Section II

Assessment and Measurement
Evidence-Based Standards and Methodological Issues in School Violence and Related Prevention Research in Education and the Allied Disciplines

Matthew J. Mayer
RUTGERS, THE STATE UNIVERSITY OF NEW JERSEY, NEW BRUNSWICK

Abstract
Standards for evidence-based research across education and allied social and behavioral disciplines are examined, with connections to research in school violence prevention and school safety promotion. The first section summarizes recent evidence-based developments across these allied disciplines, followed by an overview of major nongovernmental evidence-based organizations. Next, recent and emerging evidence-based protocols are examined. The fourth section discusses critical issues that cut across allied disciplines, including reviews of protocols and evidence-based clearinghouses, types and hierarchies of evidence debates, efficacy/effectiveness research, translational research, systematic reviews and meta-analysis, developmental issues, manualized treatments and adaptability, and journal constraints for research manuscripts. The final section considers overall implications for current and future research, policy, and practice in school violence prevention research.

Introduction
Education and allied disciplines address the needs of children and adolescents at risk for displaying or otherwise experiencing aggressive, antisocial, or delinquent behaviors. These disciplines share a common professional heritage linked to research, policy, and practice that target school violence and school safety issues. Disciplinary boundaries are often fuzzy, and tools for evidence-based evaluation of prevention and intervention programs while varying, often overlap. In turn, school safety efficacy and effectiveness research can be subject to differing evidence-based standards, depending on the disciplines involved.

This chapter provides an overview of recent developments in evidence-based standards across education and allied disciplines, connects discussion of discipline-specific developments to school violence and school safety issues, and highlights cross-cutting issues requiring further study and
attention with regard to evidence-based research in related social and behavioral fields. Section 1 summarizes recent evidence-based developments in education, special education, school and clinical psychology, public health, social work, and juvenile justice, with attention to research standards as well as translation to practice issues. Section 2 provides an overview of several nongovernmental evidence-based organizations, including the Cochrane and Campbell Collaborations. In Section 3, recent and emerging protocols aligned with the evidence-based movement are examined. Section 4 discusses critical issues that cut across allied disciplines, including a range of review protocols and evidence-based clearinghouses, types and hierarchies of evidence debates, efficacy/effectiveness research, translational research, systematic reviews and meta-analysis, developmental issues, manualized treatments and adaptability, and journal constraints for research manuscripts. Implications for research in school violence prevention are discussed for each of these issues. Section 5 considers overall implications for current and future research, policy, and practice.

Section 1: Recent Evidence-Based Developments Across Multiple Disciplines

Education

The No Child Left Behind Act (2001), using the terminology, “scientifically-based research,” marked a significant shift in educational intervention research, placing an increased emphasis on more rigorous methodology. This emphasis has been recently reinforced with requirements for evidence of effectiveness among most federally funded K–12 intervention research programs. Following enactment of the No Child Left Behind Law, the Department of Education reorganized its research branch, creating the Institute for Educational Sciences (IES). Presentations by then IES Director, Grover Whitehurst, began to emphasize a new approach to educational research, with a strong commitment to randomized controlled trials (RCT).

The IES created the What Works ClearingHouse (WWC), a federally funded venture, originally contracted to American Institutes for Research and the Campbell Collaboration. The WWC contract was awarded to the Mathematica Policy Research group in 2007. The purpose of the WWC is to develop and implement standards for evaluating educational research and disseminate evidence on educational interventions. The WWC produces three main types of informational products: (a) intervention reports, which evaluate bodies of evidence on particular interventions across identified topic areas (11 as of September, 2010); (b) quick reviews (beginning in mid-2008), which provide a summary and evaluation of recently published effectiveness studies on topics aligned with WWC work, and (c) practice guides, which synthesize best practices drawn from rigorous empirical research. The WWC also produces multimedia resources and methodological references regarding their procedures and protocols, and registries of evaluation researchers and randomized controlled trials.

WWC reports are produced through a three-stage process of initial screen, quality review, and research synthesis. As of the December, 2008 Version–2 WWC standards documentation, under eligibility screening, WWC identified four types of eligible research designs for their reviews: randomized controlled trial, quasi-experimental, regression discontinuity, and single subject (p.10). In mid-2010, the WWC published methodological standards for regression discontinuity and single case research designs.

While the WWC has produced a large body of high quality material within a relatively brief time span and has made valued contributions that have helped improve educational research, policy, and practice, it has also been criticized in several arenas. Slavin (2008) noted that WWC procedures that exclude otherwise informative studies may produce recommendations based on small-scale randomized experiments or extremely brief interventions. Slavin cited the Saxon Math evaluation report, which was based on an unpublished study of 46 students instructed by
one teacher. He also noted an example from English Language Learners research where the one program earning a positive effects report was based on a four-week study. Slavin’s criticisms of the WWC were addressed in a companion article in the same issue by Dynarski (2008). Slavin and others (Song & Herman, 2010) have criticized inclusion of studies where the outcome measures closely align with the experimental group treatment, but not the control group, which tends to inflate effect sizes. Song and Herman argued that if an experiment purports to test the effect of an intervention on subjects learning a skill, both the experimental and comparison group should be taught the same skill, but the teaching approach or method should be different between groups, to test the intervention’s effect. The Government Accountability Office (GAO) issued a report (2010) in which they generally found the WWC standards for evidence-based research appropriate, but criticized the WWC for a lack of timely process and a need for improved dissemination methods.

**Special Education**

Students in special education—particularly those with emotional disturbance—may engage in fighting and are suspended for disruptive and sometimes aggressive or violent behaviors at significantly higher rates than the general student population, signaling a need for quality research on preventing such behaviors (Mayer & Leone, 2007). While many children and adolescents who can benefit from behavior management and violence prevention efforts are served by special education, teachers of students with disabilities often use practices that are known to be ineffective (Cook, Tankersley, & Landrum, 2009). The Council for Exceptional Children Division of Research (CEC-DR) created a task force in early 2003 to (a) identify and establish quality indicators for specific research methodologies appropriate for rigorous scientific investigation in special education and (b) determine how research results from each methodology could inform effective practice. The task force identified four types of research for which quality indicators would be developed: experimental group, correlational, single subject, and qualitative (Odom et al., 2005). Odom and colleagues noted several challenges facing researchers in special education: (a) heterogeneity of special education students can make equivalent group designs problematic; (b) some disabilities are low incidence, making subject availability for large group designs impossible; (c) federal disability law may preclude the use of no-treatment groups that might deny legally mandated entitlements to students; and (d) widely used grouping practices within special education may prevent investigators from using individuals as the unit of analysis.

The CEC-DR task force crafted two sets of quality indicators (QI) governing experimental and quasi-experimental research: “essential quality indicators” and “desirable quality indicators.” These two sets of indicators were proposed for selecting “high” and “acceptable” quality research publications. High quality research must meet all but one of the 10 essential indicators and at least four of the eight desirable indicators. Acceptable research must meet all but one of the 10 essential indicators and at least one of the desirable indicators. The essential QIs require detailed information on participant description, comparable group design, intervention implementers, intervention details, fidelity of implementation, comparison condition services, multiple appropriate measures, timing of outcome measurement, appropriate data analysis techniques, and effect size calculations. The desirable QIs addressed attrition data, reliability and validity measures, follow-up measurements, qualitative aspects of fidelity of implementation, documentation on comparison conditions, audio/video documentation of intervention, and appropriate and clear presentation of results. A 2009 special issue of the research journal, *Exceptional Children*, examined evidence-based practices in special education, applying the CEC-DR quality indicators to reviews of intervention research, finding considerable variability and some problems meeting QIs (Cook et al., 2009).
Research questions that center on low incidence populations as well as other group entities such as schools, which can be treated as single units of analysis, are often studied using single case designs (SCDs), also known as single subject designs. Single case designs are often used with studies of behavioral management of aggressive behavior. The What Works Clearinghouse officially adopted SCDs with the 2008 Version-2 standards and published technical standards for SCDs in June 2010 (Kratochwill et al., 2010). Noteworthy in those standards is a requirement that for a SCD study to fully meet WWC evidence standards (a) the independent variable must be systematically manipulated; (b) outcome variables need to be measured by more than one assessor with specified calculations of interassessor agreement meeting WWC standards; (c) at least three attempts must occur across different times or phase repetitions; and (d) each phase must have at least three data points (at least 4 phases and 5 data points per phase for reversal/wrath with drawal designs, at least 6 phases and 5 data points per phase for multiple baseline designs). For purposes of combining studies to generate analysis in a WWC Intervention Report, at least five SCD research papers, conducted by at least three different research teams, with at least 20 experiments must be included (Kratochwill et al., 2010).

Clinical and School Psychology

The psychology community, particularly researchers in school and child clinical psychology, took the lead in the evidence-based standards movement in the mid-1990s. The APA Division 12 (clinical) Taskforce on Promotion and Dissemination of Psychological Procedures articulated standards for “well-established,” “probably efficacious,” and “experimental” treatments (Hoagwood & Johnson, 2003; Lonigan, Elbert, & Johnson, 1998; Ollendick & King, 2004) (see Table 19.1). The Taskforce had previously introduced the term “empirically validated,” which was contentious, especially since it could imply that the question of effectiveness was finalized (Chorpita, 2003; Ollendick & King, 2004). Both the well-established and probably efficacious criteria required at least two experimental demonstrations of effect, but the well-established criterion required superior effect compared to a placebo or alternate treatment, where the probably efficacious required the effect to be superior to a wait-list control group. Both standards required an intervention treatment manual and sample characteristic information. Additional criteria were identified for single case designs, with stricter standards for the well-established criterion.

The Division 12 Taskforce report was oriented mainly to questions of “efficacy” as opposed to “effectiveness” (Chambless & Hollon, 1998). Efficacy addresses whether the experimental treatment works, usually tested under well-controlled laboratory-like conditions. Effectiveness studies test whether and how well treatments work under more real-world conditions where the treatments are typically provided (Lonigan et al., 1998). Several researchers have spoken to the substantive differences between research-based, and real-world clinical experiments and implications for developing a knowledge base on effective interventions, demonstrating external validity (Chorpita, 2003; Kazdin, 2008; Ollendick & King, 2004; Schoenwald & Hoagwood, 2001; Weisz & Hawley, 1998).

Building on work by the APA Division 12 Taskforce in the early- to mid-1990s, as well as input from APA Division 17 (Counseling Psychology), the APA Division 16 (School Psychology) Taskforce developed the Procedural and Coding Manual for Review of Evidence-Based Interventions. There were concerns regarding the Division 12 criteria that led to an independent criteria development effort by Division 16 (Kratochwill & Stoiber, 2002), including over-reliance on randomized clinical trials, DSM diagnostic categories, and manualized treatments, as well as insufficient attention to comorbidity issues. The Division 16 approach addressed four types of research: group designs, single subject, qualitative, and program evaluation.

More recently, the APA Task Force on Evidence-Based Practice with Children and Adolescents issued a technical report: Disseminating Evidence-Based Practice for Children and Adolescents.
Evidence-Based Standards

The document presents a framework for evidence-based practice in psychology (EBPP), which acknowledges the roles of multiple stakeholders in treatment, multidisciplinary involvement with clients (schools, juvenile justice, social work, community mental health, and so forth), the collaborative nature of psychological interventions, family and cultural responsiveness, and the need to adapt interventions to local context. Clinical practice that follows principles of EBPP relies particularly on three elements:

1. assessment that guides diagnosis, intervention planning, and outcome evaluation;
2. intervention that includes, but is not limited to, those treatment programs for which randomized controlled trials have shown empirical support for the target populations and ecologies; and
3. ongoing monitoring, including client or participant feedback, conducted in a scientifically minded manner and informed by clinical expertise (p. 9).

**Social Work**

School social workers typically address needs of students engaged in violent and disruptive behaviors, provide conflict resolution and peer mediation services, and are often front line service providers in states such as Michigan, where most schools are staffed by a school social worker. Although the field of social work often interfaces with that of education and school psychology,
Evidence-based research developments in social work have been markedly different from the former fields of study. Many in the theorist, researcher, and practitioner trainer community in the field of social work have embraced a perspective of evidence-based practice (EBP), an outgrowth of the evidence-based medicine movement, typically associated with the work of David Sackett and the Evidence-Based Medicine Working Group at McMaster University (Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996). As opposed to the more discrete and targeted efforts (discussed above) in education and school psychology to create and use tools to measure the value of specific intervention research studies, EBP is a systemic approach (see Figure 19.1) that uses clinical expertise to integrate clinical circumstances, best scientific evidence, and client needs and preferences to support more well-attuned and effective service delivery (Shlonsky & Gibbs, 2004; Shlonsky & Wagner, 2005). The EBP model incorporates a multistep procedure of: (a) asking an answerable question regarding the need, (b) locating the best evidence, (c) ascertaining the validity and utility of the evidence, (d) deciding if the presenting problem can be addressed, (e) advising and conferring with stakeholders, (f) developing an integrated action plan, and (g) implementing the plan with ongoing process and outcome monitoring and evaluation. As part of the EBP process, Gibbs (2004) outlined the client oriented practical evidence search (COPES) model that assists practitioners in posing and then answering critical questions that will guide their intervention efforts. The COPES model parallels the EBP model in being client-oriented, having practical relevance in the context of daily practice, and supporting efficient searches for evidence-based methods. The model supports five types of questions, addressing: (a) effectiveness, (b) prevention, (c) assessment, (d) risk/prognosis, and (e) description.

Proponents of EBP note that a sizable segment of the social work practitioner community defer mainly to advice and directions from professional colleagues, personal experience, and professional publications for practice (Howard, McMillen, & Pollio, 2003, p. 235) and do not neces-

---

**Figure 19.1** Shlonsky and Wagner (2005) Adaptation of Revised Evidence-Based Practice Model. *Note.* Reprinted from *Children and Youth Services Review* (Vol. 27), Shlonsky, A., & Wagner, D. The next step: Integrating actuarial risk assessment and clinical judgment into an evidence-based practice framework in CPS case management, pp. 409–427, Copyright 2005, with permission from Elsevier.
Evidence-Based Standards

sarily follow “practice-related research findings” (Gibbs & Gambrill, 2002, p. 452). Rosen (2003) discussed challenges to the social work field in adopting EBP, citing three fundamental barriers: (a) personal attitudes, beliefs, and experiential knowledge among social workers that may conflict with evidence-based practice; (b) differing orientations to knowledge, related in part to a resistance to a logical positivist paradigm, along with an orientation to relative constructs of social justice and reform; and (c) intuitive applications of empirically generalized knowledge, where practitioners avoid or supplant implementation of empirically-based interventions, based on individual belief.

While EBP has been a defining feature in social work, other significant developments have occurred over the past two decades. In 1991, the Task Force on Social Work Research (created in 1988 by the Director of NIMH) presented a report—Building Social Work Knowledge for Effective Services and Policies: A Plan for Research Development—to a committee of the National Institutes on Mental Health (NIMH; Austin, 1998; Zlotnik, Biegel, & Solt, 2002). As a result of that work, several social work research centers were created as well as the Institute for the Advancement of Social Work Research (IASWR, 2003), originally, a collaboration among five social work professional bodies. The IASWR produced a report (Institute for the Advancement of Social Work Research, 2007) on evidence-based practice for NIMH—Partnerships to Integrate Evidence-based Mental Health Practices into Social Work Education and Research—as an outgrowth of a 2007 symposium hosted by NIMH.

Juvenile Justice

The field of juvenile justice has benefited from multiple distributed efforts to support evidence-based practice, but, on the whole, there has been a less well-focused effort within the field in this regard. Juvenile justice as a field has not pursued the development of instruments to evaluate specific research publications, as have the fields of education, psychology, and medicine. There have been multiple efforts along these lines in the past decade, such as the Maryland Scale of Scientific Methods (no longer in general use), and the Standardized Program Evaluation Protocol (discussed later in this chapter). Also, the field in large part has not subscribed to a framework such as evidence-based practice, as has been the case in social work. At the same time, there have been multiple developments in the research literature and the establishment of and use of evaluation research centers to guide evidence-based best practices.

The Justice Statistics and Research Association, a project funded by the Office of Juvenile Justice and Delinquency Prevention (OJJDP) in the U.S. Department of Justice, created the Juvenile Justice Evaluation Center (JJEC), which was active from 1998 to 2007. The JJEC developed a briefing series of nontechnical publications to assist administrators and practitioners in addressing evaluation issues, and JJEC also provided technical assistance to the states, online publications on evaluation issues, referral to other evaluation resources, summary information on evaluations of interventions that are organized topically, funding seed projects, and related activities.

The Center for Program Evaluation and Performance Measurement (CPEPM), an online resource for program evaluation and performance measurement available to assist state and local criminal justice agencies, field practitioners, researchers, and evaluators, is currently available (as of fall 2010) and administered by the Bureau of Justice Assistance. The Center includes an innovative online resource tool that covers a wide array of justice-related issues (e.g., adjudications, corrections, crime prevention, mental health, substance abuse), providing summaries of the research base and answering questions such as what have we learned from evaluations conducted in this area and how can evaluation findings be used for program development and improvement? This type of approach aligns well with calls from the empirical literature to better translate research to real-world settings in a practical manner. The Center offers Web links to a wide array of evidence-based clearinghouses
and other sources for information on specific interventions. The CPEPM also provides technical assistance to states and localities, such as on-site training on evaluation methods, support creating and using performance measures, crafting evaluation plans, and logic model development.

The juvenile justice research and practitioner community has also relied on a combination of resources for guidance, including the Society for Prevention Research, the Cochrane and Campbell Collaborations, the Center for Evaluation Research and Methodology at Vanderbilt University, and similar evaluation research efforts at other universities. The Campbell Crime and Justice Coordinating Group, a component of the Campbell Collaboration, is an international effort among scholars from over 10 nations to provide rigorous systematic reviews and evaluations of research in criminology and justice-related areas. More recently (2008), this group has launched a project that considers extending the CONSORT protocol (see discussion of protocols below) that is used in the healthcare field to criminal justice trials.

Howell and Lipsey (2004) discussed efforts to use research on evidence-based programs for juvenile delinquency. They identified three basic approaches: (a) replication of model programs, such as the “Blueprints” program at the Center for the Study and Prevention of Violence at University of Colorado; (b) conducting evaluations of many individual programs; and (c) defining characteristics of successful program principles and practices from research and applying them. In addition to discussing the benefits of meta-analysis, Howell and Lipsey (2004) described the Standardized Program Evaluation Protocol (SPEP), an instrument that applies principles gathered from characteristics of effective intervention programs and scores individual programs, relative to the instrument’s domains. The SPEP instrumentation has since been applied in several states, including North Carolina (Lipsey, Howell, & Tidd, 2007) and Arizona (Lipsey, 2008), demonstrating how the instrument identifies programs that foster improvements in juvenile offender outcomes (Howell, 2009).

As has been the case in allied fields (National Research Council, 2002), evaluations of intervention programs funded by the National Institute of Justice (NIJ) have come under scrutiny and criticism. For example, a Government Accountability Office (GAO) study (GAO, 2003), Justice Outcome Evaluations: Design and Implementation of Studies Require More NIJ Attention, raised serious criticism of a sample of 15 (out of 96) NIJ funded evaluation studies, where problems emerged with fidelity of implementation and insufficient outcome data. These issues are not unique to the field of juvenile justice and are indicative of the challenges facing allied disciplines in ensuring quality evaluation studies.

**Public Health**

The field of public health maintains a long tradition of ongoing monitoring and analysis of public health events along with a commitment to developing and refining research standards. For example, the Centers for Disease Control and Prevention (CDC) launched a series of 15 CDC-funded evaluation projects addressing youth violence prevention in the early 1990s (Powell et al., 1996), with a strong emphasis on rigorous science, including use of randomization and control groups. Descriptions and baseline data for these projects were presented in a 1996 special issue of the *American Journal of Preventive Medicine*. CDC has supported high quality research in violence prevention through publications of supporting materials such as the National Center for Injury Prevention and Control’s *Measuring Violence-Related Attitudes, Beliefs, and Behaviors Among Youths: A Compendium of Assessment Tools* (Dahlberg, Toal, Swahn, & Behrens, 2005).

Using the term “program” to represent any public health action, from a highly focused direct intervention with individual clients to more broad-based community-level mobilization, the CDC Evaluation Working Group developed the Framework for Program Evaluation in Public Health (Centers for Disease Control, 1999). The Framework identifies the following steps in evaluation: *
Evidence-Based Standards

practice: (a) engage stakeholders, (b) describe the program, (c) focus the evaluation design, (d) gather credible evidence, (f) justify conclusions, and (g) ensure use and sharing of lessons learned (see Figure 19.2). This framework applies the Joint Committee on Evaluation Standards’ four core concepts: utility, feasibility, propriety, and accuracy (Joint Committee on Standards for Educational Evaluation, 1994). The framework includes specific standards for research methods, including experimental, quasi-experimental, and observational designs, with an acknowledgement of the relative strengths and weaknesses of all methods and a suggestion that mixed-method approaches can offer advantages to the researcher. Unlike some recent developments in fields such as education and school psychology, where specific instrumentation has been developed to rate the strength of evidence of published research, this framework does not provide a scoring rubric as such; rather, it is a more general structure to drive the process of quality evaluation.

The CDC Office of Surveillance, Epidemiology, and Laboratory Services (OSELS) directs multiple activities including investigating and promoting best practices grounded in evidence-based research. The OSELS Epidemiology and Analysis Program Office produces the Community Guide, an online resource which provides systematic reviews of research on a wide array of public health topics, following a rigorous set of systematic review standards that were developed and influenced by the Cochrane Collaboration standards, methodological standards from the health and social science empirical literature, and expert review (Briss et al., 2000; Zaza et al., 2000). The Community Guide addresses interventions that prevent disease or injury, or promote health, for a defined population. These systematic reviews follow three highly structured procedural sections: classification information (addressing study information, including type of design); descriptive information (documenting methods and results, including setting, population, measurement tools, analytic methods, effect measures, study power, and so forth); and execution information (documenting descriptions, sampling, measurement, analysis, and interpretation.)

Figure 19.2 CDC framework for program evaluation
of results). Strength of evidence is assessed relative to evidence of effect, quality of execution, design suitability, number of studies, consistency of study results, and effect sizes (see Table 19.2).

The Substance Abuse and Mental Health Services Administration (SAMHSA) has been active in developing protocols for reviewing effective science-based prevention programs (Schinke, Brounstein, & Gardner, 2002). SAMHSA’s Center for Substance Abuse Prevention (CSAP) created the National Registry of Effective Prevention Programs (NREPP), which reviews and identifies evidence-based programs based on experimental, quasi-experimental, time-series, and ethnographic research. SAMHSA reviewed programs were classified into (a) promising programs, (b) effective programs, and (c) model programs, but that framework and set of procedures were abandoned in favor of a revised approach that went into effect in 2007 and was developed following a public call for comments posted in the Federal Register in 2005 (Hennessy, Finkbiner, & Hill, 2006).

The revised system, which took effect in 2007, uses a 16-item rubric addressing areas similar to the previous approach (see Mayer, 2006), and uses a 3-stage review process: prereview, review, and reporting. Eligible studies must demonstrate one or more behavioral outcomes ($p < .05$) in mental health or substance abuse for individuals, groups, or communities. Under the rede-

| Table 19.2 Community Guide Approach to Assessing the Strength of a Body of Evidence |
|---------------------------------|----------------|----------------|----------------|----------------|----------------|
| Evidence of effectiveness       | Execution—    | Design         | Number of     | Consistent     | Effect size    | Expert opinion |
|                                 | Good or faira | Suitability—   | studies       |                |               |                |
|                                 |               | Greatest,      |               |                |               |                |
|                                 |               | moderate, or   |               |                |               |                |
|                                 |               | least          |               |                |               |                |
| Strong                          | Good          | Greatest       | At least 2    | Yes            | Sufficient     | Not used       |
|                                 | Good          | Greatest or    | At least 5    | Yes            | Sufficient     | Not used       |
|                                 | Good or Fair  | Moderate       |               |                |               |                |
|                                 |               | Greatest       | At least 5    | Yes            | Sufficient     | Not used       |
|                                 |               |               |               |                | Large          | Not used       |
|                                 | Meet design,  |               |               |                |               |                |
|                                 | execution,    |               |               |                |               |                |
|                                 | number, and   |               |               |                |               |                |
|                                 | consistency   |               |               |                |               |                |
|                                 | criteria      |               |               |                |               |                |
|                                 | for sufficient but not strong evidence | | | | | |
| Sufficient                      | Good          | Greatest       | 1             | Not applicable | Sufficient     | Not used       |
|                                 | Good or Fair  | Greatest or    | At least 3    | Yes            | Sufficient     | Not used       |
|                                 |               | Moderate       |               |                |               |                |
|                                 | Good or Fair  | Greatest,      | At least 5    | Yes            | Sufficient     | Not used       |
|                                 |               | Moderate, or   |               |                |               |                |
|                                 |               | Least          |               |                |               |                |
| Expert                          | Varies        | Varies         | Varies        | Varies         | Sufficient     | Supports a recommendation |
| Opinion                         |               |                |               |                |               |                |
| Insufficient                    | A. Insufficient designs or executions | | | | | |
|                                 | B. Too few studies | | | | | |
|                                 | C. Inconsistent | | | | | |
|                                 | D. Small       | | | | | |
|                                 | E. Not used    | | | | | |

---

a The categories are not mutually exclusive; a body of evidence meeting criteria for more than one of these should be categorized in the highest possible category.

b Studies with limited execution are not used to assess effectiveness.

c Generally consistent in direction and size.

d Sufficient and large effect sizes are defined on a case-by-case basis and are based on Task Force opinion.

e Expert opinion will not be routinely used in the Guide but can affect the classification of a body of evidence as shown.

f Reasons for determination that evidence is insufficient will be described as follows: A. Insufficient designs or executions, B. Too few studies, C. Inconsistent, D. Effect size too small, E. Expert opinion not used. These categories are not mutually exclusive and one or more of these will occur when a body of evidence fails to meet the criteria for strong or sufficient evidence.

Evidence-Based Standards

Sign, information is organized across three dimensions—descriptive, strength of evidence, and readiness for dissemination—with the latter two scored quantitatively. The descriptive dimension includes information on intervention details, outcomes, relevant population and settings, costs, adverse effects, evaluation designs, replications, implementation history, and so forth. The revised system is aligned with a strength-of-evidence framework shown in Figure 19.3. Strength of evidence is scored on a 0–4 scale across six areas: reliability, validity, intervention fidelity, missing data and attrition, potential confounding variables, and appropriateness of analysis. The readiness for dissemination dimension includes 0–4 scales across three areas: availability of materials, availability of training and/or resources to support initial and ongoing implementation, and availability of quality assurance procedures to support initial and ongoing implementation.

**Broader U.S. Federal Approaches**

The U.S. Preventive Services Task Force (USPSTF) is an independent panel of (nonfederal employees) experts in evidence-based medicine and prevention science. The USPSTF conducts formal reviews of research in a wide array of health related areas such as medical screening, counseling, and medications. The USPSTF has a large number of national primary care partners such as the American Academy of Pediatrics, and the American Academy of Physicians. It also partners with major federal entities such as the CDC, FDA, NIH, and Office of the Surgeon General. As of May 2007, the USPSTF issued five possible grades: A—strongly recommended, B—recommended, C—no recommendation, D—not recommended, and I—insufficient evidence to make a recommendation.

In concert with and subsequent to the Government Performance and Results Act (GPRA) of 1993, the Program Assessment Performance Tool (PART) was launched in 2002. This effort, coordinated through the Office of Management and Budget (OMB), ostensibly would evaluate effectiveness of federal programs and link evaluation results to funding decisions. An early evaluation of PART by the GAO (2004) found limitations and inconsistencies in its application.

![Figure 19.3 SAMHSA NREPP strength of evidence pyramid](image_url)
Datta (2007) examined variations in evaluation practice across federal agency programs and found substantial evaluation variations across federal agencies, attributing these differences to a combination of four factors: (a) match between type of program under evaluation and appropriate evaluation designs, (b) agency culture vis-à-vis evaluation methodology, (c) lead evaluators’ preferences, and (d) politics of methodology. Interestingly, in a 2006 PART evaluation of hundreds of federal agency programs, the report found that for over half of the Cabinet departments, less than 10% of their programs were found to meet standards of being effective, but roughly 60%–90% of programs overall for all agencies were determined to be performing (defined as effective, moderately effective, or adequate).

Section 2: Nongovernmental Evidence-Based Organizations

Cochrane Collaboration

The Cochrane Collaboration, which was established in 1993, takes its name from Archibald Cochrane, who, in the 1970s, pointed to disconnects between research and effective medical practice, citing a lack of consistent application of scientific principles (Smith, 1996). The Cochrane Collaboration is an international effort among researchers from over 100 countries to provide structured reviews of interventions in the medicine and healthcare related fields. The centerpiece of the Cochrane Collaboration products is the Cochrane Library, a collection of six databases: Cochrane Database of Systematic Reviews (synthesis reviews of research on interventions effects), Database of Abstracts of Reviews of Effects (DARE; summary and commentary of quality of non-Cochrane reviews), Cochrane Central Register of Controlled Trials (brief details on RCTs from bibliographic sources), Cochrane Methodology Register (bibliography of methodological publications), Health Technology Assessment Database (details of completed and ongoing health technology assessments from international partner organizations), and the NHS Economic Evaluation Database (with cost-effectiveness and related economic analyses assembled and synthesized by partnering organizations).

A Cochrane systematic review of a collection of studies provides a summary followed by concise and accessible information in the following format: (a) background, (b) study objectives, (c) search strategy, (d) selection criteria, (e) data collection and analysis, (f) main results, and (g) reviewers’ conclusions. Cochrane reviews provide details on interventions from reviewed studies, along with conclusions regarding methodological quality and implications for research and practice, and potential conflicts of interest. The Cochrane reviews tend to focus on randomized clinical trials (Schuerman et al., 2002).

Campbell Collaboration

The Campbell Collaboration (C2), which addresses a wide range of social issues, was launched in 2000 and has emulated the work of the Cochrane Collaborative. C2 originally sought to be inclusive in its reviews of research, avoiding sources of study selection bias, including lack of international focus, and publication bias regarding nonsignificant results. Extra measures that included hand searches and communication with experts in a field of study were used to support this mission. The Campbell Collaboration began its efforts with the intent to offer a registry of evaluations of social interventions, called “C2-SPECTR” (Social, Psychological, Educational, and Criminological Trials Register; Schuerman et al., 2002). That effort brought on the order of 10,000 studies into the C2-SPECTR database, but due to relatively loose criteria that focused mainly on studies with control groups and random or quasi-random procedures, studies of lesser quality were included in the database (Petrosino, Boruch, Rounding, McDonald, & Chalmers,
As of late 2008, the Campbell Collaboration Steering Group approved a plan to end C2 involvement with SPECTR as part of a larger strategic plan for the C2. The C2 efforts have expanded in the past decade with the formation of five coordinating groups to support the systematic review process, the first three of which produce systematic reviews in their respective topic areas: Crime and Justice, Education, Social Welfare, Methods (addresses methodological standards for systematic reviews), and User Groups (fosters outreach to increase the research–to–practice connection).

**Society for Prevention Research**

The Society for Prevention Research (SPR), a multidisciplinary organization established in the early 1990s, focuses on prevention science across the education, social, behavioral, and related health sciences, including a strong interest in criminology. In addition to publishing a scholarly journal, *Prevention Science*, SPR launched multiple initiatives including, but not limited to: (a) advocacy for prevention science (promoting use of EBPs across constituencies, fostering braided funding for interventions, furthering adoption of standards of evidence, building data monitoring systems, and promoting rigorous research), and (b) advancing translational research. SPR published a rigorous standards-of-evidence document (Flay et al., 2005), which articulates specifics of research requirements to identify efficacious and effective interventions, and those suitable for broader dissemination. Efficacious interventions must demonstrate evidence from a minimum of two rigorous experiments that (a) apply to known samples representing known populations, (b) use psychometrically appropriate measures and data collection, (c) use appropriate statistical analysis, (d) demonstrate at least one desired effect without iatrogenic effects, and (e) report at least one extended time follow-up. Effective interventions must satisfy conditions for an efficacious intervention and also (a) demonstrate manuals, training, and technical support for adoption by others; (b) be based on a real-world study that used measurement in the setting and conditions natural to the population studied for experimental and control groups; (c) adequately explain the practical significance of the outcome effects; and (d) firmly establish the group to whom the intervention effects can be generalized. Interventions ready for broad dissemination must satisfy all conditions for efficacious and effective interventions, and also must provide evidence of readiness to be taken to scale, cost information, and monitoring and evaluation components.

**Section 3: Protocols for Research Synthesis, Meta-Analysis, and Program Evaluation**

Many protocols for randomized controlled trials and other methodologies have emerged from medicine and the health sciences. While at first consideration, they may seem less applicable to research on school violence in education and the allied social and behavioral disciplines, it is critical to reflect on the evolution of such standards in recent decades. For example, the Campbell Collaboration addresses a wide range of educational and social-behavioral issues. It is modeled on prior work in medicine by the Cochrane Collaboration. An increasing number of journals in education and the social-behavioral sciences, including criminology, are calling for more rigorous reporting, especially on RCTs, modeling on the CONSORT protocol (Perry, Weisburd, & Hewitt, 2010), as discussed later in this chapter. Further, research on many aspects of violence and disruption has bridged the fields of public health and medicine—particularly psychiatry—and cross-disciplinary studies that investigate developmental psychopathology and that can entail neurobiological, psychological, social, and behavioral factors are called for (Cicchetti & Toth, 2009).
**Consort Protocol**

The Consolidated Standards of Reporting Trials (CONSORT), originally drafted in 1996 and revised in 2001, is a protocol for reporting the results of randomized controlled trials (Moher, Schulz, & Altman, 2001; Schulz, Altman, Moher, & the CONSORT Group, 2010). Developed by an international task force of experts in clinical research, statistics, and epidemiology, the CONSORT 2001 protocol provided a flow chart and 22-item checklist to guide presentation of research using randomized controlled trials. CONSORT addresses a wide range of study issues, including, but not limited to sample design, allocation of participants to interventions, intervention details, specific hypotheses, details of outcome measures, randomization procedures, baseline and follow-up data procedures, and so forth. The CONSORT standard has been adopted by a large number of journals and professional organizations in the health care and allied fields. The CONSORT 2001 protocol was replaced with the CONSORT 2010 Statement (Schulz et al., 2010), a 25-item revision based on a decade of methodological evidence and application experience, and linked to over 700 studies in the CONSORT database. Recent developments with the CONSORT include a working group that held a consensus conference in Banff, Canada, in May 2009, developing a CONSORT extension for N-of-1 trials (CENT), but this group has focused primarily on A-B-A-B single subject designs and not multiple baseline research (T. Kratochwill, personal communication, September 1, 2010). Other extensions of the CONSORT under discussion include extensions to criminological RCTs (Perry & Johnson, 2008; Perry et al., 2010) and also pragmatic trials (an RCT designed to address questions of practice) (Zwarenstein et al., 2008).

Several studies have investigated quality of systematic reviews of randomized trials and use and compliance with the CONSORT protocol in the health sciences literature, as well as outcomes associated with its use. Moher and colleagues (2007) analyzed 300 systematic reviews across 132 journals in the health sciences (two third from U.S. authors) as of late 2004, examining the quality of the systematic review process, comparing Cochrane and non-Cochrane reviews. They found alarming inconsistencies and lower quality among non-Cochrane systematic reviews, including minimal reporting of the systematic review protocol used and limited reporting of funding sources in therapeutic research. Examining RCTs in criminal justice, Perry and Johnson (2008) found mixed compliance with elements of the CONSORT, based on a review of 20 RCTs in juvenile justice published between 2001 and 2006. Perry et al., (2010) analyzed 83 RCTs that were previously reviewed by Farrington and Welsh (2005) from an extensively cited article on what criminologists have learned from randomized trials during the prior two decades. They found low levels of descriptive validity, based on application of CONSORT, with low levels of information on randomization procedures, outcome measures, statistical analysis, and key findings. They advocated for adapting CONSORT for use in criminological RCTs.

**Other Evidence-Based Standards**

Several other reporting standards have been widely used across medicine and the related health sciences, with likely future applications to education and the social-behavioral sciences, including work relating to school violence prevention. While discussion of them is beyond the scope of this chapter, they merit identification. The original Quality of Reporting of Meta-Analyses (QUOROM) statement (Moher et al., 1999) was later replaced by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher et al., 2009). PRISMA can be considered a complementary approach to CONSORT, with PRISMA more oriented to meta-analyses and other systematic reviews and CONSORT to primary studies. The Meta-analysis of Observational Studies in Epidemiology (MOOSE) standards are predicated on a need for more systematic treatment of observational research in public health-related fields, as many critical questions could not be addressed by RCTs for ethical as well as practical reasons.
The Transparent Reporting of Evaluations and NonRandomized Designs (TREND) Statement grew out of a recognized need for a more formal review process with quasi-experimental and other nonrandomized research, as well as a range of program evaluations. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement (von Elm et al., 2008) addresses epidemiological studies based on observations, and cohort, case-control, and cross-sectional designs. The Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) protocol is used to evaluate the quality and strength of evidence to guide clinical decision making in medicine and has been endorsed by a wide range of journals and international medical and public health organizations, including the World Health Organization, American College of Physicians, and the Cochrane Collaboration (Guyatt et al., 2008).

Section 4: Cross-Cutting Issues in Evidence-Based Research

While developments in evidence-based intervention research review and practice have varied across disciplines, as illustrated in the previous sections, a number of broader cross-cutting issues emerge that have a bearing on school violence prevention and school safety programming. This section examines (a) the plethora of protocols and clearinghouses, (b) types and hierarchies of evidence debate, (c) efficacy/effectiveness research, (d) translational research, (e) systematic reviews and meta-analysis, (f) developmental levels and processes, (g) manualized treatments and adaptability issues, and (h) research journal space constraints on manuscripts.

A Plethora of Research Evaluation Protocols and Evidence-Based Clearinghouses

This chapter discussed eight specific protocols for research synthesis, meta-analysis, and program evaluation in addition to identifying other protocols and approaches used by a large number of governmental and nongovernmental entities. Further, there are at least five major clearinghouses (What Works Clearinghouse, SAMHSA, CDC Community Guide, Cochrane Collaboration, Campbell Collaboration) for evidence-based reviews of research, along with many other small to midsize centers. Approaches and standards across these clearinghouses vary, and inconsistencies can emerge in evaluations and recommendations. There is the dearth of empirical research on how the results from these protocols and clearinghouses are being used and their effects, an issue raised by several investigators (Bruns & Hoagwood, 2008; Kavanagh, 2009). Another serious limitation with the larger enterprise of evidence-based clearinghouses is the lack of information on transporting interventions to a variety of settings and contexts, an issue further discussed in the sections below on efficacy/effectiveness research and translational research.

Types and Hierarchies of Evidence Debates

While randomized controlled trials have often been referred to as the “gold standard” in intervention research, a growing movement across disciplines questions this assertion and would recast RCTs as an important tool in a larger methodological tool chest (Hansen & Rieper, 2009). There are several dimensions in these debates and issues within each clearly connect to the others. First, there are concerns that experimental research only provides part of the information necessary for implementation of specific interventions to benefit clients, and is neither sufficient, nor complete, when considered as an evidentiary framework. Raudenbush (2005) stressed the centrality of RCTs while arguing that they are not enough and that complementary research needs to explain how evidence-based interventions can be used to maximum effect in a resource-limited system. He wrote about improving educational research and the contributions of methodological diversity:
Experiments, while necessary, are, however, far from sufficient to support the learning required for effective instructional innovation. Other kinds of research are needed to precisely define educational aims, to identify target populations for intervention, to identify the most promising practices, and to clarify challenges and opportunities for effective implementation of those practices.

(2005, p. 27)

Similar concerns have been echoed with emphasis on extended mixed-methods approaches that are temporally focused and that address site specific and contextual factors (Chatterji, 2005). Kazdin (2008) discussed limitations in the evidence-based movement, in part, due to the divide between research and clinical practice. Kazdin and others (Cicchetti & Toth, 2009) have argued for more process research to help explicate mechanisms of change for patient improvement, as well as improved study of moderators and translation of findings to clinical care. This is particularly salient to school safety research when addressing anger and aggression issues of individual youth as well as viewing the institution as the “patient,” as school change processes associated with violence reduction and school safety promotion are not well understood.

Alternate frameworks for the evidence base have been proposed. Kellam and Langevin (2003) articulated a multidimensional framework for considering evidence based on six themes: (a) multiple lines of prevention work currently exist, applied to individual change, small social groups (classroom, peers, family and so forth), and larger political and social entities; (b) definitions of evidence vary across efficacy trials, effectiveness research, sustainability trials, going-to-scale trials, and larger system-wide change; (c) prevention programs and approaches vary as a function of types of population and types of risk, with accompanying variations in evidentiary standards; (d) given a limited resource system, cost-benefit research is essential; (e) collaborative partnerships are necessary in research to establish that interventions can actually work with relevant stakeholders; and (f) a broader group must be involved—above and beyond researchers—to enable movement to a multidimensional framework for evidence-based standards. Kellam and Langevin’s discussion essentially focused on moving the field towards translational research, an issue discussed further in the following sections.

A second dimension involves the hierarchy of evidence concept. Such hierarchies are naturally linked to the framework of research questions and the ultimate purpose of the research, which can vary considerably across disciplines, as well as across types of research, as discussed by Kellam and Langevin (2003). Hansen and Rieper (2009) examined methodological approaches in the evidence-based movement, reporting that while several major evidence-based clearing-houses use similar procedures in systematic reviews, and the WWC, and Cochrane and Campbell Collaborations use similar hierarchies of evidence with RTCs considered strongest, other international groups, such as the EPPI Centre, accept a wider array of research designs. Several researchers have presented hierarchies similar to that used by SAMHSA (see Figure 19.3), with systematic reviews and RCTs at the top, often followed by quasi-experimental designs, and some including time series research, uncontrolled case series studies, case-control observational studies, qualitative case studies, and expert consensus reports. Yet, researchers acknowledge that RCTs, while strong on internal validity, are often relatively weak on external validity (Glasgow & Emmons, 2007; Steckler & McLeroy, 2008). Hansen and Rieper pointed to work of Boaz (2006) and others who have demonstrated multiple approaches to research synthesis, presenting a typology of evidence matrix adapted from prior work by Petticrew and Roberts (2003) (see Table 19.3). The logic of the hierarchy of evidence approach has been questioned, with arguments that typologies of evidence linked to research focus are more relevant than specific hierarchies (Petticrew & Roberts, 2003). The main arguments presented by Petticrew and Roberts for moving to a typology approach are (a) empirical evidence is available that RCTs are not the
<table>
<thead>
<tr>
<th>Research Question</th>
<th>Qualitative research</th>
<th>Survey</th>
<th>Case-control studies</th>
<th>Cohort studies</th>
<th>RCTs</th>
<th>Quasi-experimental studies</th>
<th>Non experimental evaluations</th>
<th>Systematic reviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effectiveness: Does this work? Does doing this work better than doing that?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>++</td>
<td>+</td>
<td></td>
<td>+++</td>
</tr>
<tr>
<td>Process of service delivery: How does it work?</td>
<td>++</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+++</td>
</tr>
<tr>
<td>Salience: Does it matter?</td>
<td>++</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+++</td>
</tr>
<tr>
<td>Safety: Will it do more good than harm?</td>
<td></td>
<td>++</td>
<td>+</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>+++</td>
</tr>
<tr>
<td>Acceptability: Will children/parents be willing to or want to take up the service offered?</td>
<td>++</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+++</td>
</tr>
<tr>
<td>Cost of effectiveness: Is it worth buying this service?</td>
<td></td>
<td></td>
<td></td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td>+++</td>
</tr>
<tr>
<td>Appropriateness: Is this the right service for these children?</td>
<td>++</td>
<td>++</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>++</td>
</tr>
<tr>
<td>Satisfaction with the service: Are users, providers, and other stakeholders satisfied with the service?</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
</tbody>
</table>

**Note.** Reproduced from *Journal of Epidemiology and Community Health* (Vol. 57), Petticrew, M., & Roberts, H. Evidence, hierarchies, and typologies: Horses for courses, pp. 527–529, Copyright 2003, with permission from BMJ Publishing Group Ltd.
only method to produce valid and reliable results, citing work by Concato, Shah, and Horwitz (2000) and others; and (b) different methods are best suited for different research questions (e.g., applying interventions to different real-world contexts). Chou and Helfand (2005) reasoned that evidence syntheses in medicine needs to address not only clinical effect, as best done with RCTs, but investigation of possible harm, which typically requires observational studies. A traditional hierarchy of evidence that rates observational studies low in evidentiary value will tend to exclude such research. This concern translates from medicine to social-behavioral research—and particularly violence prevention approaches—where possible iatrogenic effects of an intervention will likely remain undiscovered in classic RCT designs, but would more likely be detected thorough mixed methods approaches that include observational and qualitative case study data.

A third dimension involves broader questions regarding understanding social phenomena and developing and implementing not only specific evidence-based interventions, but fostering large scale public health interventions and systemic and cultural changes that are likely to directly impact outcomes of interest. Cornell and Mayer (2010) argued that above and beyond specific interventions aligned with varying theoretical orientations, qualitative methodologies offer critical contextual understanding of life experiences, belief systems, and thought processes that can drive student behavior—knowledge essential to effective school safety programming. Rychetnik, Frommer, Hawe, and Shiell (2002) differentiated public health interventions from more specific interventions for individuals or small groups designed to prevent illness or injury, or alter behaviors, arguing that public health interventions are necessarily “complex, programmatic, and context dependent.” Accordingly, RCTs can answer some of the key questions in public health intervention research, but a broader understanding of factors related to successful change at the community level will require multiple methods, including observational (case-control) and qualitative techniques.

**Efficacy/Effectiveness Research**

There are substantial differences in experimental results from efficacy and effectiveness research, with research-based therapy (efficacy research) usually faring better than real-world clinical therapy (effectiveness research; Hoagwood & Johnson, 2003; Lonigan et al., 1998; Weisz, Donenberg, Han, & Weiss, 1995; Weisz & Hawley, 1998). Ollendick and King (2004) and others have noted differences in characteristics of study samples, interventionists, settings, and research conditions (see Table 19.4). At the same time, Ollendick and King, using a brief discussion of three studies, explicate how the lines between efficacy and effectiveness research can become blurred. There are many implications to consider for intervention research generally as well as school safety research with regard to efficacy-effectiveness issues. Three of these issues are highlighted below.

First, study subjects often differ, with efficacy research participants often demonstrating milder presenting problems and limited comorbid conditions, as compared to most subjects receiving interventions in real-world settings (Chambless & Hollon, 1998; Lonigan et al., 1998). School violence prevention efforts at the indicated level (Institute of Medicine, 1994) target youngsters with the highest degree of need. For children and adolescents with entrenched patterns of antisocial behavior, comorbid conditions (e.g., CD, ODD, ADHD), and a combination of harsh family and school failure experiences, there are many variables in play that may not even be approximated in much of the efficacy research on school violence prevention measures. As a result, the generalizability of such efficacy research to these high-needs students remains questionable.

Second, intervention training and caseloads of the interventionists can differ considerably between efficacy and effectiveness studies (Chorpita, 2003; Ollendick & King, 2004). Chorpita suggested that most organizations that deliver interventions lack the resources to train staff in
Evidence-Based Standards

Table 19.4 Characteristics of Research and Clinic Therapy

<table>
<thead>
<tr>
<th>Research therapy</th>
<th>Recruited cases (less severe, study volunteers)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Homogeneous groups</td>
</tr>
<tr>
<td></td>
<td>Narrow or single-problem focus</td>
</tr>
<tr>
<td></td>
<td>Treatment in lab, school settings</td>
</tr>
<tr>
<td></td>
<td>Researcher as therapist</td>
</tr>
<tr>
<td></td>
<td>Very small caseloads</td>
</tr>
<tr>
<td></td>
<td>Heavy pretherapy preparation</td>
</tr>
<tr>
<td></td>
<td>Preplanned, highly structured treatment (manualized)</td>
</tr>
<tr>
<td></td>
<td>Monitoring of therapist behaviour</td>
</tr>
<tr>
<td></td>
<td>Behavioural methods</td>
</tr>
<tr>
<td>Clinic therapy</td>
<td>Clinic-referred cases (more severe, some coerced into treatment)</td>
</tr>
<tr>
<td></td>
<td>Heterogeneous groups</td>
</tr>
<tr>
<td></td>
<td>Broad, multiproblem focus</td>
</tr>
<tr>
<td></td>
<td>Treatment in clinic, hospital settings</td>
</tr>
<tr>
<td></td>
<td>Professional career therapists</td>
</tr>
<tr>
<td></td>
<td>Very large caseloads</td>
</tr>
<tr>
<td></td>
<td>Little/light pretherapy preparation</td>
</tr>
<tr>
<td></td>
<td>Flexible, adjustable treatment (no treatment manual)</td>
</tr>
<tr>
<td></td>
<td>Little monitoring of therapist behaviour</td>
</tr>
<tr>
<td></td>
<td>Nonbehavioural methods</td>
</tr>
</tbody>
</table>


intervention delivery in as thorough a manner as for staff working for an efficacy research study. The implications of this difference for school safety interventions relates directly to fidelity of implementation. With less training and less time to devote (as a result of relatively heavy caseloads), interventionist staff in schools and clinics are prone to offer less thorough service delivery, resulting in diminished outcome effects. This theme is paralleled in some of the meta-analytic findings suggesting that fidelity of implementation was a critical issue for effective programs (Durlak & DuPre, 2008; Wilson, Lipsey, & Derzon, 2003).

Third, intervention settings can differ greatly, resulting in varied research outcomes (Chambless & Hollon, 1998; Hoagwood, Burns, Kiser, Ringiesen, & Schoenwald, 2001; Ollendick & King, 2004; Schoenwald & Hoagwood, 2001). Chambless and Hollon noted: “All things being equal, those studies that most faithfully reproduce the conditions found in actual clinical practice are most likely to produce findings that generalize to those settings” (p. 15). Greenberg (2004) noted the importance of focusing on the “quality and nature of adaptations,” rather than engaging in the “polarizing debate regarding fidelity versus adaptation” (p. 9). The lack of fit between efficacy studies and real-world settings is reflected in the mismatch between (a) the strictly defined intervention and related research requirements and (b) the philosophy, policies, procedures, and staff scheduling and service delivery requirements found in real-world settings (Hoagwood et al., 2001; Shoenwood & Hoagwood, 2001). Similar setting-related problems can easily compromise school violence intervention research. For example, given the pressures of demonstrating adequate yearly progress under the mandates of the No Child Left Behind Act, for purposes of testing a school violence intervention, schools would be hard pressed to take time and resources away from focused academic programs aimed at raising school-wide scores despite the obvious advantages of reducing school violence and disruption.

Translational Research

Translational research, while sometimes having unique meaning within different disciplinary boundaries, is generally defined as studying ways of transporting evidence-based tools and
methods to the real world of practice in naturally occurring settings, along with their typical populations and contextual factors. Translational research has become a high priority in the health and allied sciences as evidenced by major NIH investment. NIH began the Clinical Translational Science Award program in 2006, funding academic centers engaging in translational research. Projections for 2012 suggest total funding of about $500 million, supporting approximately 60 translational research centers (Woolf, 2008).

The Institute of Medicine (IOM) Clinical Research Roundtable identified two fundamental barriers to translational research in medicine (Sung et al., 2003): (a) taking results of basic science research into clinical studies and (b) taking results of clinical research into daily medical practice and health decision-making. These findings can be considered relative to research in school violence prevention by mapping basic research in anger/aggression, social bonding, social communication and problem-solving, self-regulation, and so forth—areas of more basic scientific research in the social and behavioral disciplines—and taking these findings to the level of interventions to help improve individual behaviors and social relationships. This has already been accomplished to some degree, as seen in interventions such as Multisystemic Therapy (Henggeler, Schoenwald, Borduin, Rowland, & Cunningham, 2009) and Lochman's Anger Coping Program (Larson & Lochman, 2010). Glasgow and Emmons (2007) examined challenges to the translational research enterprise and identified key barriers, seen in Figure 19.4. These barriers to dissemination apply to the fields of education and its allied disciplines, in their efforts to promote school safety and reduce school violence and disorder.

Astor, Guerra, and Van Acker (2010) discussed research challenges in school violence prevention, the need for investing in translational science, and complementing those efforts, increased activity nationally developing local and regional surveillance and monitoring systems connected
to school events and behaviors, in much the same way as established public health mechanisms are used for injury prevention, including violence prevention. Schools and other institutions sometimes launch violence prevention programs and specific interventions, not necessarily based on an explicit linkage to data-demonstrated need, but rather, as a result of funding availability, administrative mandates, or other political pressures. Yet taking evidence-based interventions to scale and adapting to local need and context require reliable and valid data on local conditions.

**Systematic Reviews and Meta-Analysis**

Systematic reviews involve explicit objectives, transparent procedures, clear eligibility criteria (inclusion and exclusion), a well-defined and structured search protocol that reduces bias, screening procedures, exhaustive coverage of the literature, full extraction of relevant study data, quantitative analysis as appropriate (may or may not include meta-analysis), and a structured and detailed report (Welsh & Farrington, 2007). Meta-analysis can be considered a subset of the systematic review process. Multiple approaches to systematic reviews exist. For example, Boaz and colleagues (2006) performed five different types of systematic reviews, demonstrating that variation in types of research across multiple disciplines could require alternate methods for systematic reviews and that the conventional approach as exemplified by the Cochrane Collaboration, was not the only appropriate methodology available. But currently, most leading evidence-based organizations rely heavily on meta-analysis for systematic reviews.

Meta-analysis has been widely used to assess the efficacy of interventions in education, psychology, and allied fields (Lipsey & Wilson, 1993). Examining the results of over 300 meta-analytic studies, Lipsey and Wilson demonstrated how broad patterns in intervention research could inform the research and practitioner community and how meta-analysis has repeatedly produced clear evidence regarding treatment efficacy. Lipsey and Wilson (2001) articulated four specific strengths of meta-analysis: (a) meta-analysis uses a well-structured approach to synthesizing research, with documented criteria and procedural steps that can be fully audited and evaluated; (b) meta-analysis uses sophisticated techniques for integrating effect size and significance statistics that help differentiate contributions among included studies, more so than conventional reviews of research; (c) meta-analysis provides the tools to detect effects and relations that may be hidden or difficult to determine using other research review approaches; and (d) meta-analysis offers a systematic and well-defined approach and set of data management tools to process large volumes of information from studies under review.

Along a similar vein, Hunter and Schmidt (1990) discussed contributions of meta-analysis to the advancement of scientific knowledge: (a) meta-analysis in the social and behavioral sciences has demonstrated that prior research findings have not been as ambiguous or conflicting as previously thought and that valuable conclusions could be drawn from these bodies of research, (b) the “cumulativeness of research findings in the behavioral sciences is probably as great as in the physical sciences” (p. 37), and (c) findings from meta-analysis provide critically important direction for future theory development and research efforts.

While appealing in its ability to provide syntheses of prior research, concerns have been raised with regard to this technique. Lipsey and Wilson (1993) offered cautionary discussion of the following: (a) assessing methodological quality of included studies, (b) availability bias, (c) small sample bias, and (d) generalized placebo effect. The authors also commented that pretest and posttest designs typically overestimate effects compared to random samples, and publication bias regarding excluded studies with negative results can inflate effect size results. Later research by Wilson and Lipsey (2001) found that among 319 meta-analyses, study methods explained almost as much variability in meta-analysis results as did the analyzed data on the respective interventions.
Other researchers have also raised concerns about use of meta-analyses in effectiveness research (Kazdin & Weisz, 1998; Kratochwill, 2002; Weisz, Donenberg, Han, & Weiss, 1995; Weisz & Hawley, 1998). These concerns included: (a) screening criteria for including studies; (b) imbalanced coverage of specific techniques (e.g., behavioral/cognitive-behavioral more so than psychodynamic); (c) varying approaches to effect size calculation; (d) representativeness of subjects, interventionists, and treatment conditions to real-world practice; (e) common omission of single subject research; (f) confounding of treatment-outcome interactions; and (g) outcome reporting sources relative to knowledge of subject treatment status.

Meta-analysis, by its very nature, often sifts through large volumes of data. As a result, differences across study characteristics can constrain the fineness of common detail found, resulting in more broad-based findings. For school safety and related research, there can be a tradeoff between (a) a more inclusive approach to meta-analysis that taps into a large body of relevant research, often yielding only general findings; and (b) preserving the ability to uncover the effectiveness of more specific intervention components, based on an analysis of a more focused subset of the research literature. Also, obscuring the view provided by meta-analytic techniques into specific intervention mechanisms that can effectively reduce school violence and disruption problems are confounds among variables that influence outcomes (Kazdin & Weisz, 1998; Weisz & Hawley, 1998). For example, where a meta-analysis may determine that a stand-alone anger management curriculum may have resulted in reduced incidents of violence in school, an interaction effect involving increased teacher attention (outside of the intervention) to, and in support of participating students, may also have significantly contributed to the outcomes.

**Developmental Levels and Processes**

Many psychotherapeutic interventions for children and adolescents have not been developmentally targeted, taking into account specific age-related tasks and milestones that account for developmental differences (Holmbeck, Devine, & Bruno, 2010; Weisz & Hawley, 2002). Holmbeck and colleagues (2010) suggested that effect sizes for child and adolescent interventions that have been in the moderate range would be even larger if these interventions were developmentally oriented. Interestingly, many interventions for adolescents have been adapted upward from those originally designed for children, and downward from those originally designed for adults (Weisz & Hawley, 2002).

Several meta-analyses of cognitive-behavioral interventions have demonstrated significantly larger (about double) effect sizes for interventions targeted at adolescents, compared to younger children. These results may reflect differences in cognitive abilities, or possibly other related developmental attributes. Holmbeck et al. (2010) explicited models in which developmental attributes may serve as either moderating or mediating variables, affecting outcomes. For example, according to Piagetian theory, a student’s cognitive-developmental level would be linked to his or her ability to engage in perspective taking and develop empathy for others, core skill components common to several empirically supported school violence reduction programs. As such, a student’s developmental level could act as a moderating variable. A mediating variable is tied to some type of mechanism that leads from a causal variable to an outcome. As discussed by Holmbeck and colleagues (2010), Kendall and Treadwell (2007) demonstrated how negative self-statements acted as a mediator between intervention and the outcome measure—level of anxiety. Investigation into the degree to which developmental levels have been incorporated in cognitive-behavioral research with adolescents found noteworthy increases from about 26% of reviewed empirical articles for the period, 1990–1998, to about 70% of the reviewed empirical research for articles from 1999–2004, demonstrating a pronounced shift among researchers in this field (Holmbeck, O’Mahar, Abad, Colder, & Updegrove, 2006).
Developmental processes are also a critical aspect of the research on evidenced-based practices, particularly as applied to emotional/behavioral difficulties, more serious psychopathology, and aggressive and violent behaviors in school. For example, Masten and colleagues (2005) explicated a developmental cascade that linked academic difficulties to emerging internalizing and externalizing problems. Loeber and Stouthamer-Loeber (1998) discussed developmental origins of juvenile aggression and presented a now well-accepted empirically derived model for boys that included three developmental pathways: overt, covert, and authority conflict. Dodge and Pettit (2003) considered the complexities associated with conduct disorder and explicated a biopsychosocial model that incorporated biological predispositions, early life experiences, social-ecological factors, and emotional/behavioral social information processing. These examples are congruent with discussion by Cicchetti and Toth (2009), stressing the importance of understanding how individuals diverge from and regain positive functioning over time within a life-course framework. Accordingly, research needs to better identify the dynamic interplay over time of influences and processes that drive pathological trajectories, using multiple levels of analysis, investigating across the ecology of youth, and addressing genetic, neurobiological, social, psychological, and pre- and postnatal environmental influences. Yet, this view suggests a possible dilemma, where such complex approaches to research would seem to defy efforts to systematically identify evidence-based methods. For example, Mayer and Van Acker (2008) identified trade-offs in more complex cognitive-behavioral interventions where new challenges emerged with cost, training, access to clients, implementation fidelity, client attrition, and difficulty teasing out specific effects.

Coverage of developmental processes across the collective body of evidenced-based interventions has been uneven with many lines of research essentially ignoring critical developmental factors as part of their design. The implications for school violence prevention, particularly with interventions that involve cognitive-behavioral techniques, are clear. Specific techniques that are manageable to implement and well-aligned with the child’s development are more likely to succeed.

**Manualized Treatments and Adaptability Issues**

Documentation of treatment protocols in manual form (a) provides a clear definition of the treatment, (b) standardizes procedures, (c) facilitates assessment of implementation fidelity, (d) provides researchers and practitioners with a clear understanding of what was done, and (e) allows others to engage in replication research (Kratochwill & Stoiber, 2000; Ollendick & King, 2004). As discussed previously in this chapter, treatment manuals were required by the APA Division 12 Task Force draft document for an intervention to be classified as either “probably efficacious” or “well established.” The Division 16 protocols required use of a manual along with other conditions in order to earn the higher component scores of “2” or “3” for implementation fidelity.

Durlak and Dupre (2008) reviewed over 500 quantitative studies involving physical and social-emotional-behavioral prevention and promotion interventions for children and adolescents, with 483 studies analyzed across five meta-analyses, finding strong outcome effects linked to implementation fidelity. This research also indicated that implementation is not always a matter of precisely scripted procedures, but also involves adjusting to contextual factors and ensuring fit to the local setting. Durlak and Dupre discussed the challenges in finding the optimal balance between fidelity and adaptation, noting, interventions vary in their ability to be adapted; empirical evidence exists for improved outcomes where strong fidelity with adaptation were used; researchers can improve interventions by learning from local practitioners about the implementation while measuring aspects of implementation; and the most theoretically salient aspects of the intervention must be closely monitored for fidelity and adaptability to enable researchers to inform future research. They also remarked that the more theoretically important aspects of the
intervention should be maintained with the maximum fidelity feasible, while more of the secondary aspects should be tailored to improve ecological fit; and monitoring of implementation fidelity and adaptations are particularly important for subgroups and with reference to culturally bound issues, where fit and outcomes need to be analyzed relative to the fidelity-adaptation balance.

Practitioners have expressed concerns that mandatory manualization of treatment, prescribing a “lockstep” process, is impractical and will harm clients. Real-life clinical practice also involves working with clients and families who may not fully invest in treatment, requiring modification of procedures (Kratochwill & Stoiber, 2000; Weisz & Hawley, 1998). Kendall, Chu, Gifford, Hayes, and Nauta (1998) argued that treatment manuals can be effectively used in a flexible manner. Strict manualization of interventions in school violence prevention programs could compromise effectiveness in cases where the local context and changing conditions demand some flexibility. As discussed by Kratochwill and Stoiber (2002) and Ollendick and King (2004), applied intervention settings typically have barriers to treatment that must be accommodated as well as idiosyncratic attributes of local populations. For example, an after-school social skills training program could be faced with a situation where a group of students must leave early due to parents pulling them out for an alternate activity. The treatment manual may be designed around working with groups of students for a given minimum time and with a given number of students making up the group, for purposes of group activities that comprise part of the training. Without the ability to modify the program to accommodate to changing circumstances, outcomes may be severely compromised.

Research Journal Space Constraints on Manuscripts

The What Works Clearinghouse Study Review Standards, APA’s Division 16 Procedural and Coding Manual, as well as SAMHSA’s NREPP protocols and requirements of the Cochrane and Campbell collaborations, all articulate varying, yet significant levels of documentation of high quality research and strict adherence to scientific standards. Authors of prospective manuscripts seeking publication of intervention research results must develop extensive and lengthy written explanations to satisfy these reviewing organizations. While a laudable goal and supportive of establishing quality reviews of evidence-based intervention research, these requirements may clash with publishers’ typically strict limitations on manuscript length. Many journals advise prospective authors to limit submitted manuscripts to about 25–30 pages, double-spaced. It seems questionable whether the research review protocols can be satisfied, given such constraints. Moher et al. (2007) echoed these concerns in their analysis of publication compliance with CONSORT guidelines. Research journal publishers and affiliated professional groups may need to revisit these guidelines in view of this issue, and also consider alternative solutions, such as linking published articles to online publisher depositories for authors’ methodological appendices.

Section 5: Implications for Current and Future Research and Practice

There are several important implications for research and practice to be drawn from this review (see Table 19.5). Considering the many clearinghouses and other evidence-based organizations, and varied protocols in use across multiple disciplines, a professional seeking clarification on particular evidence-based approaches may feel overwhelmed. This problem connects to the historical fragmentation across the human service disciplines, each embracing different theoretical frameworks, and organizational philosophies, targeting different goals, utilizing different service delivery models, employing varying measurement and accountability tools, and using similar sounding, yet distinct technical terminology.
This issue of differing orientations and practices is not new to the school violence research community. Definitional and measurement issues have led to varying reports of school violence data (Mayer & Furlong, 2010; Mayer & Leone, 2007; Reiss & Roth, 1993; U.S. Library of Congress, 1994). These definitional and measurement challenges remain, as does the need to reconcile the admissibility of varying research designs in evidence-based practice. For example, single subject research has contributed key understandings of effective behavioral interventions with low incidence and other groups in special education environments that have major implications for school safety research and practice (Horner et al., 2005). Yet, it took great effort over many years before the What Works Clearinghouse accepted this methodology for inclusion in reviewing evidence-based interventions.
The implications for researchers in school violence also relate to studying interventions with those students who are most often involved in school violence. Students who are at highest risk and are receiving indicated interventions (Institute of Medicine, 1994) may comprise about 1%–7% of a school’s population (Sugai et al., 1999). Discussing adolescent violence, Dodge and Pettit (2003) noted, “over 50% of violent behaviors are perpetrated by only 6% of the population” (p. 350). This points to a need to examine highly focused interventions for a relatively small and unique group within the school. Due to a lack of heterogeneity and limited availability for group assignments, evaluation for group differences using randomized clinical trials that depend on group assignment would not be practical for these high-risk students in many educational settings. This presents serious challenges for effectiveness research in real-world settings.

Given the relatively bleak state of K–12 funding nationally, combined with mandates to produce improved test scores for adequate yearly progress pursuant to the requirements of the No Child Left Behind Act (and likely, subsequent reauthorized legislation, regardless of the name), schools are struggling to use their personnel in the most efficient ways possible. This leaves little room for releasing and paying for staff training in new intervention programs designed to reduce school violence. This will severely compromise effectiveness research in school violence prevention, as problems may occur with fidelity of implementation, outcome measurements, and related concerns tied to appropriate staff training.

Practitioners face challenges on several levels with regard to violence prevention programming and service delivery. First, although there are a growing number of information clearinghouses available to the practitioner community, the standards of research evaluation underpinning these endorsed programs vary. The end user is forced to proceed in uncertain waters, assuming that, despite different approaches, the approved programs are worth implementing. Second, research design constraints may limit the availability of evidence for interventions addressing special education populations, and especially, the 6% or so of chronic, high-risk adolescents. However, chronic, high-risk students are precisely the group for whom we need to develop more effective school violence prevention programs. Third, continuing disconnects between efficacy and effectiveness research, the difficulty of conducting effectiveness research in settings where high-risk students typically exist, and the tendency favoring efficacy research, all totaled, will likely limit the availability of proven intervention approaches that are rigorously tested in real-world environments. As previously noted by Chorpita (2003), research in transportability, generalizability, and system effectiveness is rare. In sum, the consuming public will probably have limited choices of thoroughly researched interventions that are ready for large-scale implementation. These points are not made to suggest a purely negative scenario. Rather, they draw attention to continuing challenges the research and practice community face in more strategically responding to competing requirements and constraints in evidence-based research.

Developments over the last decade in building improved intervention research evaluation protocols and models have yielded several viable approaches that continue to evolve. While still problematic, and far from perfect, they signal an important shift in thinking. Given limited resources and significant social problems, especially in reducing school violence and promoting school safety practices, the field can no longer afford to implement interventions without a solid foundation of research-based evidence. The research and practitioner community needs to continue to invest in the concept of evidence-based practice and to support continued development efforts so that future research will benefit the clients who need these interventions.

References

Evidence-Based Standards


Evidence-Based Standards


