONGRESS PURSUANT TO 49 U.S.C. 5309(k)(1) (2009), *available at* http://www.fta.dot.gov/documents/NewStarts_mainText_Jan_2010.pdf. "New Starts projects are those whose sponsors are requesting $75 million or more in New Starts funds, or anticipating a total capital cost of $250 million or more (49 USC 5309(d)). Small Starts projects are those whose sponsors are requesting less than $75 million in Small Starts funds and anticipating a total capital cost of less than $250 million (49 USC 5309(e))." *Id.* at 1.

29 FED. TRANSIT ADMIN., REPORT TO CONGRESS ON FY 2008 AND FY 2009 FTA RESEARCH PROJECTS, REPORT NO. FTA-TRI-20-2009.1, at 1 (2009) [hereinafter FTA Report].

30 *Id.* at 22. This pilot project had already been funded since 2006 for a total of $1,829,774. *Id.* Other examples of smaller pilot projects in 2009 included the Pilot Marin Headlands Shuttle in the Golden Gate National Recreation Area for $405,000 and the $100,000 allocated for the Planning Study to Evaluate a Pilot Partnership Transit System in the

the Human Services Transportation Coordination project, funded for $1,600,000 each year for two years, and receiving $6,400,000 in funding since 2006.[31]

B. Congressionally Mandated Assessments of Federal Transit Administration Projects

In 2005 Congress passed the Safe, Accountable, Efficient Transportation Equity Act: A Legacy for Users (SAFETEA-LU) by amending Section 5309(*l*)(1) of Title 49 United States Code to require the FTA to submit annual reports to Congress regarding its projects. The reporting requirements include an analysis of both the projects themselves, as well as an analysis of the quality of the work performed by the project contractors.[32] How effective have the accountability requirements of SAFETEA-LU been, and what lessons have we learned from its first six years of implementation?

First, ironically, we have learned that it takes several years before we can have sufficient data to analyze the effects of statutorily mandated accountability measures. In its first Before and After Studies report to Congress, the FTA explained that because project data is often not available until several years after the funding is approved, and because it "can then take several additional months for the project sponsor to synthesize and evaluate all the information collected over the period of a project's development," reports under new accountability standards will not be available for several years after their pilot program implementation.[33] In addition, after Congress promulgates accountability measures, it takes some time for the agency to draft and publish policy guidelines and a notice of proposed rulemaking for implementation by the project managers and contractors.[34]

Guilford Courthouse National Military Park. FED. TRANSIT ADMIN., ANNUAL REPORT ON FUNDING RECOMMENDATIONS: FISCAL YEAR 2012 CAPITAL INVESTMENT AND PAUL S. SARBANES TRANSIT IN PARKS PROGRAMS: REPORT OF THE SECRETARY OF TRANSPORTATION TO THE UNITED STATES CONGRESS PURSUANT TO 49 USC 5309(K)(1), at A6–A7 (2011), http://www.fta.dot.gov/documents/Annual_Report_main_text_FINAL_2_11_11(1).pdf.

31 FTA Report, *supra* note 29, at 28.

32 Title 49 U.S.C. §5309(g)(2)(C) (2013), requires Before and After Studies of each of the projects, and 49 U.S.C. § 5309(1)(2) (2013), requires Contractor Performance Assessment Reports.

33 Letter from James S. Simpson, Adm'r, U.S. Dep't of Transp., Fed. Transit Admin., to The Honorable Richard C. Shelby, Chairman, Comm. on Banking, Hous., and Urban Affairs, U.S. Senate (Aug. 10, 2006), *available at* http://www.fta.dot.gov/documents/06_Before_and_After_-_Final_Letter.pdf.

34 *Id.* There is also an ongoing process of promulgating and providing notice for changes in the transparency and reporting requirements for both the federal agency and the prospective and selected contractors. For examples, see, e.g., Dep't of Transp., Fed. Transit Admin., Notice of Availability of Proposed Guidance for New Starts/Small Starts Policies and

Second, experience in implementing the accountability standards led to substantial modifications of those standards. The changes have both a positive and a problematic impact. First, implementation led to several cost/benefit conclusions that resulted in a more nuanced regulatory scheme. For instance, between 2007 and 2008, regulations were modified according to the size of the project award so that smaller projects required a lesser scale of economy regarding the cost and scope of reporting rules. In 2008 the two-category approach of classifying projects as "New Starts" or "Small Starts" was modified to a three-tier system by adding "Very Small Starts" projects.[35] The 2008 report explained that "[a] subset of very simple and low cost transit projects, termed "Very Small Starts" projects, will be evaluated and rated using an even more simplified process."[36] A second modification to the assessment instruments was to change from a system that gave each variable identical weight to one that provided "comparable but not necessarily equal weights."[37] For instance, the new system provided the following weights: "cost effectiveness, 20 percent; land use, 20 percent; economic development, 20 percent; mobility improvements, 20 percent; operating efficiencies, 10 percent; and environmental benefits, 10 percent."[38] Although these changes may have reduced excessive administrative costs for smaller projects and better refined the values of assessment criteria, those changes have made it almost impossible to conduct a longitudinal examination of the effectiveness of the assessment instruments themselves or relative success and failure of different types of pilot projects over a defined period of time because of the paucity of required reported data.

In addition to the nuanced changes in assessment criteria, we have learned many lessons from the individual FTA projects analyzed under those criteria. In this sense, the assessment mechanisms have been extremely important in shaping future assessment instruments, not just in the FTA, but also in generalizing to other federal and state regulatory

---

Procedures and Request for Comments, 74 Fed. Reg. 23,776 (May 20, 2009); Final Guidance on New Starts/Small Starts Policies and Procedures, 73 Fed. Reg. 46,352 (Aug. 8, 2008).

35 *Compare* FED. TRANSIT ADMIN., ANNUAL REPORT ON FUNDING RECOMMENDATIONS: PROPOSED ALLOCATIONS OF FUNDS FOR FISCAL YEAR 2007, app. at B-3–B-24 (2006), *and* FED. TRANSIT ADMIN., BEFORE AND AFTER STUDIES OF NEW STARTS PROJECTS: REPORT TO CONGRESS (2007) [hereinafter BEFORE AND AFTER STUDIES 2007], *with* FED. TRANSIT ADMIN., ANNUAL REPORT ON FUNDING RECOMMENDATIONS: PROPOSED ALLOCATIONS OF FUNDS FOR FISCAL YEAR 2008, app. at B-3–B-26 (2007), [hereinafter FY 2008 NEW STARTS]e

36 FY 2008 NEW STARTS, *supra* note 35, at B-5. Very Small Starts projects require funding of "less than $50 million (including all project elements) and less than $3 million per mile, exclusive of rolling stock." FED. TRANSIT ADMIN. ANNUAL REPORT ON FUNDING RECOMMENDATIONS: FISCAL YEAR 2011 app. at B-5 (2010) [hereinafter FY 2011 NEW STARTS]e *available at* http://www.fta.dot.gov/12304_11092.html.

37 FY 2011 NEW STARTS, *supra* note 36, app. at B-3.

38 *Id.*

schemes. For example, in the initial reports in 2007, several factors made project evaluation extremely difficult. First, in a report on the Utah Transit Authority—Medical Center Extension of the TRAX Light Rail System project, two serious flaws in the project system were noted: (1) organizations other than the FTA created predictive models which did not meld with FTA criteria; and (2) those involved in the project did not retain critically important data necessary for project evaluation.[39] The "inconsistency and lack of archived data was troublesome and made an accurate comparison [between predicted and actual ridership] over time impossible."[40] We can generalize these findings to all pilot project assessment instruments by requiring standardization of terminology and the archiving of all specifically defined relevant data.

Justice Oliver Wendell Holmes argued that judges, by building an experiential base during several years of judging, could replace jurors in deciding negligence issues.[41] In effect, Justice Holmes argued that judges could determine how the world operates and from that base of knowledge could develop generalizations or customs regarding how the world reasonably should operate. By analogy, those in charge of assessing pilot projects can develop a sense of what predictions within proposals are realistic or unrealistic. The FTA report on the Tri-County Metropolitan Transportation District of Oregon project serves as an excellent example.[42] Several of the predictions in the "Before" study proposal were not realized.[43] For instance, the predicted number of passengers who would use the rail project was substantially greater than the actual number of riders; the number of jobs created in the sector was 27% less than predicted; the predicted percentage of riders who would be commuters was substantially higher than the actual ridership; the park and ride estimates were excessive; and the actual rail speeds were much lower than those predicted.[44] Because these inaccurate predictions affect Congress's and the FTA's decision regarding whether to proceed with a

---

39 BEFORE AND AFTER STUDIES 2007, *supra* note 35, at 5–8.

40 *Id.* at 8.

41 *See* Baltimore & O.R. Co. v. Goodman, 275 U.S. 66, 69–70, 48 S. Ct. 24, 25 (1927) (deciding whether a driver had a duty to exit the car at a railroad crossing where his view of oncoming trains was obscured to determine whether it is safe to cross the tracks—colloquially termed the "stop, look, and listen rule"). Justice Holmes stated that although the "question of due care very generally is left to the jury . . . when the standard is clear it should be laid down once for all by the Courts." *Id.* at 70. Holmes' "stop, look, and listen" court-formed negligence rule was clarified in *Pokora v. Washbash Ry. Co.*, 292 U.S. 98, 54 S. Ct. 580 (1934), in which the court returned to the traditional view that "questions of care [are left] to the jury to be decided under the broad, unelaborated standard of negligence," not as absolute standards set by judges. Trevino v. Union Pac. R.R. Co., 916 F.2d 1230, 1235 (7th Cir. 1990).

42 *See* FED. TRANSIT ADMIN., U.S. DEP'T OF TRANSP., BEFORE AND AFTER STUDIES OF NEW STARTS PROJECTS (2008) [hereinafter BEFORE AND AFTER STUDIES 2008].

43 *See, e.g.*, *id.* at 5.

44 *See id.*

pilot project, the government may require more accurate predictions or additional assessment controls to combat unrealistic estimates.[45] By increasing the accuracy of proposal predictions, agencies can assure more accurate data in deciding whether to fund pilots, thus increasing the chance of success for selected proposals.

C. *Congressionally Mandated Assessments of Federal Transit Administration Selected Project Contractors*

SAFETEA-LU amendments to 49 U.S.C. § 5309(*l*)(2) require the FTA to file a "Contractor Performance Assessment Report" with Congress annually.[46] The purpose of the contractor estimates is to help increase "the consistency and accuracy of cost and ridership estimates made by each contractor."[47]

The Contractor Performance Assessment reports identified a number of themes in the reliability of contractor project predictions.[48] For example, in comparing predicted ridership with actual ridership, the FTA discovered the following recurring patterns: (1) ridership forecasts for initial projects are much less accurate than for expansions of existing

---

[45] If these types of inaccurate predictions are observed in a substantial number of pilot projects, those involved in choosing which pilot projects to fund may require modifications to project projections before the funding is approved, rather than learning after the fact that "actual capital costs were higher than the predicted costs, while the actual ridership was lower than predicted ridership"—or a result that cost more and delivered much less than promised. BEFORE AND AFTER STUDIES 2008, *supra* note 42, at 8. In the Tren Urbano Heavy Rail Project, the Puerto Rico Highway and Transportation Authority predicted a ridership of 114,500 that was only 26,900, and noted that the error was due to the following misconceptions: (1) overconfidence that riders would give up automobiles for rail travel; (2) the assumed flat fare was substantially less than the actual fare; (3) the predicted time advantages for rail travel was lower than actual travel time; and (4) the projected population growth of 19 percent was actually only 5.4 percent. *See* BEFORE AND AFTER STUDIES 2008, *supra* note 42, at 8. Another assessment found that programs receiving multiple sources of funding made it "difficult to isolate the impacts" of the different programs. Letter from Mary E. Peters, U.S. Sec'y of Transp., to The Honorable Christopher J. Dodd, Chairman, Comm. on Banking, Hous., and Urban Affairs, U.S. Senate (Nov. 2, 2007), *available at* http://www.fta.dot.gov/documents/5310_Report_for_Website.pdf.

[46] *See* 49 U.S.C. § 5309(*l*)(2)(A) (2013).

[47] *Id.* Although this statute does not require contractor assessments in all pilot projects and instead focuses on larger New Start projects, the lessons learned from these assessments could also be applied to pilot projects. In its first contractor assessment report to Congress, the FTA stated that it would not conduct contractor assessments for "Very Small Starts [projects] since such projects are expected to be pre-qualified based on meeting certain easily calculated ridership and cost criteria and not based on an extensive analysis by contractors." FED. TRANSIT ADMIN., U.S. DEP'T OF TRANSP., CONTRACTOR PERFORMANCE ASSESSMENT REPORT 2 (2006), *available at* http://www.fta.dot.gov/documents/06_Contractor_Performance_Assessment_Report.pdf.

[48] *See* FED. TRANSIT ADMIN., U.S. DEP'T OF TRANSP., CONTRACTOR PERFORMANCE ASSESSMENT REPORT (2007) [hereinafter 2007 REPORT], *available at* http://www.fta.dot.gov/documents/CPAR_Final_Report_-_2007.pdf.

projects;[49] (2) downtown people-movers have poor ridership forecasts;[50] (3) variations in projected and actual frequency of service directly affect ridership predictions;[51] and (4) overestimation of employment and population trends often seriously affect ridership projections.[52]

These pilot project result patterns enable reviewers of new pilot project proposals to require that estimates take into consideration variables that consistently overestimate ridership and skew cost-benefit analyses away from properly balancing costs with realistic estimates of social benefit. More methodologically sound proposals will produce three major effects: (1) Congress, state legislatures, and administrative agencies will approve fewer unrealistic proposals; (2) more resources will be available to more accurately predict pilot project results; and (3) the percentage of pilot projects that result in success as defined in their proposals should increase, thus reducing economic waste.

D. An Overview of Three FTA Pilot Projects

One might expect that with the stringent, congressionally mandated accountability requirements that most FTA pilot projects would be successful in relation to project projections. Because the statutory requirements were not promulgated until 2005, however, and because almost no pilot projects were finalized until 2011, it is still too early to determine the effectiveness of those administrative controls. In addition, because pilot projects are merely experiments by definition, the predicted success rate is likely to be much lower than for expansion or extension of existing successful programs.[53] The following three FTA pilot projects demonstrate some of the assumptions and predictions that have resulted in project failure. These three pilot project examples range from one that required almost no reliability standards, no assessment or measurements of success, and no formal reporting mechanism to a project that was well defined by data-based outcome measures.

1. Debt Service Reserve Pilot Program

What if the FTA created a debt service pilot project, but no one applied for the debt service? "The Debt Service Reserve Pilot Program (DSRPP) seeks to provide credit support and reduce overall project expenses by providing public transportation project sponsors that plan to issue bonds in the capital markets with the additional flexibility to use

---

[49] *See id.* app. at 39.
[50] *Id.*
[51] *See id.* at 41.
[52] *See id.* at 110.
[53] *See, e.g.*, *id.* app. at 39.

federal grants to establish and maintain a debt service reserve fund."[54] Not one transit agency applied for the program because the user projections and the attractiveness of the financial instruments were based upon four oversights: (1) there are very few potential applicants because only transit agencies that issue debt would be interested; (2) there are existing competitive programs; (3) the financial product is not perceived as a significant benefit over other alternatives; and (4) there is competition with broker commissions.[55]

The DSRPP is an example of a pilot project that should have failed the "smell test" because the flaws in the projections should have been obvious to those with expertise in market and financial instruments. It is unlikely that Congressional assessment and accountability mandates would correct this type of mistake, which appears to be a product of failing to have appropriate experts review the underlying suppositions of the pilot concept, rather than isolating consistently unreliable predictability criteria. But had the project proposal been required to explicitly articulate outcome measures, methods and standards of assessment, and report requirements, the project might have been either rejected or amended prior to implementation.

2. Cooperative Procurement Pilot Program

The Cooperative Procurement Pilot Program involves an example of a program with low to moderate accountability, assessment, and reporting requirements. "Section 166 of the Transportation, Treasury, and Independent Agencies Appropriations Act, 2004 directed the FTA to establish the Cooperative Procurement Pilot Program (CPPP) to determine the benefits of encouraging cooperative procurement of major capital equipment."[56] The CPPP is an excellent example of a modestly structured pilot program with limited methodologically sound accountability measures that failed to deliver expected results. Success or failure, however, should not be the litmus test for pilot projects where one of the primary goals of these experiments is to determine feasibility of new ideas and procedures, not to assure implementation. As long as the project produces a clear analysis of the variables that interfered with the predicted results, the study can be useful in helping to predict future pilot

---

54 Letter from Mary E. Peters, U.S. Sec'y of Transp., to The Honorable James L. Oberstar, U.S. House of Representatives, Comm. on Transp. and Infrastructure, at 1 (Sept. 22, 2008), *available at* http://www.fta.dot.gov/documents/Debt_Service_Reserve_Pilot_Program__-Final_for_Web.pdf (discussing the DSRPP and 49 U.S.C. § 5323(d)).

55 *Id.* at 2.

56 FED. TRANSP. ADMIN., U.S. DEP'T OF TRANSP., REPORT TO CONGRESS ON THE RESULTS OF THE COOPERATIVE PROCUREMENT PILOT PROGRAM 1 (2010). *See generally* Consolidated Appropriations Act of 2004, Pub. L. No. 108–199, § 166, 118 Stat. 309.

projects that may be more successful after implementing those modifications.

The enabling legislation for the CPPP included the following required assessment and accountability measures: (1) cooperative development of specifications;[57] (2) joint procurement proposals that may include multiple agencies; (3) proportionality of agency procurement expenses; (4) technical outreach assistance by the Secretary of the FTA; and (5) a report that must include the savings realized and the benefits discovered from the implementation of cooperative procurement.[58]

The pilot project created five consortia of transportation authorities in different states, with a goal of determining whether they could reduce the cost of purchasing new buses if they pooled their purchases—much like consumers attempting to gain a fleet discount.[59] The results demonstrated that cooperative procurement of buses by transit authorities neither produced savings in purchases nor provided any meaningful benefits. "Only three pilot projects successfully completed bus purchases, and the results for these pilots were mixed."[60] The experiment of providing more federal financing for bus purchases proved inadequate to incentivize city transit authorities to use cooperative procurement, and interviews with bus manufacturers revealed that cooperative procurement does not realize savings because each city transit authority requires different bus characteristics, thus eliminating scales of economy through standardization.[61] The report does not specify why these weaknesses were not reasonably knowable prior to pilot project implementation, but this new information can be factored into cost-benefit analyses to determine whether or not to fund analogous pilot studies.

---

57 One of the most successful federal transportation pilot programs, the Nonmotorized Transportation Pilot Program provided $100 million to four cities to study methods of increasing bicycle usage. *See* FED. HIGHWAY ADMIN., U.S. DEP'T OF TRANSP., REPORT TO THE U.S. CONGRESS ON THE OUTCOMES OF THE NONMOTORIZED TRANSPORTATION PILOT PROGRAM SAFETEA-LU SECTION 1807, at 1 (2012). In order to assure comparability of research methodology among the four research sites, the pilot program "created a Working Group (WG) composed of representatives from the administrating agencies in each of the communities . . . [which] held regular teleconferences and annual meetings to discuss progress and challenges and coordinate efforts across the pilot communities." *Id.* at 3. Creating the WG helped "resolve technical issues and implement a common methodology for data collection and analysis." *Id.* at 6. "The key successes of the WG have been to develop a collaborative approach to data collection and evaluation, to maintain a coordinated national program, to establish consistent and credible reporting of results, and to share the progress of the Pilot Program to multiple audiences throughout the years of its existence." *Id.* at 7.

58 *See* FED. TRANSP. ADMIN., *supra* note 56, at 2–3.

59 *See id.* at 4.

60 *See id*. at 27 ("Only three pilot projects successfully completed bus purchases, and the results for these pilots were mixed.").

61 *See id.* at 27–28.

3. The Remote Infrared Audible Signage Pilot Program

"RIAS [Remote Infrared Audible Signage] technology is a remote infrared communication system designed to eliminate barriers to accessibility for people who are blind, visually impaired, or cognitively or developmentally disabled by assisting them with both orientation and mobility (O & M)."[62] After a competitive proposal process, the FTA selected the Central Puget Sound Regional Transportation Authority for the pilot project of installing RIAS technology.[63] The total budget was $2.283 million to install 303 transmitters in six transit stations.[64]

The legislation creating the pilot project listed four specific required assessments: (1) the effect of the pilot program on accessibility of the public transit system; (2) the effect of the program on public transportation operators and passengers; (3) the effect public transit has on people with visual, cognitive, and learning disabilities; and (4) the effect of the program on the quality of life of the target population.[65] Unlike the first two pilot projects discussed,[66] the RIAS pilot project involved a multifaceted methodological assessment plan that included: (1) surveys, (2) focus groups, and (3) qualitative interviews.[67] This inclusive, methodological fact-investigation permitted the evaluation of both quantitative and qualitative results and made the decision of whether to expand the pilot into a full program much more accurate and nuanced.[68]

The pilot data demonstrated that the RIAS system, through its methodological assessment plan, increased transit accessibility for the target population, and increased users' confidence and efficiency in using the system.[69] The interviews discovered, however, that some of the placements of the RIAS transmitters were much less meaningful than others, and that too few were placed in strategic locations, especially in ways

---

62 MARGARET PETRELLA ET AL., U.S. DEP'T OF TRANSP., REMOTE INFRARED AUDIBLE SIGNAGE PILOT PROGRAM: EVALUATION REPORT 1 (2009).

63 *Id.*

64 *Id.*

65 *Id.*

66 Letter from Mary E. Peters to The Honorable James L. Oberstar, *supra* note 54; FED. TRANSP. ADMIN., *supra* note 56.

67 PETRELLA ET AL., *supra* note 62, at 2.

68 *See id.* at 2–3.

69 *Id.* at 2. "[A]necdotal evidence can 'vividly complement' statistical evidence of discrimination." Jeffrey M. Hanson, *Hanging by Yarns?: Deficiencies in Anecdotal Evidence Threaten the Survival of Race-Based Preference Programs for Public Contracting*, 88 CORNELL L. REV. 1433, 1448 (2003) (quoting Concrete Works of Colo., Inc. v. City of Denver, 36 F.3d 1513, 1520 (10th Cir.1994)). "Convincing statistical and anecdotal evidence can make for a 'potent' combination, because anecdotal evidence can bring 'cold numbers convincingly to life.'" *Id.* (quoting Int'l Bhd. of Teamsters v. United States, 431 U.S. 324, 339 (1977); Coral Constr. Co. v. King Cnty., 941 F.2d 910, 919 (9th Cir. 1991)). *See also* Dawn Freshwater et al., *Qualitative Research as Evidence: Criteria for Rigour and Relevance*, 15 J. OF RES. NURSING 497, 498 (2010).

that provided sufficient information about approaching buses.[70] Yet because of the pilot study's short duration, the data regarding projections about the effect of the RIAS system on ridership provided insufficient evidence to permit an accurate prediction on the effects of an expanded system.[71] Based upon the pilot project data, the report recommended: (1) a continuation of the pilot project in order to obtain longitudinal data; (2) further analysis by a group of experts to determine whether or not other technology might be better than the RIAS system in helping the target population with their transportation needs and confidence; and (3) an expanded publication of the pilot project information so that more users might participate in the pilot project.[72]

The RIAS pilot project encapsulates most of the important variables necessary for designing a pilot project that will yield sufficient depth and breadth of data from which a reasoned decision about whether to expand the pilot to a full project can be made.[73] However, the RIAS pilot involved an experiment in which much of the data to be discovered was objective––whether the transponders worked and provided better access to the transit for the target audience, or whether there were a sufficient number of transponders and if they were strategically placed—and only a few questions concerned subjective qualitative criteria, such as riders' satisfaction with the system or the effect on riders' confidence in using the transit system.[74] Even so, the methodology and assessment instruments used in the RIAS pilot project comprise some of the most sophisticated analytical variables that offer benefits for future project accountability standards.

III. STATE LEGISLATIVE PILOT PROJECTS: POORLY REGULATED AND NON-STANDARDIZED

Unlike the congressionally required assessment and accountability standards in federal pilot projects, such standards are rarely mandated in state pilot projects. This Part compares two state pilot projects that were designed to determine whether previously confidential child dependency, child abuse, and neglect court hearings should be opened to the press and public. As will be demonstrated, pilot projects that begin with a substantial disagreement between constituents and users regarding the social utility of the desired pilot project need to be designed somewhat differently than pilots that are less concerned with political and human rights

70 PETRELLA ET AL., *supra* note 62, at 2.

71 *Id.* at 3.

72 *See id.* at 29–31.

73 *See id.*

74 *See id.* at 27–28.

issues, since anecdotal and subjective opinions might overwhelm data-based outcome measurements.

The debate over whether civil child dependency courts that litigate whether children have been abused should be presumptively open to the press and public has been vibrant for several decades. Historically, dependency court proceedings were confidential in a supermajority of states.[75] In the late 1990s, however, the debate shifted from one almost exclusively based on politics, ad hoc personal observations, and speculation to a debate based on system accountability. The debate began to shift after the Minnesota Supreme Court published its "empirical" study of its Open Court Pilot Project.[76] Although the Minnesota pilot project, which was researched and written by the National Center for State Courts (NCSC), found that opening the courts had little impact on improving the public's knowledge of dependency courts and that the predicted systemic improvements did not materialize, the result that drew the most attention was that opening the proceedings to the press and public did not place abused children at further risk of being psychologically harmed from the publicity.[77] The NCSC pilot project report had a cascading effect on the open/closed court debate. First, the largest organization of juvenile court judges in the United States, the National Council of Juvenile and Family Court Judges, issued a policy statement supporting the opening of child dependency courts to the press and public.[78] Second, as the claims of potential psychological harm to children lost vitality in the face of the Minnesota pilot results, states such as California and Connecticut proposed open court legislation.[79] Third, writers relied on

---

75 I have been researching and writing on the topic of open versus confidential juvenile dependency cases for approximately fifteen years. For background data on this debate, and for extensive discussions about the Minnesota and Connecticut Pilot Projects, see William W. Patton, *Pandora's Box: Opening Child Protection Cases to the Press and Public*, 27 W. ST. U. L. REV. 181 (1999); William W. Patton, *Revictimizing Child Abuse Victims: An Empirical Rebuttal to the Open Juvenile Dependency Court Reform Movement*, SUFFOLK U. L. REV. 303 (2005) [hereinafter Patton, *Revictimizing Child Abuse Victims*]; William W. Patton, *When the Empirical Base Crumbles: The Myth that Open Dependency Proceedings Do Not Psychologically Damage Abused Children*, 33 LAW & PSYCHOL. REV. 29 (2009) [hereinafter Patton, *When the Empirical Base Crumbles*]; William W. Patton, *The Connecticut Open-Court Movement: Reflection and Remonstration*, 4 CONN. PUB. INT. L.J. 8 (2005).

76 MINN. SUP. CT. ADVISORY COMM., INTRODUCTION TO FINAL REPORT OF NATIONAL CENTER FOR STATE COURTS (Vol. 1, 2001) [hereinafter NCSC REPORT V. 1]. *See also* FRED L. CHEESMAN, II, KEY FINDINGS FROM THE EVALUATION OF OPEN HEARINGS AND COURT RECORDS IN JUVENILE PROTECTION MATTERS (Vol. 2, 2001) [hereinafter NCSC REPORT V. 2].

77 NCSC REPORT V. 2, *supra* note 76, at 16, 32.

78 NAT'L COUNCIL OF JUVENILE AND FAMILY COURT JUDGES, RESOL. NO. 9, IN SUPPORT OF PRESUMPTIVELY OPEN HEARINGS WITH DISCRETION OF COURTS TO CLOSE (2005), *available at* http://www.ncjfcj.org/sites/default/files/resolution%2520no.%25209%2520open%2520hearings.pdf.

79 The California Legislature debated three successive open court measures; however, none of them received the necessary political support and/or votes to pass. *See, e.g.*, S.B. 1391, 1999-2000 Reg. Sess. (Cal. 2000). In Connecticut, two open court bills failed to pass,

the results as gospel for the conclusion that open courts do not harm abused children.[80] And finally, the Arizona Legislature commissioned its own open pilot project study that was modeled upon the methodology of the Minnesota pilot study.[81]

But the inevitable occurred. The methodology, assessment instruments, and results of the Minnesota pilot project were subjected to serious scholarly attack, and the person in charge of that pilot project report admitted the report's methodological flaws while testifying under oath in a juvenile court.[82] First, it became evident that the NCSC had been forbidden from interviewing the children and parents who were appearing in the child dependency proceedings because those interviews might harm the abused children.[83] The irony of the Pilot Advisory Committee permitting abused children to undergo in-court examination before strangers and the press, but not permitting questioning in a much less threatening environment by the NCSC staff, impeached the pilot's findings that children are not harmed by merely discussing their abuse or neglect with researchers in a controlled and safe environment. In addition, the failure to interview those in the best position to know how the open court process affected the abused children—the children themselves and their parents—demonstrated the speculative nature of the conclusion regarding the children's safety in open courts. But the methodological flaws ran deeper. The assessment instruments were not designed to even register examples or evidence of harm to the abused children. Rather, the instruments only measured "extraordinary harm," a term that was never defined in the report and which was not defined for those who completed the pilot project surveys.[84] Another major methodological flaw was that the study did not take into consideration the psychological and psychiat-

---

and a third bill created a pilot project to be discussed. *See infra* note 90 and accompanying text.

80 *See* Barbara W. Stack, *The Dangers of Identifying Children*, THE QUILL, Sept. 2002, at 32; Barbara White Stack, *Edited Transcript of Comments*, 4 CONN. PUB. INT. L.J. 24, 25 (2004); Kathleen Blatz, *Transcript of Remarks*, 4 CONN. PUB. INT. L.J. 24, 29 (2004); Sara VanMeter, *Public Access to Juvenile Dependency Proceedings in Washington State: An Important Piece of the Permanency Puzzle*, 27 SEATTLE U. L. REV. 859, 879 (2004); CHRISTINA D. GHIO, TESTIMONY OF THE CENTER FOR CHILDREN'S ADVOCACY IN SUPPORT OF RAISED BILL NO. 555: AN ACT CONCERNING JUVENILE COURTS (2004); DIONNE MAXWELL ET AL., NAT'L COUNCIL OF JUVENILE AND FAMILY COURT JUDGES, TO OPEN OR NOT TO OPEN: THE ISSUE OF PUBLIC ACCESS IN CHILD PROTECTION HEARINGS (2004).

81 GREGORY B. BROBERG, ARIZONA OPEN DEPENDENCY HEARING PILOT STUDY (2006).

82 *See* Patton, *Revictimizing Child Abuse Victims*, *supra* note 75, at 309–16; Patton, *When the Empirical Base Crumbles*, *supra* note 75, at 30; Transcript of Proceedings, *In re* San Mateo County Human Services Agency (2005) (on file with author) [hereinafter Transcript].

83 Transcript, *supra* note 82, at 16.

84 *Id.* at 34, 45–46. When asked on the witness stand to define the term "extraordinary" harm, Dr. Cheesman stated not only that he chose that standard, but also that he guessed that it meant a case in which "you're able to demonstrate embarrassment or psychological trauma." *Id.* at 45–46.

ric evidence that trauma to the abused children from public disclosure, embarrassment, guilt, and shame might not manifest for weeks or months after the open court hearing, and that longitudinal surveys would be required to validate a conclusion that open hearings do not harm abused children.[85] Finally, the NCSC pilot study did not even survey the mental health experts in the child dependency system regarding whether or not the open hearings could or did harm the abused child victims.[86] Based upon the methodological weakness in the Minnesota pilot project report, the National Council of Juvenile and Family Court Judges issued a report that concluded:

> The NCSC report and its findings are now widely referenced by proponents for open hearings as supporting the view that open hearings do not produce the negative effects that have been argued for by opponents to this practice . . . . [A] number of methodological and other design flaws have been identified in the study by other researchers in this area that may further limit the scope and applicability of these findings to other jurisdictions.[87]

Unlike the Congressionally defined assessments of FTA pilot projects, there were no statutorily-mandated assessment or methodological standards imposed on the Open Court Pilot Project. Had such controls been implemented in the design and approval of the study, the methodological flaws and resulting impeachment of the study may have been eliminated. In addition, the flawed conclusion that open hearings do not harm abused children would not have solidified from opinion to supposed empirical truth, and some of the open court legislation from after the Minnesota pilot study might have resulted in a greater percentage of failure, or a closer inspection of the issue of harm to abused children, and mandatory prophylactic protections for children might have been implemented.[88]

---

85 Patton, *Revictimizing Child Abuse Victims*, *supra* note 75, at 312–15.

86 Transcript, *supra* note 84, at 44.

87 DIONNE MAXWELL ET AL., *supra* note 80, at 12–14, 15 n.26 (2004).

88 For a discussion of the difficulties of designing an accurate pilot project assessment tool for programs that involve multiple samples, see the discussion of the assessment of the Florida coordinated school health pilot project, that identified the following assessment problems: (1) the start of a pilot project before the finalization of the assessment instruments; (2) the failure to establish uniform goals and assessment criteria among multiple sites; (3) the problems inherent in setting assessment goals where no prior baseline data has been accumulated for comparison; and (4) the lack of pairings between comparison and control groups. Robert M. Weiler et al., *Evaluation of the Florida Coordinated School Health Program Pilot Schools Project*, 73 J. SCH. HEALTH 3, 4 (2003).

Once the Minnesota report was impeached, however, more methodologically-sound instruments and assessments were soon developed for future state open court pilot projects.[89] The best example is Connecticut's Juvenile Access Pilot Program.[90] Whether to open the dependency courts in Connecticut was an extremely contentious political issue, and two earlier bills to open the courts were defeated prior to the 2009 legislation.[91] The political debate was even more problematic since those on both sides of the issue worked together on a daily basis in the dependency courts: (1) the attorneys representing the Department of Child and Family Services; (2) parents' lawyers; (3) children's lawyers, guardians *ad litem*, and child and adolescent psychologists and psychiatrists; and (4) judges. Thus, any open pilot project in Connecticut would not only have to cure the methodological flaws inherent in the Minnesota study, but also assure the disparate contingencies that they would have an equal role in the investigation and analysis of the study results to create sufficient credibility to support expanding the open court system beyond the pilot project. The legislative response was to form a Juvenile Access Pilot Program Advisory Board, which not only included representation of the broad scope of professionals working in and with the courts in dependency cases, but it also provided a seat for both proponents and opponents of opening the courts.[92]

The Connecticut pilot study mandated both quantitative and qualitative examination of the open court pilot project. For instance, the quanti-

---

89 One of the problems with some court pilot projects is that they approve a courtroom experiment *before* agreeing to the methodological techniques and outcome measures. For instance, after years of judicial debate, in September 2010 "the Judicial Conference authorized a three-year pilot project to evaluate the effect of cameras in district court courtrooms, video recordings of proceedings therein, and publication of such video recordings by making them available" through court websites. Press Release, United States Courts, Judicial Conference Comm. on Court Admin. and Case Management, Guidelines for the Cameras Pilot Project in the District Courts 1, *available at* http://www.uscourts.gov/uscourts/News/2011/docs/Cameras Guidelines.pdf. However, the Conference directed another committee to "conduct a study of the pilot" and to "prepare interim reports after the first and second years of the pilot." *Id.* Starting pilot programs in several different locations without first establishing a unified methodology and consistent outcome measures decreases the chances that valid conclusions can be generated out of the cumulative data.

90 H.B. 6419, 2009 Reg. Sess. (Conn. 2009).

91 *See* H.B. 5555, Reg. Sess. (Conn. 2004) ("An Act Concerning Public Access to Proceedings in Certain Juvenile Matters"); H.B. 5219, Reg. Sess. (Conn. 2005) ("An Act Concerning the Hearing of Juvenile Matters").

92 The advisory board included the Chief Court Administrator (a proponent), a center for child advocacy proponent, as well as attorneys from Connecticut Legal Services, the Division of Criminal Justice, the Bureau of Child Welfare Services, the Office of the Child Advocate, Office of the Chief Child Protection Agency, Office of the Chief Public Defender, Office of the Attorney General, a social worker representative for the Department of Child & Family Services, and a representative from the media. *Juvenile Access Pilot Program Advisory Board*, STATE OF CONNECTICUT JUDICIAL BRANCH, http://www.jud.ct.gov/Committees/juv_access/default.htm (last visited Nov. 14, 2014).

tative data included the number of hearings closed to protect the best interests of children and the number and success of motions by attorneys to close the presumptively open hearings.[93] Although the Connecticut qualitative interviews included comments from the same system professionals included in the Minnesota report, the Connecticut pilot study also surveyed abused children's parents and mental health experts regarding their attitudes toward opening the courts to the press and public.[94] Although the Advisory Board was co-chaired by two professionals who favored opening the courts at the beginning of the pilot study, when the Board reviewed all of the Pilot Project's empirical data, it recommended presumptively closing all of the juvenile dependency courts to the press and public because "significant concerns remained . . . that opening child protection proceedings could potentially harm children."[95] Unlike the Minnesota Pilot Project report, there have so far been no published attacks on the methodological and assessment validity of the Connecticut study.

Unlike the statutorily mandated outcome measures required under federal law, most state pilot projects share one or more of the following characteristics: (1) no mandatory general statutory data-based outcome measures; (2) an ad hoc or sui generis research design for each specific pilot project; and (3) very few objective outcome measures defined before implementation, so that determining success, failure, or the desirability of expanding the pilot project state-wide is usually more politically than empirically based.[96]

---

93 *See* JUVENILE ACCESS PILOT PROGRAM ADVISORY BD., REPORT TO THE CONNECTICUT GENERAL ASSEMBLY 20–21, 25 (2010).

94 *Id.* at 22.

95 *Id.* at 15, 30.

96 One might think that comparing accountability measures in federal transportation pilot projects with state accountability controls in child abuse pilot projects is inappropriate since testing transportation effects are much easier than determining the effects of changes in child abuse prevention and treatment results. However, federal outcome measure and compliance standards are equally rigorous for transportation projects and for federally supported child abuse prevention and treatment programs. *See* Social Security Act, Title IV-E § 422, 42 U.S.C. § 622 (2012); 45 C.F.R. § 205.50(a)(1)(i), (a)(2)(ii) (2014); 42 U.S.C.A. § 5106a(b)(2)(B)(viii) (2012); U.S. DEP'T OF HEALTH AND HUMAN SERVS., FEDERAL CHILD AND FAMILY SERVICES REVIEWS: AGGREGATE REPORT, ROUND 2, FISCAL YEARS 2007–2010, at 1 (2011), *available at* http://www.acf.hhs.gov/sites/default/files/cb/fcfsr_report.pdf. For an extensive analysis of federal accountability standards for federally supported state child abuse programs funded under the CAPTA and the Social Security Act, see William Wesley Patton, *Bringing Facts Into Fiction: The First "Data-Based" Accountability Analysis of the Differences Between Presumptively Open and Closed Child Dependency Systems*, 44 U. MEM. L. REV. 831 (2014).

CONCLUSION: LESSONS LEARNED

It is important to properly define "success" and "failure" for pilot projects. Sufficiently defining success enables evaluators to both "[u]nderstand why actual organizational conduct differs from official policy" and to better recommend "implementation strategies for change," and provides a more accurate determination for whether the pilot project should be expanded or made a permanent program.[97] Obviously, a methodologically sound, unbiased[98] report that produces a roadmap for successfully replicating the pilot project on a larger scale is a success under any definition.[99] But, we must not lose sight of the nature of pilot

---

97 Mary E. Poulin et al., *The Significance of Definitions of Success in Program Evaluation*, 24 EVALUATION REV. 516, 516–18 (2000). Heterogeneity among multiple pilot projects can substantially decrease the chances of successfully expanding to full scale programs. "[W]ell-matched comparison groups offer the best chance for identifying the effects of any particular intervention." CONG. BUDGET OFFICE, ISSUE BRIEF: LESSONS FROM MEDICARE'S DEMONSTRATION PROJECTS ON DISEASE MANAGEMENT, CARE COORDINATION, AND VALUE-BASED PAYMENT 7–8 (2012).

98 Project evaluations must avoid the potential for bias through a too limited or selective evaluation population. "It is essential to gain a comprehensive understanding of the motivations, constraints, and considerations of all parties involved in the program (policy makers, program operators, and the target population)." Yonatan Eyal, *Examination of the Empirical Research Environment of Program Evaluation: Methodology and Application*, 34 EVALUATION REV. 455,e476 (2010). The term "trapped administrators" describes those involved in the analysis of pilot projects who "have so committed themselves in advance to the efficacy of the reform that they cannot afford honest evaluation." Judith Clark Turner & William J. TenHoor, *The Nimh Community Support Program: Pilot Approach to a Needed Social Reform*, 4 SCHIZOPHRENIA BULL. 319, 341 (1978). Turner and TenHoor identified the following common weaknesses in project reports: (1) inadequate definition of service system goals; (2) fragmentation and confusion of responsibility among participants and agencies; and (3) lack of a systemic approach to financing community-based services. *Id.* at 322–25.

99 Drawing generalizations about pilot projects is difficult because there is a bias in the publication of pilot studies. "Publication bias may occur because of a tendency for journals to accept only papers that have statistically significant results and not to report non-significant effects . . . ." van Teijlingen & Hundley, *supra* note 2. Pilot studies also involve Type III errors: the failure to adequately and frequently assess target populations implementation of pilot goals. "Two frequently overlooked issues in implementation involve documenting the rate and intensity of participation of the target population, and monitoring alternative programs delivered to control or comparison groups." Joseph A. Durlak, *Why Program Implementation Is Important*, 17 J. PREVENTION & INTERVENTION COMMUNITY 5, 13 (1998). Differing levels of participation may involve logistical problems, differing levels of participation stress, and differences among participants' abilities to access and use program materials which can be analyzed using "[f]idelity reports [that] show that all leaders were able to cover the material and differences in group composition." Tali Raviv & Martha E. Wadsworth, *The Efficacy of a Pilot Prevention Program for Children and Caregivers Coping with Economic Strain*, 34 COGNITIVE THERAPY RES. 216, 225 (2010). The failure to adequately account for participation rates and "drop out" participants can seriously bias pilot project results.

> For example, if persons with better future prospects were more likely not to bother showing up for the program, participants would overrepresent persons with less favorable prospects. Thus the control group outcome would overestimate the outcome for participants in the absence of the program, and the difference between the observed participant and control group outcomes would underestimate the effect of

projects—they are experiments, and even if they fail they can successfully provide data on how to redesign projects, or evidence of why a larger scale project should not be pursued.[100] Even if the pilot project appears to be initially successful, however, if the methodology and assessment tools present an inaccurate and/or biased report, expansion beyond the pilot project may ultimately prove to be a costly mistake:

> [W]here a pilot study requires a significant investment of resources, making it difficult for the study team to call a halt to the research after an unsuccessful pilot study . . . [they] might be tempted to make considerable changes in the main study, rather than deciding that the proposed study is not possible with the available resources, time, population, etc.[101]

---

the program. The reverse would occur if persons with less favorable prospects (perhaps because of weak motivation) were most likely not to show up for the program. Howard S. Bloom, *Accounting for No-Shows in Experimental Evaluation Designs*, 8 EVALUATION REV. 225, 226 (1984). Assessment design must account for the effects of subject selection. For instance, in studies involving children, a shift from passive parental consent (students become study subjects unless parents object) to active parental consent (students cannot participate without active, and often written, parental consent) can result in: (1) a biased sample substantially underrepresenting minority children; (2) reduced response rates; and (3) a lower incidence of antisocial behavior among the test group. Mathew W. Courser et al., *The Impact of Active Consent Procedures on Nonresponse and Nonresponse Error in Youth Survey Data: Evidence from a New Experiment*, 33 EVALUATION REV. 370, 371–75 (2009). For an interesting solution to the problem of designing a study when "random allocation is not possible" between control and reference groups, see Jose E. Urquieta-Salomón et al., *Poverty and Gender Perspective in Productive Projects for Rural Women in Mexico: Impact Evaluation of a Pilot Project*, 33 EVALUATION REV. 27, 31 (2009).

100 The most vital pilot project reviews include evaluative and summative, or "sunset," review. To be successful, sunset review, which requires justification of the continuation of a program, "should be accompanied by realistic information requirements that are tailored to specific program types and by early and detailed guidance concerning the designs and methods that will produce the desired information." Linda Berry, *Providing Guidance for Program Evaluations: Sunset Reviews Versus Evaluation Plans*, 10 EVALUATION REV. 757, 772 (1986). The quality of evaluative reviews is increased by "reviewing evaluations of similar programs, by consulting with program managers, and by defining the tasks required to achieve the desired results." *Id.* at 758.

101 van Teijlingen & Hundley, *supra* note 2. Program evaluators are not immune from temptations toward unethical conduct. In one study, "over two thirds (68%) of those who had done external evaluations exclusively said that they had faced ethical conflicts, whereas less than half (49%) of those who had conducted only internal evaluations reported any ethical challenges." Michael Morris & Robin Cohn, *Program Evaluators and Ethical Challenges: A National Survey*, 17 EVALUATION REV. 621, 636 (1993). Over 59% of evaluators listed pressure applied on evaluators regarding the presentation of the project findings as the most common ethical conflict they face. *Id.* The ethical conflicts strike at the heart of the program examiners' role to "seek the truth and communicate it . . . evaluators frequently feel pressured to compromise their role as scientists." *Id.* at 639. The process of defining goals and assessment instruments during the formative phase is a deeply political process in which different constituencies attempt to realize their often conflicting goals. *See* Judy L. Fitzpatrick, *Roles of the Evaluator in Innovative Programs: A Formative Evaluation*, 12 EVALUATION REV. 449, 449, 452, 458 (1988).

The definition of pilot project "success" must, therefore, include the quality and reliability of the data generated by the pilot project, not just the conclusion that the experiment produced positive results.[102]

In order to produce valid comparative and longitudinal assessments, pilot projects need to provide: (1) consistency of terminology and assessment instruments;[103] (2) baseline definitions;[104] (3) realistic budgetary

---

102 Drafting unambiguous pilot project performance measures is essential in reducing contract litigation. *See, e.g.*, OTI Am. v. United States, 68 Fed. Cl. 108, 112 (2005) (discussing a conflict surrounding the Government Printing Office [GPO] pilot program that created a four-stage head-to-head competition among companies to design a new passport. The GPO alleged that OTI America failed to "meet the contract requirements" and OTI sued for breach of contract under the Tucker Act, 28 U.S.C. §1491(b) (2012), arguing that the conditions that led to an inferior product had been out of its control and were corrected). *See also* Wilco Floor Serv., Inc. v. United States, 197 Ct. Cl. 902, 903 (1972) (holding that failure to pay costs for the first six weeks on the initial pilot project phase constituted a breach of contract); Phoenix C & D Recycling v. Des Moines Metro. Area Solid Waste Agency, No. 0-235/09-0712, 2010 Iowa App. LEXIS 508, at *3 (Iowa Ct. App. May 26, 2010) (addressing a dispute over the costs of waste cleanup in a pilot project based upon agency determination that contractor "did not meet minimum standards set by MWA"); Sargent v. Block, 576 F. Supp. 882, 882 (D.D.C. 1983) (addressing a dispute over the legality of the pilot program verification procedures). For cases involving pilot project statutory and constitutional violations, see *In re* D.W., 155 P.3d 682, 682–83 (Haw. Ct. App. 2007) (holding that the Family Court pilot program providing parents in child abuse cases with guardians *ad litem* rather than attorneys did not violate due process); Smith v. L.A. Cnty. Bd. of Supervisors, 128 Cal. Rptr. 2d 700, 700 (Cal. Ct. App. 2002) (holding that the county's Home Call Visitation Pilot Project which required home visits as a condition for receiving welfare benefits did not violate the Fourth Amendment); Int'l Bhd. of Teamsters v. U.S. Dep't of Transp., 724 F.3d 206, 206 (D.C. Cir. 2013) (holding that the Federal Motor Carrier Safety Administration pilot project to permit Mexico-domiciled trucking companies to operate in the United States did not violate federal law).

103 *See, e.g.*, U.S. DEP'T OF EDUC., PARTNERSHIPS IN CHARACTER EDUCATION: STATE PILOT PROJECTS, 1995–2001 LESSONS LEARNED 7 (2008), *available at* http://www2.ed.gov/programs/charactered/lessons.pdf (arguing that the federally-funded state character education pilot projects completed between 1991 and 2006 could not be effectively compared because "[t]he assessment process was different from state to state because no uniform assessment model was available"). Although standardized assessment tools provide for reliable comparisons among pilots, standardization may lead to less effective pilot programs in which "cookie-cutter" approaches may not meet local or regional needs. *See id*. at 15 (arguing that some states, such as Vermont and California, found a need to design different assessment methods because "each program was unique to its specific school environment," but, due to the lack of standardization programs, "could not be assessed adequately across these different environments, which made it difficult to draw quantitative conclusions"). In addition, researchers cannot just intuitively generalize program models among different subject populations. *See, e.g.*, Jodi Lane & Lonn Lanza-Kaduce, *Before You Open the Doors: Ten Lessons from Florida's Faith and Community-Based Delinquency Treatment Initiative*, 31 EVALUATION REV. 121, 145 (2007) (finding that the Office of Juvenile Justice and Delinquency Prevention (OJJDP) Florida pilot project to design and implement a faith and community based program for incarcerated juveniles between the ages of thirteen to seventeen simply did not work for children six to eleven years old because it did not account for the difference among family members' participation rates and scheduling difficulties when younger children were studied).

104 U.S. DEP'T OF EDUC., *supra* note 103, at 8 (finding that "[i]n some cases, schools lacked baseline data to make comparisons and, thus, measure improvement."). *See also* Johanna D. Birckmayer & Carol Hirschon Weiss, *Theory-Based Evaluation in Practice: What Do We Learn?*, 24 EVALUATION REV. 407, 408 (finding that theory based evaluation requires a

costs;[105] (4) defined outcome measures and explicit definitions and measures of success;[106] and (5) the focus and freedom to critically assess whether the pilot program needs modification, whether the pilot experiment should be expanded into a permanent program, or whether the pilot program should end without expansion.[107] These accountability controls

---

determination of "what activities are being conducted, what effect each particular activity will have, what the program does next, what the expected response is, what happens next" until the expected outcomes are analyzed); *id.* at 408 (stating that "[a]n evaluation that investigates the theories underlying . . . [a] program . . . seeks to find out whether the theories on which the program is based are realized in action"); *id.* at 429 (finding that the quality of project evaluations will be increased if the project theory is well-defined, the project reflects "the assumptions embedded in the theory," and there is sufficient time and resources to properly evaluate the pilot study); *id.* at 426 (discussing whether the theoretical assumptions underlying pilot projects need to be made explicit since implicit assumptions will require the evaluator to construct assessment variables).

105 Since one of the salient issues inherent in pilot projects is to determine feasibility of extending the limited pilot in size and duration, the pilot must determine realistic costs for providing the promised services and/or pilot goals. *See, e.g.*, U.S. DEP'T OF EDUC., *supra* note 103, at 8 (determining that in the federal character education, state pilots many could not be sustained without continuing federal budgetary support, and "[a]lthough the Pilot Project grants funded activities for up to five years, sustaining any significant education program thereafter required new or reallocated funding"). In order to help sustain the state programs funding had to be increased by almost 300%. *Id.* at 10. Some of the educational pilot programs lacked money to sustain and expand the pilots. *Id.* at 19. Other pilot project studies do not even mention the cost of the pilot or project the costs of expansion. *See, e.g.*, JUDY CASHMORE & LILY TRIMBOLI, NSW BUREAU OF CRIME STATISTICS AND RES., AN EVALUATION OF THE NSW CHILD SEXUAL ASSAULT SPECIALIST JURISDICTION PILOT (2005), *available at* http://www.bocsar.nsw.gov.au/agdbasev7wr/bocsar/documents/pdf/r57.pdf (describing a ninty-two-page analysis of a child sexual assault pilot program without any discussion of the cost of the pilot or what the cost of expanding the pilot would be on court proceedings).

106 *See* Maggie Worth, *How to Write a Proposal for a Pilot*, EHOW, http://www.ehow.com/how_6611523_write-proposal-pilot-project.html (last visited Nov. 16, 2014) (stating that a pilot project proposal should "[e]xplain how the project will be evaluated . . . [and] what results constitute a successful program"). *See generally* Eric Brown et al., *Design and Analysis of the Community Youth Development Study Longitudinal Cohort Sample*, 33 EVALUATION REV. 311, 312 (2009) (providing an excellent analysis of a large-scale longitudinal pilot study analysis of an evidence-based program to reduce adolescent health and behavior problems). *See also* Peter Z. Schochet, *An Approach for Addressing Multiple Testing Problem in Social Policy Impact Evaluations*, 33 EVALUATION REV. 539, 539 (2009) (providing a discussion of multiple testing problems in statistical analyses conducted across multiple outcomes and subgroups); Richard Berk, *Evidence-Based Versus Junk-Based Evaluation Research: Some Lessons from 35 Years of the Evaluation Review*, 35 EVALUATION REV. 191, 196 (2011) (providing an excellent discussion of the inherent problems in casual modeling in survey instruments); Wade Horn & Joel Heerboth, *Single-Case Experimental Designs and Program Evaluation*, 6 EVALUATION REV. 403,e405 (1982) (stating that proving causation in single case experiments is more direct since it involves "taking repeated measures of one or more dependent variables and systematically applying, and in some designs withdrawing, an independent variable. If the application or withdrawal of the independent variable is associated with systemic changes in the dependent variable, then it is inferred that the independent variable has caused the changes.").

107 *See* Fitzpatrick, *supra* note 101, at 459 (arguing that evaluators and researchers of pilot projects "must protect themselves against . . . personal identification with the design or delivery format of the program when measuring outcomes"). *See also id.* at 459–60 (stating that assessment during the experimental pilot program should ask questions involving (1) ma-

will better ensure that limited governmental resources will be intelligently spent on pilot experiments that have the best odds of success and expansion to full-scale governmental programs.

jor revisions, (2) modification and elimination of unattainable goals, (3) certain "radically alter[ed] or expand[ed] . . . delivery components to achieve problematic objectives," and (4) expanding successful components). In deciding whether to move to a full-scale project "[o]ne approach is to estimate all possible benefits and costs in terms of monetary amounts and weigh benefit against cost." Pal et al., *supra* note 14, at 272. Alternatively, relative importance of various decision criteria may be established and the ratings under each of the criteria can be assimilated into overall normalized scores. *See also id.* (arguing that "[i]rrespective of the evaluation scheme the essence of the pilot study remains important to take the next big step, i.e., implementation of the full scale project").