Report of the National Panel for Evidence-Based School Counseling: Outcome Research Coding Protocol and Evaluation of Student Success Skills and Second Step

John C. Carey

Center for School Counseling Outcome Research,

University of Massachusetts, Amherst

Carey Dimmitt

Center for School Counseling Outcome Research,

University of Massachusetts, Amherst

Trish A. Hatch

School Counseling Program,

San Diego State University

Richard T. Lapan

Department of Educational & Counseling Psychology,

University of Missouri-Columbia, Columbia

Courtland C. Lee

Department of Counseling and Personnel Services,

University of Maryland, College Park

Susan C. Whiston

Department of Counseling and Educational Psychology,

Indiana University

Abstract

The National Panel for School Counseling Evidence-Based Practice was established by the Center for School Counseling Outcome Research to improve the practice of school counseling by helping to develop the research base that is necessary for responsible and effective practice. This article presents the Panel's *Outcome Research Coding Protocol* and its evaluation of the scientific research evidence supporting the effectiveness of the Student Success Skills and Second Step interventions.

Report of the National Panel for Evidence-Based School Counseling: Outcome Research Coding Protocol and Evaluation of Student Success Skills and Second Step

The 2003 ASCA-ACES Research Summit determined the need for the establishment of a National Panel for Evidence-Based School Counseling in order to conduct continuous reviews of the evidence base for the profession, identify needed research studies, and document the consistency of school counseling practice with the *No Child Left Behind* Standards for Evidence-Based Practice. The National Center for School Counseling Outcome Research at the University of Massachusetts Amherst agreed to form this panel and facilitate its work. Since that time, the Panel has reviewed the school counseling outcome research literature (and related literatures from other disciplines); studied the operation of other similar Panels in human service disciplines; developed an Outcome Research Coding Protocol, and has begun using this protocol to evaluate the research base of school counseling interventions. The work of the Panel has been presented at the 2004 and 2005 American School Counselor Association (ASCA) conferences and at the 2005 American Counselor Association (ACA) conference where input from the profession was sought. This input was used to revise Panel procedures and standards.

The National Panel for School Counseling Evidence-Based Practice was established to improve the practice of school counseling by helping to develop the research base that is necessary for responsible and effective practice. The Panel is striving to: 1) provide school counselors, school leaders, policymakers, and the public with independent, unbiased information on the extent to which school counseling

practices are supported by scientific evidence; 2) provide information to practitioners on promising practices; 3) provide school counseling researchers with suggestions about critically needed areas of inquiry; and 4) provide practicing professional school counselors and researchers guidance about measurement and research methodology. The Panel reviews research literature in school counseling and related journals with the intention of locating relevant research findings wherever they exist and connecting school counseling research to the broader interdisciplinary, social science research context. In the future, the panel will disseminate its work through books and monographs as well as continue to make yearly reports at the ASCA national conference and contribute to *Professional School Counseling*. Panel documents and reports are also available on the Center for School Counseling Outcome Research website: www.cscor.org.

The Panel is an independent body that strives to provide comprehensive and unbiased reviews and analyses. The Panel seeks to determine levels of existing evidence and to support the development of research-based school counseling practice by identifying school counseling interventions or approaches that have a demonstrated beneficial causal relationship to important student outcomes. To say that an intervention *caused* an increase in a student outcome such as student achievement, it is necessary to rule out other plausible explanations for the increase in student achievement. Our level of *confidence* that the causal relationship is valid depends on how well this goal is met. The Panel therefore applies rigorous, systematic, and transparent research review methods in order to evaluate the scientific evidence for causal relationships between interventions and outcomes. Increasingly, school counseling models have emphasized the importance of systemic interventions (e.g. interventions that affect a whole school's ability to educate

children). Well-controlled research that leads to causal inferences about systemic interventions is very difficult to accomplish. Nonetheless, such research is vitally important. With regard to systemic interventions, the Panel will work to both evaluate the current state of knowledge by identifying promising practices and to facilitate the development of scientifically credible outcome studies of systemic interventions.

The current state of research in school counseling is that there are too few well-controlled studies of outcomes (Whiston and Sexton, 1998). The Panel is committed to evaluating our current state of knowledge through comprehensive, interdisciplinary reviews and to encourage the development of well designed outcome studies within professional school counseling.

Many forms of research are needed in school counseling. Different types of research serve different functions. The Panel is primarily interested in the studies that attempt to evaluate the extent to which interventions or approaches have an impact on student outcomes. Experimental and certain quasi-experimental research designs are among the most appropriate methodologies for this purpose (Shadish, Cook, & Campbell, 2002) and are the "gold standard" for determining confidence in any claim of causality between intervention and student outcomes. At the same time, experiments and quasi-experimental studies are very difficult to implement in school settings. There is much to be learned through "less rigorous" designs in terms of which interventions or approaches are promising. The Panel will evaluate evidence on the effectiveness of an intervention from many different research methodologies. It will evaluate the strength of scientific evidence of a causal relationship to student outcomes and, (in cases where strong

evidence is not found) identify specific studies that need to be conducted to supply the missing evidence.

Outcome Research Coding Protocol

In order to do its work the Panel needed a standardized method for evaluating the extent to which an intervention is supported by a body of research that yields strong evidence that the intervention leads to better academic, career and/or social/emotional outcomes for students. In constructing this protocol, the Panel reviewed the work of several other professional panels. The resulting Outcome Research Coding Protocol presented in Table 1 was strongly influenced by the What Works Clearing House's *Study Design and Implementation Device (*Valentine & Cooper, 2003) and the Task Force on Evidence-Based Interventions in School Psychology's *Procedures and Coding Manual* (Task Force on Evidence-Based Interventions in School Psychology, 2003).

The primary purpose of the Protocol is to help the Panel evaluate the degree to which a body of research literature related to a specific intervention meets the rigorous, scientific standards needed to conclude that a causal relationship exists between an intervention or approach and student outcomes. The Protocol was also designed to help the Panel identify the limitations in the existing scientific evidence and determine which critical studies need to be conducted to bolster the research base. The Protocol consists of seven domains that include: Measurement, Comparison Groups and Statistical Analysis of Outcome Variables, Implementation Fidelity, Replication, Ecological Validity, and Persistence of Effect.

To begin its work, the Panel selected the Student Success Skills (Brigman & Campbell, 2003; Brigman & Webb, 2004; Webb, Brigman, & Campbell, 2005) and Second Step: A Violence Prevention Curriculum (Committee for Children, 1997a,b) interventions because of the quality of the available research and the importance of the major outcomes (academic achievement enhancement, violence prevention) associated with each intervention. Student Success Skills is a structured group and classroom guidance approach to teaching cognitive/meta-cognitive skills, social skills and self-management skills (e.g. using a memory peg system to remember key facts, active listening, building optimism, and managing test anxiety) that have been identified by research as being related to academic success and that is designed to be most appropriate for grades 5 through 9. Second Step is a structured social-emotional learning program for students in grades K-8. It is designed to reduce violence through increased empathy and social skills and through reductions in aggressive and bullying behaviors.

The Panel's Evaluation of Student Success Skills

The Panel reviewed three studies that support the efficacy of Student Success Skills (SSS) (Brigman and Campbell, 2003; Campbell and Brigman, 2005; Webb, Brigman, and Campbell, 2005).

Domain 1: Measurement. All three studies used the Florida Comprehensive Achievement Test (FCAT) as the primary outcome measure. This test is a criterion referenced state achievement test and meets standards of psychometric rigor. All three studies also used teacher ratings of students using the School Social Behavior Scale (SSBS) which also meets rigorous psychometric standards for reliability and validity.

FCAT scores were used as pre and post-test measures of achievement for both the experimental and control groups. SSBS scores were used as pre and post-test measures of appropriate school social behavior for only experimental group subjects.

In the Measurement domain, the Panel judged that SSS research reflects Strong Evidence based on the FCAT. The Panel noted that, in spite of the psychometric properties of the SSBS, the use of this instrument added little either to the understanding of the mechanisms by which SSS may impact academic achievement or to confidence in the impact of SSS. Specifically, the convincing linkage between process and results data was missing as the study lacked the perception data (i.e. impact of SSS on knowledge, attitudes, skills learned through the SSS process) that may have contributed to the increases in the FCAT. Additionally, the achievement-related data (i.e. actual improvement in students' actual academic skills, social skills and self-management skills) was not measured against a control group. Consequently, the logical links between the nature of the SSS intervention and the constructs measured by the SSBS are neither obvious nor explicitly linked to their impact on student achievement data, specifically their performance on the FCAT. We strongly recommend that future studies of the SSS include measure that reflect the specific constructs targeted by the SSS interventions (e.g. cognitive/meta-cognitive skills, social skills and self-management skills) so that the impact of SSS on these outcomes and the relationships between changes on these variables and increases in academic achievement can be ascertained.

The Panel strongly recommended that future studies of the SSS include measures that reflect the specific constructs targeted by the SSS interventions (e.g. cognitive/meta-cognitive skills, social skills and self-management skills) so that the impact of SSS on

these outcomes and the relationships between changes on these variables and increases in academic achievement can be ascertained. We also recommend that, if adult ratings are employed, raters should be blind to treatment condition.

Domain 2: Comparison Groups. All three SSS studies compared a Treatment Group to an untreated Control Group with academic achievement as the dependent variable. Elements of random assignment were employed to ensure initial group equivalence and covariance analyses were employed to statistically equate groups. All three studies used untreated control groups, meaning that they did not include Active Comparison Groups with alternative treatments. As a result, it is impossible to determine the potential impact of attention or a placebo effect on the outcome measures. With regard to school behavior, all three studies only measured pre-post intervention changes in the Treatment group. The lack of a control group for the school behavior severely limits the confidence that the observed changes in school behavior are caused by the intervention. The Panel judged that, in the Comparison Groups domain, SSS research reflects Promising Evidence with respect to academic achievement. Follow-up studies with placebo control groups are needed to ensure that the achievement effects are related to the learning that takes place related to the intervention rather than to attention or expectations. Strong experimental research designs are also needed to assess the impact of SSS on variables related to achievement (e.g. school behavior, self-management, and self-efficacy) that may mediate or moderate the impact of the intervention on achievement.

Domain 3: Statistical Analysis of Outcome Variables. . In all three studies an ANCOVA using the previous year's FCAT as the covariate and the post treatment FCAT

as the dependent variable found statistically significant results. The three studies included data from 222, 306, and 418 students. In two studies significant effects were reported on both FCAT reading and math tests (Brigman & Campbell, 2003; Campbell & Brigman, 2005). The third study found a significant effect for math but not reading (Webb, Brigman, & Campbell, 2005). Effect sizes for the impact of SSS on FCAT scores were not reported, however Cohen's d effect size statistic was used by the Panel to estimate effect sizes from the post-test FCAT data. All three studies used an ANCOVA design to control but did not report the adjusted means. The Panel had to estimate effect sizes from unadjusted post-test means and decided to do so only for data where significant pretest differences were not found. Effect sizes ranged between .176 and .216 for the FCAT Reading Test and between .142 and .154 for the FCAT Math Test. These effect sizes would be considered small (Welkowitz, Ewen & Cohen, 1971). Based on the effect size, The Panel concluded that SSS research to date reflect Promising but not Strong evidence of effectiveness. A small effect with respect to academic outcomes measured by a state achievement test is not particularly surprising given the multitude of factors that impact achievement. Additional studies documenting the impact of SSS on intermediate variables related to student skills and development (e.g. self-management skill acquisition) that are in turn related to academic achievement is needed. Larger effect sizes would be expected with such outcome variables.

Domain 4: Implementation Fidelity. The SSS is a well-documented, structured intervention that can be delivered with fidelity by trained facilitators (Brigman & Webb, 2004). In all three studies, fidelity was assured through training, peer coaching, weekly checks of content delivery and weekly logs. The three studies used a number of

experienced school counselors to deliver the intervention (10, 25, and 25). The panel concludes that a Strong Evidence rating is more than justified in this domain.

Domain 5: Replication. Three independent studies found equivalent significant Control-Treatment Group differences for FCAT math scores. Two out of the three studies found significant effects for FCAT reading scores. All three SSS studies were conducted by the same research team, who are also involved in the development of the intervention. The Panel finds Promising Evidence of effectiveness in this domain with the caution that the effects of SSS on math achievement may be more robust than on Reading Achievement. In order to achieve a rating of Strong Evidence, SSS should be evaluated by an independent research team.

Domain 6: Ecological Validity. All three studies of SSS were based upon regular public school implementations. Two studies report participant samples with limited racial/ethnic diversity and with a range of socio-economic diversity (82% White, 60% free or reduced lunch; 85% White, 45% free or reduced lunch). In all three studies, the researchers selected participants from students who had scored average or below average on the previous year's FCAT (25th-50th percentile or 25^{th}-60^{th} percentile). The relatively small numbers of non-White students made it impossible to determine whether SSS is more effective with some groups of students. These research studies did not address the issue of whether SSS effectiveness was related to socioeconomic status. Based upon the public school implementations and the clear delineation of generalizability limitations in the research reports, the Panel finds Strong Evidence in this domain with the caveat that evidence exists for the effectiveness of SSS with average to below average achieving students in predominately White schools. The Panel strongly

feels that replication of these SSS outcome studies in more diverse schools and with specific attention to determining whether SSS effectiveness is related to ability, racial/ethnic and/or socio-economic diversity is needed.

Domain 7: Persistence of Effect. None of the three studies investigated the persistence of the effects of SSS on academic achievement beyond the school year in which the intervention occurred. In fact, the research studies are unclear about the time period between the last SSS session and FCAT testing. In the Brigman and Campbell (2003) study the 8-week-session group intervention was completed at the beginning of December but students received booster sessions in January, February, March and April. In Campbell and Brigman (2005) and Webb, Brigman and Campbell (2005), eight weekly group sessions began in October and were followed by with booster sessions in January through April. No study reported the FCAT test date but it is highly likely that there was only a very short time period between the last booster session and the test. Given this short time period, the Panel concludes that Strong or Promising evidence for the persistence of effect for SSS is not yet available.

Summary. SSS is developing a solid research base and has some very important findings. It is particularly important that the positive effects of the SSS intervention on academic achievement as measured by state test scores have been demonstrated. The Panel's evaluations of the evidence supporting the effectiveness of SSS are presented in Table 2. SSS achieved a standard of Strong Evidence in the domains of Measurement, Implementation Fidelity, and Ecological Validity. SSS achieved a standard of Promising Evidence in the domains of Comparison Groups, Replication, and Statistical Analysis of Outcome Variables. SSS failed to achieve a standard of Promising evidence in terms of

Persistence of Effects. The Panel strongly recommends additional research be undertaken to investigate whether the beneficial effects of SSS persist and are generalizable to more diverse student populations. A two year follow up study would provide convincing evidence of persistence but would require very large numbers of participants if the FCAT were the only outcome measure used (given the small effect sizes noted in the existing research).

The Panel recommends expanding the outcome measures used in SSS research to include psychometrically sound measures of variables related to the SSS target skills of cognitive/meta-cognitive skills, social skills and self-management skills. While it is important to document the ultimate impact of school counseling interventions on standardized state test scores, it is likely that the effect sizes of such interventions will be relatively small since test scores are influenced by a multitude of variables. Large N studies with diverse sample will be needed if the only outcome variable is academic achievement measured by standardized test scores. As SSS is designed to develop cognitive/meta-cognitive skills, social skills, and self-management skills, it would be helpful to examine outcome measures related to these skills. It is likely the SSS effect sizes for these outcomes would be moderate or strong since they are closely linked to the intervention. Subsequent research demonstrating that SSS leads to lasting increases in cognitive/meta-cognitive skills, social skills and self-management skills that are in turn related to academic achievement would help establish the mechanisms by which SSS works and offer some very pragmatic advantages to future research. Many of the remaining questions related to SSS efficacy could be investigated without the necessity of using state achievement tests as the outcome variable. The Panel recommends replication

of the SSS findings by independent research teams. Finally, the Panel recommends additional studies that employ "placebo" controls and studies that investigate how effective SSS is with different student subgroups.

The Panel's Evaluation of Second Step

The Second Step Violence Prevention Curriculum (Committee for Children, 1997a,b) is a social and emotional learning program for students K-8. Class plans are for approximately 30 minutes, and are usually taught 1-2 times a week throughout the academic year. The developmentally progressive contents include identifying feelings, solving problems, developing social skills, building empathy, reducing anger, managing stress, resisting peer pressure, dealing with bullying, and defusing potentially violent situations.

The Panel reviewed seven research studies that have appeared in juried journals about Second Step (Frey, Nolen, Van Schoiack-Edstrom, & Hirschstein, 2005; Grossman, et al., 1997; McMahon & Washburn, 2003; McMahon, Washburn, Felix, Yakin, & Childrey, 2000; Orpinas, Parcel, McAlister, & Frankowski, 1995; Taub, 2002; Van Schoiack-Edstrom, Frey, & Beland, 2002). Additional evaluations done by the Committee for Children was not included in the review because of the potential for bias. The Second Step research is well developed and is included in this review as an example of the kinds of research that school counseling interventions will need in order to be considered evidence-based research. The Panel gave the Second Step research Strong Evidence ratings in every domain, although there are certainly still suggestions about future research.

The Grossman et al. (1997) research (supported by a grant by the CDC) is particularly important because it demonstrates that high-quality field research is possible (Rosenberg, Powell, & Hammond, 1997). Schools in the study were randomly assigned to treatment or control groups, several outcomes were measured, behavior observation as well as teacher and parent reports were utilized, and the outcomes were measured over time.

Domain 1: Measurement. Orpinas et al. (1995) used the Aggressive Behavior Scale; Grossman et al. (1997) used the School Social Behavior Scale (SSBS), the Achenbach Teacher Report Form, the Achenbach Child Behavior Checklist, and the Parent-Child Rating Scale; McMahon et al. (2000) used the teacher ratings from The Social Skills Rating System; Taub (2002) used the School Social Behavior Scale; Van Schoiack-Edstrom, Frey, and Beland (2002) used the Endorsement of Aggression Scale and Perceived Social Difficulty Scale; McMahon and Washburn (2003) used the Psychological Sense of School Membership Questionnaire and the Aggressive Behavior Scale. All of these measures meet rigorous psychometric standards for reliability and validity.

Most researchers also used behavioral observations and pre-and post-test content knowledge and skill surveys. Some researchers also used self-report surveys and child interviews to determine self-awareness of behavioral or knowledge changes.

The Panel judged in the Measurement domain, Second Step research reflects Strong Evidence. The intervention has been shown to make a difference on measurement scales which are reliable and valid, and which map onto the constructs which the intervention purports to impact (behavior, empathy, social skills, aggression, and so on).

Most of the studies used behavioral observation methods where the observers and behavioral coders were blind to the status of the children they were observing.

To date, no studies have looked at whether or not Second Step impacts academic achievement. We strongly recommend that future studies of Second Step determine whether the social and behavioral skill development found with this intervention impacts academic outcomes as well.

Domain 2: Comparison Groups. Grossman et al. (1997) used a randomized control trial design with schools as the unit of randomization. Taub (2002) used a non-randomized control group comparison, and Frey et al. (2005), used randomly assigned control and intervention groups. Other studies mostly used repeated measures designs with pre- and post-test measurements. To the extent that it is possible in school-based research, much of the Second Step research to date has effectively utilized control groups either within schools or has used comparable schools as the unit of comparison.

Domain 3: Statistical Analysis of Outcome Variables. All seven reviewed studies used statistical analyses that were thorough, sophisticated, and highly appropriate for assessing change in the outcome variables. Examples of analyses employed included: ANOVA and MANOVA main and interaction effects, as well as repeated measure designs; the generalized estimating equation regression method to deal with specific issues related to covariates because schools (and not individual students) were randomized; and intra-class correlations to establish inter-observer agreement. Of special note, the authors in each of these studies were very clear in pointing out both when and where outcomes did not support the hypothesis that the treatment led to better results for students. In general, these studies adequately controlled for experiment wise error and

had adequate sample sizes. In addition, the researchers are to be commended for including time intensive, direct behavioral observations of student behavior in multiple contexts (e.g., classroom and the playground).

Several concerns were noted. First, many of the reported findings were described as "modest positive effects." While most of these studies employed a control group, an absence of credible competing alternative treatments was evident. A consistent finding in student self-reported knowledge gain and attitude changes was reported but a lack of change in both teacher and parent reports of student behavior was also found. And finally, some researchers pointed to the possible existence of contextual factors (e.g., teacher receptivity, willingness, and engagement throughout the implementation process) that may hamper a clear understanding of construct validity issues related to treatment implementation and the identification of what leads to the reported changes. However, in spite of these limitations the quality of the studies to date led the Panel to conclude that Second Step meets the criteria for Strong Evidence of program effectiveness in this area.

Domain 4: Implementation Fidelity. In all seven research articles about Second Step, the curriculum was taught by teachers, school counselors, school psychologists, or doctoral level graduate students in psychology who had completed between 4 and 16 hours of training in the materials with Committee for Children trainers. In all of the research the intervention occurred over the course of several months as is recommended. In most studies the Second Step information was presented in 30 minute lesson plans one or two times a week.

In the Grossman et al. (1997) study, the entire curriculum was completed by all participating teachers and the quality of implementation was monitored two times. In the

other studies the curriculum completion rate was 95-100%, with the exception of Frey et al. (2005), which had a completion rate between 42% and 100%. Treatment fidelity was monitored in most studies through the use of logs, support teams, administrator supervision, and/or self-report.

Domain 5: Replication. The Second Step research has found consistent evidence across studies that the curriculum increases social skills and pro-social behavior, and decreases anti-social and/or aggressive behavior. Most of the studies evaluated these domains and some have also replicated findings about self-reported attitudes, knowledge and skills. The Frey et al. (2005) article makes some initial links between social cognitions and behavior that have not yet been replicated.

Domain 6: Ecological Validity. All of the Second Step research under consideration in this review has occurred in public schools, and findings have been successfully replicated across contexts with different ages (PreK-8), a range of racial/ethnic groups, with both males and females, and with economically and socially diverse populations. Research has occurred in different parts of the U.S. and Canada, in both urban and rural settings, in large and small schools and with at-risk populations.

Domain 7: Persistence of Effect. The Grossman et al. (1997) study collected data prior to implementation, post-implementation, and 6 months later. They found that there was persistence in effect for behavioral observations, with fewer negative interactions and more neutral interactions among students on the playground and in the cafeteria. Taub (2002) also looked at persistence of effect, conducting evaluations prior to intervention, post-intervention and 1 year later. She found that most, though not all,

effects were maintained or increased over time. Orpinas et al. (1995) conducted pre-intervention, post-intervention and 3-month follow-up evaluations, and found that many of the effects found post-intervention did not continue at the 3 month point.

Summary. Second Step has been endorsed by the U.S. Department of Education as an exemplary Safe, Disciplined, and Drug-Free Schools program (U.S. DOE, 2005). The Substance Abuse and Mental Health Services Administration of the U.S. Department of Health and Human Services has named Second Step a promising program in the domain of violence prevention in schools (U.S. DHHS, 2005). The Committee for Children, which publishes the materials, has a research department that is actively conducting research in conjunction with the University of Washington. The research done to date on Second Step warrants ratings of Strong Evidence in the 7 domains of the Outcome Research Coding Protocol for the National Panel for School Counseling Evidence-Based Practice, but the Panel still has several suggestions and concerns about the research.

Measuring programs in naturalistic settings such as schools creates unique challenges. No comparison site/school, even if similar in demographics and size, can truly replicate another, as each has a unique climate, culture and context. Control schools/sites, unless they have something to gain from participation, may be hesitant to engage in research. Replication is also tricky, as even if all those who implement the intervention receive the equivalent training, they still may not teach it in the same way. It is hard to be blind to condition, so if teacher or parent ratings are used, they are necessarily subjective and not blind.

With the Second Step research, there were efforts to be consistent in the implementation of the curriculum, and most studies reported that most, if not all, of the curriculum was taught. However, complete fidelity of implementation is hard to obtain in schools, where there are multiple demands on teacher time and student attention. The research to date has not been able to identify what it is about the intervention that actually creates change. Hard to measure constructs such as commitment, hopefulness, and belief in the intervention may be as important as the curriculum materials themselves, and may account for some of the differences in outcomes across the studies.

Increasingly, the Second Step research is conducted by researchers supported at least in part by the Committee for Children. The studies discussed in this paper were all published in journals with peer reviews, so potential bias is controlled for to an extent, but this is still an area of concern.

Conclusions

The Panel's evaluations of both the Student Success Skills and Second Step interventions is based upon stringent criteria for the quality of evidence needed to conclude that an intervention causes a positive change in student outcomes. Second Step is an exceptionally well-researched intervention. The scope and quality of its research base has been greatly enabled by federal funding for violence prevention.

Student Success Skills has three strong studies that support its effects on student achievement. Additional research is needed to provide strong evidence of the lasting nature of SSS's impact on student behavior and achievement. Hopefully, the positive findings of the existing research and the importance documenting the impact of school

counseling interventions on student achievement will encourage school counseling researchers to study this intervention and encourage grant funding agencies to support these efforts. Meanwhile, the Panel is continuing it work evaluating the research base of additional interventions and will continue to publish its analyses and suggestions for needed research.

References

Brigman, G., & Campbell, C. (2003). Helping students improve academic achievement and school success behavior. *Professional School Counseling, 7*, 91-98.

Brigman, G., & Webb, L. (2004). *Student Success Skills: Classroom Manual*. Boca Raton, FL: Atlantic Education Consultants.

Campbell, C.A. & Brigman, G. (2005). Closing the achievement gap: A structured approach to group counseling. *The Journal for Specialists in Group Work, 30,* 1-16.

Committee for Children. (1997a*). Second step: A violence prevention curriculum grades 1-3* (2nd ed.). Seattle, WA: Author.

Committee for Children. (1997b). *Second step: A violence prevention curriculum grades 4-5* (2nd ed.). Seattle, WA: Author.

Frey, K.S., Nolen, S.B., Van Schoiack Edstrom, L., & Hirschstein, M.K. (2005). Effects of a school-based social-emotional competence program: Linking children's goals, attributions, and behavior. *Applied Developmental Psychology, 26*, 171-200.

Grossman, D. C, Neckerman, H.J., Koepsell, T.D., Liu, P.Y., Asher, K.N., Beland, K., Frey, K., & Rivara, F.P. (1997). Effectiveness of a violence prevention curriculum among children in elementary school: A randomized controlled trial. *Journal of the American Medical Association, 277(20)*, 1605-1611.

McMahon, S.D., & Washburn, J. J. (2003). Violence prevention: An evaluation of program effects with urban African American students. *Journal of Primary Prevention, 24,* 43-62.

McMahon, S. D., Washburn, J., Felix, E.D., Yakin, J., & Childrey, G. (2000). Violence prevention: Program effects on urban preschool and kindergarten children. *Applied & Preventive Psychology, 9,* 271-281.

Orpinas, P., Parcel, G.S., McAlister, A., & Frankowski, R. (1995). Violence prevention in middle schools: A pilot evaluation. *Journal of Adolescent Health, 17*, 360-371.

Rosenberg, M.L., Powell, K.E., & Hammond, R. (1997). Applying science to violence prevention. *Journal of the American Medical Association, 277 (20),* 1641-1642.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin Company.

Task Force on Evidence-Based Interventions in School Psychology (2003). *Coding and Procedures Manual*. Madison, WI: Author.

Taub, J. (2002). Evaluation of the second step violence prevention program at a rural elementary school. *School Psychology Review, 31*. 186-200.

U.S. Department of Education (2005). Retrieved 6/21/05 from http://www.ed.gov/admins/lead/safety/exemplary01/report_pg6.html

U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration (2005). Retrieved 6/21/05 from http://mentalhealth.samhsa.gov/schoolviolence/Irenelis.asp

Van Schoiack-Edstrom, L., Frey, K. S., & Beland K. (2002). Changing adolescents' attitudes about relational and physical aggression: An early evaluation of a school-based intervention. *School Psychology Review, 31,* 201-216.

Webb, L.D., Brigman, G.A., & Campbell, C. (2005). Linking school counselors and student success: A replication of Student Success Skills approach targeting the academic and social competence of students. *Professional School Counseling, 8*, 407-413.

Welkowitz, J., Ewen, R., & Cohen, J. (1971). *Introductory statistics for the behavioral sciences.* New York, NY: Academic Press.

Whiston, S.C. and Sexton, T.L. (1998). A review of school counseling outcome research: Implications for Practice. *Journal of Counseling and Development, 76*, 412-426.

Valentine, J. C., & Cooper, H. (2003). *What Works Clearinghouse Study Design and Implementation Assessment Device* (Version 1.0). Washington, DC: U.S. Department of Education.

Table 1: Outcome Research Coding Protocol

Coding Studies and Rating the Level of Evidence for the Causal Effect of an Intervention

School counseling interventions will be evaluated by the Evidence-Based Practice Panel to determine the level of evidence that exists in outcome studies that supports the contention that the intervention causes a change in an important student outcome. The following seven domains will be used in this evaluation. Each domain has threshold criteria for two levels of strength: Strong Evidence and Promising Evidence. To be considered an Evidence-Based Practice, an intervention must exceed the Strong Evidence threshold in all seven areas. To be considered Promising Practice, an intervention must exceed Promising Evidence threshold in all seven areas.

Three Panel Members will independently review the outcome research related to a given intervention and independently rate each intervention on all seven criteria. Consensus in ratings will be achieved through consultation. The panel will disseminate its overall rating and, in cases where interventions fail to achieve Evidence-Based Practice or Promising Practice status, an analysis of deficiencies in the evidence base will be offered.

The seven domains and criteria are included below:

Domain **1. Measurement**

Principle: Important academic, career and/or personal/social outcomes are measured using reliable and valid instruments.

Strong Evidence:

1. Outcomes measures have established high reliability and validity characteristics.
2. Outcome measures are established to be appropriate for the population under study.

Promising Evidence:

Outcome measures have been used in previous studies,

1. Reliability characteristics are evaluated in the study and show adequate reliability.
2. Logical argument supports the appropriateness of the measures for the population under study.

Domain 2. Comparison Groups

Principle: Comparison groups with adequate controls are included so that resulting group differences can be attributed to the intervention.

Strong Evidence:

1. Active Comparison Groups (alternative treatment) with Adequate Controls (attention, placebo) are included in an outcome study.
2. Initial Group Equivalence is assured through random assignment.
3. Group Equivalence in Mortality/Attrition is established.

Promising Evidence:

1. Groups equated through matching or statistical procedures (e.g. ANCOVA) or strong pre-post-test designs are used with adequate controls.

Domain 3. Statistical Analyses of Outcome Variables

Principle: Statistical analysis documents low probability of Type 1 error and potency of intervention.

Strong Evidence:

1. Statistically significant finding using appropriate test.
2. Control for Experiment-wise error rate.
3. Adequate N.
4. At least a Moderate Effect size for critical outcome variables.

Promising Evidence:

1. Statistically significant finding using appropriate test.
2. Control for Experiment-wise error rate.
3. Adequate N.
4. At least a Small Effect size for critical outcome variables.

Domain 4. Implementation Fidelity

Principle: Intervention can be delivered with fidelity across contexts and is not contaminated by implementer.

Strong Evidence:

1. Intervention is extensively documented (manual or protocol) so that it can be reliably replicated.
2. Intervention is delivered by multiple people with adequate training and checks for adherence to protocol.

Promising Evidence:

1. Intervention is standardized and can be delivered across contexts.
2. Intervention is delivered by multiple people with adequate training.

Domain 5. Replication

Principle: The same intervention independently implemented with an equivalent population results in equivalent outcomes.

Strong Evidence:

1. Independent evaluators find equivalent outcomes with a similar population.

Promising Evidence:

1. Same evaluator finds equivalent outcomes with same population

Domain 6. Ecological Validity

Principle: The intervention can be implemented effectively in a public school with consistent effects across all student subgroups or with known differences between student subgroups. Limitations of the generalizability of results are clearly explicated.

Strong Evidence:

1. Study conducted in a diverse public school.
2. Outcomes are assessed across different subgroups of students or clearly specified as valid for a specific subgroup.

Promising Evidence:

1. Study conducted in a private, laboratory, or charter school or in a public school with limited diversity.

Domain 7. Persistence of Effect

Principle: The intervention results in a lasting effect on an important outcome measure.

Strong Evidence:

1. Treatment-Comparison group differences are demonstrated to persist for a practically significant time period.

Promising Evidence:

1. Treatment-Comparison Group Differences are demonstrated to persist beyond the immediate implementation.

Table 2: Quality of Evidence for Student Success Skills and Second Step

Protocol Domain	**Student Success Skills**	**Second Step**
1. Measurement	Strong Evidence	Strong Evidence
2. Comparison Groups	Promising Evidence	Strong Evidence
3. Statistical Analysis of Outcome Variables	Promising Evidence	Strong Evidence
4. Implementation Fidelity	Strong Evidence	Strong Evidence
5. Replication	Promising Evidence	Strong Evidence
6. Ecological Validity	Strong Evidence	Strong Evidence
7. Persistence of Effect	Weak Evidence	Strong Evidence