Administration and Policy in Mental Health and Mental Health Services Research
https://doi.org/10.1007/s10488-020-01043-0

ORIGINAL PAPER

An Innovator and a Disruptor: Leonard Bickman on Program Theory, Null Findings, and Advice to Future Child Mental Health Services Researchers

Susan Douglas[1]

Abstract
With a career spanning more than half a century, Leonard Bickman has contributed to improving children's mental health through research on innovative interventions, such as measurement-based care, and service-level initiatives, including systems of care. Len's highly productive career in children's mental health services research is founded in his influence on the science of program evaluation, particularly in the area of program theory. This article provides an abridged and combined version of a video interview with Len dated April 16, 2019 and written responses to a series of questions posed to Len in advance of the festschrift held in his honor at Johns Hopkins Bloomberg School of Public Health on May 11, 2018.

Keywords Leonard Bickman · Program evaluation · Program theory

Over a decade ago, Leonard Bickman (2008; Douglas Kelley and Bickman 2009) defined a measurement feedback system as a technology for administering measures of treatment progress and process systematically and frequently throughout treatment, with results interpreted and delivered to clinicians as timely and clinically useful feedback. Under Len's leadership at the Center for Evaluation and Program Improvement (CEPI) at Vanderbilt University, a team of researchers developed one of the early measurement feedback systems called Contextualized Feedback Systems (CFS; Bickman et al. 2011, 2016). CFS was based on a comprehensive theory of how feedback influences clinician practice (Riemer and Bickman 2011; Riemer et al. 2005). Because a measure set using brief and clinically relevant scales did not yet exist to assess treatment progress and process in child mental health services, the research team collaborated with clinicians to create the Peabody Treatment Progress Battery (PTPB; Bickman et al. 2007, 2010; Riemer et al. 2012).

More recently, the field has moved to the terminology of measurement-based care to describe the systematic assessment of patient symptoms at each clinical encounter to inform care (Scott and Lewis 2015). Len saw the potential of measurement-based care as a promising intervention that was not specific to any particular treatment modality (e.g., cognitive behavioral therapy), but rather operated under some basic principles of human behavior and the influence of feedback on decision-making (Riemer and Bickman 2011; Riemer et al. 2005). As researchers in the field of children's mental health services, the CEPI team was interested in the non-specific nature of the intervention given the groundswell of findings at the time that evidence-based treatments were not affecting outcomes in community settings to the degree they had done so in the original empirical studies (Douglas Kelley et al. 2010; Karver et al. 2006). After first establishing that measurement-based care positively impacted treatment effectiveness (Bickman et al. 2011), the team also contributed to the growing interest in measurement-based care with work on implementation (Bickman et al. 2016; Douglas et al. 2015, 2016; Gleacher et al. 2016; Lewis et al. 2019) and mechanisms of action (Douglas et al. 2015). In 2010, Len and the CEPI team began an interesting project that blended measurement-based care with an evidence-based treatment, with the intent of exploring the additive effect of feedback on both treatment fidelity and outcomes. Unfortunately, Len had to close down the initiative when CEPI discontinued using the CFS technology due to continuing technological challenges and related expenses.

✉ Susan Douglas
susan.douglas@vanderbilt.edu

[1] Department of Leadership, Policy, and Organizations, Peabody College, Vanderbilt University, 230 Appleton Place, PMB #414, Nashville, TN 37203-5721, USA

Published online: 22 April 2020

As Len (Bickman et al. 2016) later wrote, measurement feedback systems meet the definition of a 'disruptive innovation' (Rotheram-Borus et al. 2012) as both a technological intervention and a quality improvement tool. Recent meta-analyses and systematic reviews have affirmed the role of measurement-based care as an evidence-based practice in mental health (see summary in de Jong et al. 2019). Measurement-based care is now included as part of usual care in several countries and is increasingly becoming a standard for practice and accreditation in the United States (Joint Commission 2011). There is a growing literature on the implementation of and mechanisms of action of measurement-based care to better inform the development and support of this tool in practice.

With a career spanning more than half a century, and no intention of stopping any time soon, Len's work on measurement-based care is only one example of his influence on the field of children's mental health services. Len is very well known for his seminal contributions in the area of program theory (Bickman 1987, 1989; Bickman and Peterson 1990) that significantly influenced the formation of the science of program evaluation. In addition to his innovative ideas that bridge disciplines, Len's methodological rigor is his hallmark, which stood him in good stead back in the 1990s during his years of "null findings" on the Fort Bragg Evaluation Project and the Stark County Evaluation Project (Bickman 1996a, b, 1997, 1999; Bickman et al. 1999). Len's research set the standard for incorporating implementation monitoring into any project to facilitate better discrimination between program theory failure and implementation failure.

Len was interviewed on April 16, 2019 and also provided written responses to a series of questions in advance of the festschrift held in his honor at Johns Hopkins Bloomberg School of Public Health on May 11, 2018. Below is an abridged and combined version of both interviews.

Program Theory Gives Us the Freedom to Answer, "How Does a Program Work"?

Susan: In the late 1980s and early 1990s, you wrote seminal articles about the use of program theory in evaluation, which contributed to the approach of theory-based evaluation (Bickman 1987, 1989; Bickman and Peterson 1990). What drew you to be interested in bringing a theory-based understanding to the field of program evaluation?

Len: Well I guess part of it is my background. People doing evaluation now come with an educational background in evaluation, right? For people of my era, there were no evaluation courses, there were no mentors. We were essentially the first generation of evaluators and we lacked credentials in evaluation that would have typically come in the form of coursework and projects and so on. So, we came to it from our own PhD discipline training background—to some extent, it was about not being trained in evaluation.

My background was in social psychology, which was a factor that influenced how I looked at program evaluation. Now, my perspective was quite different than most social psychologists. Two of my mentors—and I give them credit—one was a gentleman named Harold Proshansky who co-developed the field of environmental psychology (Rivlin 1992), which is still a major field. My other advisor was a psychologist named Stanley Milgram, who was famous for his obedience studies but actually was much broader than that in his thinking about how psychology can contribute to society (Blass 2005). So those two scholars had a strong influence on me, and one of their influences was that they, and subsequently me, weren't happy with being a social psychologist in the traditional mode which, back in the 1970s and 80s, was essentially a laboratory researcher who did research on a particular area and focused on either questionnaires or simulations that occur in laboratory. That just wasn't me, so even as a graduate student I published research on field experiments in the real world (Bickman and Henchy 1972). At least, that's what I thought was the real world. In the first two years of my career, I focused on an area that was a particular interest of mine, which was bystander intervention in emergency situations (Bickman 1984). My dissertation was actually a laboratory study, one of my last (Bickman 1971). And then we did a number of field experiments, funded by the National Science Foundation, looking at people intervening in mild crime situations (Bickman 1984). It took me a while to realize that those situations were in some sense as artificial and as removed from reality as the laboratory studies.

The field of program evaluation was just starting then. Neither of my advisors were involved in that area. But as a young assistant professor I did get involved. As I look back, another influential person was Marcia Guttentag, who was also a social psychologist by training, and one of the leaders in the field. She passed away too early in her career, but she produced a handbook on program evaluation (Guttentag and Struening 1975). I attended the first national meetings on program evaluation and there were several societies back then. I got involved in program evaluation because I wanted a greater sense of reality than even social psychological field experiments could bring, because they didn't deal with real problems with real people.

Now to get back to the question of where did program theory come in. There was one aspect of program evaluation that I found really boring and that was how people conceptualized evaluations. Back then there wasn't one society of evaluators. There was something called the Evaluation Research Society[1] and the Evaluation Network. The Evaluation Network was composed primarily of educational evaluators, and their approach was to evaluate based on the objectives of the program. As educators, their background led them to develop a system that focused on objectives, and sub-objectives, and sub-sub-objectives. Each one of those had to be evaluated and it was really very repetitive, very narrow and very uninteresting. Huey Chen and I independently came up with the idea of looking at program theory—and it is theory, as I point out in my writing, with a small "t" (Bickman 1987, 1996b; Bickman and Peterson 1990).

I come from a tradition in psychology of big "T's". Going back to the 1930s, before my time, it was what I called the 'great white men theories of everything.' Essentially, these are theories that are named after people, typically white males, and that's how people identify themselves (e.g., "I'm a Rogerian," or "I'm a Skinnerian"). And those weren't the theories that I wanted to address, because they had little relevance to what programs were doing. The issue really became, and I think this is a particularly academic perspective more than a practitioner perspective: "Why are we doing this? Why do we do program evaluations?"

I still think people sometimes miss the big picture here. Most people evaluate individual programs. Is this program meeting its goals and objectives? Was this program implementing the plan correctly? Was this program meeting the needs of the people it is designed to serve? They were all individual instances of something I thought. And what was it? Well, it's that there should be a theory underlying every program. Why does it work?

For example, why do you think that changing the color of these walls will increase educational attainment? If you look at many of the major interventions that our government has invested billions of dollars in, you find that people were not conceptualizing their interventions at a very high conceptual level. People don't get degrees or training in program development, as far as I can tell. They do get training in program evaluation, but how do you develop a program? What are the key ingredients? What are you trying to accomplish? What are the things that you think will make this work? What has to happen—and why? It is the why part that is the theory part. Why should this program work? And if you don't think that through, you probably are not going to have a successful program. If you're not basing a program on the extant research literature, then that program is based on what I call 'notions' and not a strong conceptual framework.

I remember a program that I was involved in evaluating, which was later turned into a multibillion-dollar program at the federal level. First, the program was supported by a research foundation that was attempting to reduce drug use. When I interviewed the foundation leadership, I learned their efforts were based on a notion that was very different from the major thrust to reduce the supply of drugs. Their notion was to reduce the demand. Sensible, yes. Supply and demand, if you cut both of them or either one of them, you should reduce drug use. But that was as far as that thinking went. It was novel back then, but how they then went from that notion to what they actually supported was to me a real gap. So, what did they support? They said, well—and I really saw this in terms of blaming the victim as opposed to coming up with a creative solution—communities are complaining that we don't give them enough support and resources to deal with drug use, so our intervention is essentially to give money to communities so locals can come up with their own solutions to reduce demand.

Susan: How did that work out?

Len: Terrible, as you'd expect it would. Why would you expect anyone to be able to come up with successful solutions based on no training, no background, and little understanding of the underlying issues? This is where I look at it as 'blame the victim.' "All right you guys, the complaining that the government doesn't provide solutions has to stop. Here's $300,000, develop a solution." And that was basically it! What we found was that the community groups didn't develop any solutions. Instead, they spent most of their time organizing and reorganizing and worrying about representation. They had lots of meetings and the people were naïve, trying things like having a day to celebrate being drug free to generate publicity. But they didn't come up with a theory that said how these interventions should reduce demand. Instead, what they actually came up with was a bunch of activities. They never developed a theory of action. You have to always look at the full range of how a program is supposed to operate. How is this part going to influence that part?

An unfortunate thing I'm seeing now is more and more evaluations that are backing away from really looking at effectiveness. One, because the money may not be there to support it and, two, it's very uncomfortable for people to evaluate effectiveness. Instead, these evaluations are focusing on processes, like how well networks are integrated or whether people are communicating with each other. But, how does communicating better lead to better outcomes? At the bottom line is the client or the person who's suffering and evaluators always have to keep their eye on that bottom line. While evaluations can be done in phases, that raises

[1] The Eastern Evaluation Research Society (https://eers.org/about-eers/) is not the same organization.

the concern that all the latter phases won't happen. In her dissertation, one of my students, Debra Rog, studied evaluability assessments, where you explore whether a program is ready to be evaluated (Leviton et al. 2010). Is there substantial evidence that the program is based on sound concepts? Is it feasible to evaluate the program? These are just a few of the questions that should be asked before committing to an evaluation. She found that those programs that first did evaluability assessments usually never ended up doing evaluations because they thought they were done, regardless of the results of the evaluability assessment.

One of the big surprises for me was that when you write something, you get feedback from people. I mean, no one writes you fan letters, but on one of my first sabbaticals in Australia, I was approached by a group of evaluators that thanked me for the article or chapter I wrote on program theory. I asked why this was this interesting to them, and was told that before, all that people would do in evaluations was take the educational approach of looking at objectives one after the other. They went on to say that it freed them to look at the broad picture of things and understand the basis of the programs they were evaluating. It gave them a reference they could identify and defend to their sponsors.

Evaluation as a Disruptive Force

Susan: I recall that something like 80 articles, books, book chapters were produced about the Fort Bragg Evaluation and Stark County Evaluation Projects. There were special issues of the American Psychologist and symposia where leaders in the field of child mental health services talked about the null findings. Yet, the system of care continued to be funded. What is important for program evaluators to consider about the political or contextual factors that influence program evaluation and its value?

Len: Two things. One, what's the alternative? People are reluctant to move away from something that works in terms of funding mechanisms unless they have an alternative. No one had one. I went off, as you know, into a different area of research for the last 20 years, looking at feedback as an alternative way of dealing with the lack of effective treatments in child mental health services. We're still doing research on measurement-based care and it seems to be increasingly adopted. But systems of care, similar to Fr. Bragg, are still being funded by the Substance Abuse and Mental Health Services Administration (SAMHSA), with over a billion dollars spent since its inception.

People need solutions. You just can't say to people, 'Oh, that doesn't work.' What were we supposed to do, stop funding kid's services? That's one of the reasons I never testified before Congress on this. Because what would happen? The best that would happen is that the government would take the money away from kids and put it into something else. I never got involved in political aspects of trying to stop the funding of systems of care for that reason, because the money was going towards children's mental health. Until we have something that we can offer as a substitute, it is not going to change much. You need an alternative and it is not necessarily the responsibility of evaluators to provide alternatives. Although I personally felt I had to do that as a researcher, but not as an activist. The paradox now is that I was just appointed to SAMHSA's National Advisory Committee for the Center for Mental Health Services, which funds systems of care. I am hoping that my presence there will result in the consideration of alternative service approaches.

Susan: You consider the effect on the stakeholders if the program goes away.

Len: Exactly. I do say to people, "Look if you do this effectiveness evaluation, are you prepared for failure?" It is my responsibility to make sure the stakeholders go in with their eyes wide open about the risks they're taking. And if they can't appreciate those risks, they shouldn't do an outcome evaluation, but instead do a developmental evaluation. You could say that you don't think the program's ready to be tested this way, or that the field is not ready.

Most program people believe in their treatment. They have to. So, you're a disruptive force as an evaluator. You are disrupting their thinking and questioning some of the basic assumptions underlying a program. Pointing out gaps in a logic model is not usually a comfortable situation for people because program people by and large are advocates, and advocates have difficulty acknowledging and questioning the unspoken assumptions that they're making. So, you have to have some awareness of the larger context, you have to have some understanding of the problem, and recognize there are people involved.

Susan: It's like a partnership and keeping the people in mind is so important.

Len: You can't be too doubtful about the efficacy of your own activities and still be motivated. There's a balance there and the evaluator comes along and is disrupting to that balance. So, the question is—can you afford to really understand your intervention? Can you really afford to know whether or not you're having the positive outcomes that are motivating you to do whatever this is? Because the problems most programs deal with are difficult and complex, it is likely that many interventions will not be effective. Thus, in many cases evaluations will find evidence that is at best equivocal, occasionally effective, but a lot of times there will be no evidence that a program works. If you see the evaluative process as a building block to improvement, where the data are used to help guide changes, that holds out hope for the future.

Len's Advice for Future Generations: Follow Your Interests and Ask Questions

Susan: What is your advice for the next generation of child mental health services researchers?[2]

Len: **Be a skeptical optimist.** My skepticism has been developed over many years of mentoring, experimentation, and experience, and should not be mistaken for cynicism. The former is a tendency not to accept things as they appear to be, while the latter connotes contempt and suspiciousness of the motives of others. Skepticism is not something that is valued by the orthodox. I know as a youth, my incessant questioning about religion drove my religiously observant father to distraction. I think all researchers should be skeptical, especially of their own work. The optimistic part is what keeps me going after so many failures to find effective solutions to the research problems that I study. Without the sense that solutions will be found, it would have been easy just to give up.

Enjoy Positive Marginality

Another aspect of being in an interdisciplinary field such as program evaluation and services research is that you never really are in the mainstream of any field. The late Clara Mayo (1982) wrote a wonderful essay on the advantages of being marginal and not in the mainstream. One advantage is the ability to see assumptions that people who are immersed in that area have difficulty seeing because it has always been there, hence why fish will be the last to discover water. For example, I have done research that focused on some of the core concepts of therapy. We studied supervision and therapeutic alliance, both from an experimental perspective. What I was surprised to discover is that among the thousands of articles and books on those topics, there was almost no randomized experiments. Treatment planning has loads of books on how to do it but almost no research on whether it makes a difference in outcomes. As an outsider, these gaps were obvious to me but seemed hidden in plain view to the clinical field.

Recognition, Rewards, and What Really Matters

Looking back at the earlier stages of my career I found it was interesting, fun, challenging and moreover, someone was willing to pay me to do research. I enjoyed the process of doing social psychology. It was both fun and challenging to do field experiments. Publication in good journals and continued external funding was sufficient recognition. Later, I became more committed to doing research in the real world. I wanted to work on serious problems, but I was not committed to becoming an advocate for research-based amelioration of those problems. Publication of my own research was still highly rewarding. However, I believe that after the first 75 publications or so, they do not count as much towards one's own career goals but can serve as good opportunities for advancing the careers of graduate students and junior colleagues. This is very gratifying. In the midpoint of my career, I also found that disseminating other's research as an editor of a journal and several books were significant activities. Attempting to help shape the field through leadership roles was also rewarding. In the later stages of my career I continue to have these goals.

It has been both a challenge and a pleasure to respond to these questions. What were the major events and people that I thought shaped my career? How much of what happens during a lifetime is due to chance? If I had done more careful planning or made other choices how would my career have been different? How much was determined by my environment in contrast to my personality and values? I do not think being in the right place at the right time is the whole story, but it certainly played a role. More important is persistence and integrity in doing what you believe in and seeing your surroundings as filled with opportunities and challenges. I strive to be open-minded yet skeptical and believe that scientific rigor and compassion should co-exist.

Many of my colleagues my age have retired. Almost all my friends outside of academia have already retired. I think people retire for primarily three reasons: poor health, desire to travel or not loving their work. I love what I am doing, I travel as much as I want and, in the past, usually somebody else pays for it, and I am healthy enough to work but I realize, only too well, that good health is difficult to predict. I see myself similar to the old Maine farmer who, when asked by a tourist, "Lived here all your life?" answered, "Not yet."

Conflict of interest Susan Douglas reported receipt of compensation related to the Peabody Treatment Progress Battery and a financial relationship with Mirah; there is a management plan to ensure that this conflict does not jeopardize the objectivity of her research. No other disclosures were reported.

References

Bickman, L. (1971). The effect of another bystander's ability to help on bystander intervention in an emergency. *Journal of Experimental Social Psychology, 7*, 367–379. https://doi.org/10.1016/0022-1031(71)90035-7.

Bickman, L. (1984). Bystander intervention in crimes: Research and application. In J. Karylowski, J. Rekowsky, E. Staub, & D. Bar-Tal

[2] Len's written response to this question dated May 9, 2018 included material from previously published interviews (Fitzpatrick 2002; Fitzpatrick et al. 2009) and an autobiographical essay (Bickman 2006).

(Eds.), *Development and maintenance of prosocial behavior: International perspectives* (pp. 457–470). New York: Plenum.

Bickman, L. (1987). The functions of program theory. *New Directions for Program Evaluation, 1987*(33), 5–18. https://doi.org/10.1002/ev.1443.

Bickman, L. (1989). Barriers to the use of program theory. *Evaluation and Program Planning, 12*(4), 387–390. https://doi.org/10.1016/0149-7189(89)90056-6.

Bickman, L. (1996a). A continuum of care. More is not always better. *The American Psychologist, 51*(7), 689–701. https://doi.org/10.1037/0003-066x.51.7.689.

Bickman, L. (1996b). The application of program theory to the evaluation of a managed mental health care system. *Evaluation and Program Planning, 19*(2), 111–119. https://doi.org/10.1016/0149-7189(96)00002-x.

Bickman, L. (1997). Resolving issues raised by the Fort Bragg evaluation. New directions for mental health services research. *The American Psychologist, 52*(5), 562–565. https://doi.org/10.1037/0003-066x.52.5.562.

Bickman, L. (1999). Practice makes perfect and other myths about mental health services. *American Psychologist, 54*(11), 965–978. https://doi.org/10.1037/h0088206.

Bickman, L. (2006). My life as an applied social psychologist. *Current Psychology: Developmental, Learning, Personality, Social, 25*(2), 67–92. https://doi.org/10.1007/s12144-006-1005-5.

Bickman, L. (2008). A measurement feedback system (MFS) is necessary to improve mental health outcomes. *Journal of the American Association of Child and Adolescent Psychiatry, 47*, 1114–1119. https://doi.org/10.1097/chi.0b013e3181825af8.

Bickman, L., & Henchy, T. (Eds.). (1972). *Beyond the laboratory: Field research in social psychology*. New York: McGraw-Hill.

Bickman, L., & Peterson, K. A. (1990). Using program theory to describe and measure program quality. *New Directions for Program Evaluation, 1990*(47), 61–72. https://doi.org/10.1002/ev.1555.

Bickman, L., Athay, M. M., Riemer, M., Lambert, E. W., Douglas Kelley, S., Breda, C., et al., Tempesti, T., Dew-Reeves, S. E., Brannan, A. M., & Vides de Andrade, A. R. (Eds.). (2010). *Manual of the peabody treatment progress battery* (2nd ed.). Nashville, TN: Vanderbilt University.

Bickman, L., Douglas Kelley, S., Breda, C., de Andrade, A. R., & Riemer, M. (2011). Effects of routine feedback to clinicians on mental health outcomes of youths: Results of a randomized trial. *Psychiatric Services, 62*(12), 1423–1429. https://doi.org/10.1176/appi.ps.002052011.

Bickman, L., Douglas, S. R., De Andrade, A. R., Vides, T. M., Gleacher, A., … Hoagwood, K., (2016). Implementing a measurement feedback system: A tale of two sites. *Administration and Policy in Mental Health and Mental Health Services Research, 43*(3), 410–425. https://doi.org/10.1007/s10488-015-0647-8.

Bickman, L., Noser, K., & Summerfelt, W. T. (1999). Long-term effects of a system of care on children and adolescents. *The Journal of Behavioral Health Services & Research, 26*(2), 185–202. https://doi.org/10.1007/bf02287490.

Bickman, L., Riemer, M., Lambert, E. W., Douglas Kelley, S., Breda, C., Dew, S. E., et al., Brannan, A. M., & Vides de Andrade, A. R. (Eds.). (2007). *Manual of the peabody treatment progress battery*. Nashville, TN: Vanderbilt University.

Blass, T. (2005). The urban psychology of Stanley Milgram. *Journal of Urban Distress and the Homeless, 14*(1–2), 12–22. https://doi.org/10.1179/105307805807066293.

de Jong, K., Barkham, M., Wolpert, M., Douglas, S., & Delgadillo, J. et al. (2019). The impact of progress feedback on outcomes of psychological interventions: An overview of the literature and a roadmap for future research. *Manuscript submitted for peer review.*

Douglas Kelley, S., & Bickman, L. (2009). Beyond outcomes monitoring: Measurement feedback systems in child and adolescent clinical practice. *Current Opinion in Psychiatry, 22*, 363–368. https://doi.org/10.1097/yco.0b013e32832c9162.

Douglas Kelley, S., Bickman, L., & Norwood, E. (2010). Evidence-based treatments and common factors in youth psychotherapy. In B. L. Duncan, S. D. Miller, B. E. Wampold, & M. A. Hubble (Eds.), *The heart and soul of change: What works in therapy* (2nd ed., pp. 325–355). Washington: American Psychological Association. https://doi.org/10.1037/12075-011.

Douglas, S., Button, S., & Casey, S. E. (2016). Implementing for sustainability: Promoting use of a measurement feedback system for innovation and quality improvement. *Administration and Policy in Mental Health and Mental Health Services Research, 43*(3), 286–291. https://doi.org/10.1007/s10488-014-0607-8.

Douglas, S. R., Jonghyuk, B., de Andrade, A. R. V., Tomlinson, M. M., Hargraves, R. P., & Bickman, L. (2015). Feedback mechanisms of change: How problem alerts reported by youth clients and their caregivers impact clinician-reported session content. *Psychotherapy Research, 25*(6), 678–693. https://doi.org/10.1080/10503307.2015.1059966.

Fitzpatrick, J. (2002). A conversation with Leonard Bickman. *American Journal of Evaluation, 23*(1), 69–80. https://doi.org/10.1016/s1098-2140(01)00164-3.

Fitzpatrick, J., Christie, C., & Mark, M. M. (2009). The evaluation of the Ft. Bragg and Stark County systems of care for children and adolescents: An interview with Len Bickman. In *Evaluation in action: Interviews with expert evaluators* (pp. 69–92). Thousand Oaks, CA: Sage. https://doi.org/10.4135/9781412990288.d17.

Gleacher, A. A., Olin, S. S., Nadeem, E., Pollock, M., Ringle, V., Bickman, L., … Hoagwood, K. (2016). Implementing a measurement feedback system in community mental health clinics: A case study of multilevel barriers and facilitators. *Administration and Policy in Mental Health and Mental Health Services Research, 43*(3), 426–440. https://doi.org/10.1007/s10488-015-0642-0.

Guttentag, M., & Struening, E. L. (Eds.). (1975). *Handbook of evaluation research (v 2)*. Thousand Oaks, CA: Sage Publications.

Joint Commission. (2011). *Comprehensive accreditation manual for behavioral health care*. Oakbrook Terrace, IL: Joint Commission Resources.

Karver, M. S., Handelsman, J. B., Fields, S., & Bickman, L. (2006). Meta-analysis of therapeutic relationship variables in youth and family therapy: The evidence for different relationship variables in the child and adolescent treatment outcome literature. *Clinical Psychological Review, 26*(1), 50–65. https://doi.org/10.1016/j.cpr.2005.09.001.

Leviton, L. C., Khan, L. K., Rog, D., Dawkins, N., & Cotton, D. (2010). Evaluability assessment to improve publich health policies, programs, and practices. *Annual Review of Public Health, 31*, 213–233. https://doi.org/10.1146/annurev.publhealth.012809.103625.

Lewis, C. C., Boyd, M., Puspitasari, A., Navarro, E., Howard, J., Kassab, H., et al. (2019). Implementing measurement-based care in behavioral health: A review. *JAMA Psychiatry, 76*(3), 324–335. https://doi.org/10.1001/jamapsychiatry.2018.3329.

Riemer, M., Athay, M. M., Bickman, L., Breda, C., Douglas Kelley, S., & Vides de Andrade, A. R. (2012). The peabody treatment progress battery: History and methods for developing a comprehensive measurement battery for youth mental health. *Administration and Policy in Mental Health and Mental Health Services Research, 39*(1–2), 3–12. https://doi.org/10.1007/s10488-012-0404-1.

Riemer, M., & Bickman, L. (2011). Using program theory to link social psychology and program evaluation. In M. M. Mark, S. I. Donaldson, & B. Campbell (Eds.), *Social psychology and evaluation* (pp. 102–139). New York, NY: Guilford Press.

Riemer, M., Rosof-Williams, J., & Bickman, L. (2005). Theories related to changing clinician practice. *Child and Adolescent Psychiatric Clinics of North America, 14*(2), 241–254. https://doi.org/10.1016/j.chc.2004.05.002.

Rivlin, L. G. (1992). A tribute to Harold M. Proshansky (1920–1990). *Journal of Environmental Psychology, 12*(1), 1–4. https://doi.org/10.1016/s0272-4944(05)80292-0.

Rotheram-Borus, M. J., Swendeman, D., & Chorpita, B. F. (2012). Disruptive innovations for designing and diffusing evidence- based interventions. *American Psychologist, 67*(6), 463–476. https://doi.org/10.1037/a0028180.

Scott, K., & Lewis, C. C. (2015). Using measurement-based care to enhance any treatment. *Cognitive and Behavioral Practice, 22*(1), 49–59. https://doi.org/10.1016/j.cbpra.2014.01.010.