

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

ASCA National Conference  
Orlando, FL  
June 28, 2005

## Panel Members

- ▶ John Carey
  - University of Massachusetts, Amherst
- ▶ Carey Dimmitt,
  - University of Massachusetts, Amherst
- ▶ Trish Hatch
  - San Diego State University
- ▶ Rich Lapan
  - University of Missouri, Columbia
- ▶ Courtland Lee
  - University of Maryland, College Park
- ▶ Susan Whiston
  - University of Indiana, Bloomington

## Need

- ▶ 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.

## Meta-Analysis of School Counseling Interventions: Information on Studies

- ▶ 117 studies
  - 69.2% published
  - 30.8% thesis/dissertations
- ▶ Participants
  - 16,296 total
- ▶ Age
  - Elementary = 59 studies (50.4%)
  - Middle/Junior High = 21 studies (17.9%)
  - High School = 29 studies (24.8%)
  - Mixture = 7 studies (6%)
  - Unreported = 1 study (.9%)

# Interventions

- 153 school counseling interventions
- Types of interventions
  - Guidance Curriculum = 44 studies/57 interventions
  - Individual Planning = 9 studies/10 interventions
  - Responsive Services = 58 studies/74 interventions
  - Program-Wide Services = 8 studies/9 interventions
  - Other = 2 studies/3 interventions
- Length of interventions
  - Number of sessions  $M=12.44$ ,  $SD = 15.53$ , Range 1-120
  - Number of hours  $M=8.41$ ,  $SD = 6.87$ , Range .42-40
  - Duration  $M=71.76$ ,  $SD = 87.52$ , Range 1-720
- Providers of interventions
  - Experienced counselor = 37.9% (58 interventions)
  - Counselor in training = 9.8% (15 interventions)
  - Teacher = 8.5% (13 interventions)
  - Student peer = 3.3% (5 interventions)
  - Student independent = 5.9% (9 interventions)
  - Other = 5.9% (9 interventions)
  - Combination = 13.7% (21 interventions)
  - Unreported = 15% (23 interventions)

# Outcomes

- ▶ 785 outcomes
  - Cognitive = 12.2% (96 outcomes)
  - Behavioral = 34.8% (273 outcomes)
  - Affective = 39.0% (306 outcomes)
  - Effective Role Functioning = 9.4% (74 outcomes)
  - Satisfaction = .1% (1 outcome)
  - Global evaluation of school counseling program = .4% (3 outcomes)

## Effect Sizes

- Unweighted = .46
- Weighted = .29 (based on 153 interventions)
  - ▶ CI = .26 to .32
  - ▶  $Q = 616.124, p < .001$

## Grade Level and Type of Intervention

- Grade level effect sizes
  - ▶ Elementary ES = .25
  - ▶ Middle School ES = .39
  - ▶ High School ES = .34
- Type of intervention
  - ▶ Guidance Curriculum ES = .33 (44 studies/57 interventions)
  - ▶ Individual Planning ES = .27 (9 studies/10 interventions)
  - ▶ Responsive Services ES = .35 (58 studies/74 interventions)
  - ▶ Program Wide Evaluation = .19 (8 studies/9 interventions)

## Program component by Grade level

- **Guidance Curriculum**
  - Elementary ES = .29
  - Middle School ES = .41
  - High School ES = .39
- **Individual Planning**
  - Elementary ES = -.137 (1 intervention)
  - Middle School ES = 1.01 (3 interventions)
  - High School ES = .22 (6 interventions)
- **Responsive Services**
  - Elementary ES = .40 (45 interventions)
  - Middle School ES = .22 (10 interventions)
  - High School ES = .34 (20 interventions)

## Cognitive Outcomes

- ▶ **Overall ES = .19 (94 outcomes)**
  - ▶ GPA ES = .19 (49 outcomes)
  - ▶ Academic Achievement tests ES = .16 (16 outcomes)
  - ▶ Career knowledge ES = .61 (12 outcomes from 3 studies)
  - ▶ Other ES = .73 (19 outcomes from 11 studies)

## Behavioral Outcomes

- ▶ Overall ES = .39 (274 outcomes)
  - ▶ Behavior rating scales ES = .24 (123 outcomes)
  - ▶ No study examined change in enrollment (e.g., taking more difficult classes) as an outcome measure
  - ▶ Attendance ES = .30 (19 outcomes)
  - ▶ Interpersonal or social skills ES = .34 (55 outcomes)
  - ▶ Aggressive behavior/fights ES = .26 (12 outcomes)
  - ▶ Problem-solving skills ES = .96 (26 outcomes)
  - ▶ Peer counseling skills ES = 1.12 (15 outcomes)
  - ▶ Number of disciplinary referrals ES = .82 (9 outcomes)
  - ▶ Other ES = .25 (14 outcomes)

## Affective Outcomes

- ▶ Overall ES = .22 (306 outcomes)
  - ▶ Self-esteem ES = .18 (172 outcomes—22% of total outcomes)
  - ▶ Personal/social adjustment ES = .24 (27 outcomes)
  - ▶ Anxiety ES = .39 (35 outcomes)
  - ▶ Depression ES = .35 (6 outcomes)
  - ▶ Other ES = .24 (66 outcomes)

## Effective Role Functioning

- Overall ES = .12 (74 outcomes)
  - Career maturity/decidedness ES = .20 (30 outcomes)
  - Academic functioning ES = -.17 (9 outcomes)
  - Other ES = .12 (35 outcomes)

## Conclusion

- ▶ School counseling interventions moderately effective
- ▶ Individual planning at middle school level has a large effect size (small number of studies)
- ▶ Guidance curriculum more effective at middle and high school level; whereas, responsive services have a larger effect size at elementary level
- ▶ Large effects
  - Career knowledge
  - Problem solving
  - Disciplinary referrals
- ▶ Contributes to academic achievement
- ▶ Moderately effective in reducing anxiety and depression, increasing interpersonal social skills, and positively influencing attendance

## Panel Goals

- ▶ 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;
- ▶ 4) provide practicing professional school counselors and researchers guidance about measurement and research methodology.

## Outcome Research Protocol

- ▶ 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.

## Outcome Research Protocol

- ▶ Seven Domains of Quality of Evidence
  - Measurement
  - Comparison Groups
  - Statistical Analysis of Outcome Variables
  - Implementation Fidelity
  - Replication
  - Ecological Validity
  - Persistence of Effect

## Outcome Research Protocol

- ▶ Specific criteria in each domain related to:
  - Promising Evidence
  - Strong Evidence

## Outcome Research Protocol

- ▶ Three Panel Members independently review the outcome research related to a given intervention and independently rate each intervention on all seven criteria.
- ▶ Consensus in ratings is 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 with suggestions for research.

## Second Step: A Violence Prevention Curriculum

### Second Step

- ▶ Social/emotional learning program
- ▶ Developed by the Committee for Children
- ▶ PreK-8
- ▶ Curriculum and assessment materials
- ▶ Family component
- ▶ School-wide implementation plans

## Second Step Curriculum

- ▶ Identifying feelings
- ▶ Problem solving
- ▶ Social skills
- ▶ Building empathy

## Second Step Curriculum

- ▶ Anger reduction
- ▶ Stress management
- ▶ Resisting peer pressure
- ▶ Dealing with bullying
- ▶ Defusing potentially violent situations

## Second Step Curriculum

Lesson cards with:

- ▶ Lesson objectives
- ▶ Discussion questions
- ▶ Role plays and other activities
- ▶ Assessment materials

Videos

## Second Step Research

- ▶ Seven articles between 1995 and 2005 in juried journals
- ▶ Key article in 1997 by Grossman et al. in *Journal of American Medical Association*

## EBP Panel Review of Second Step Research

1. Measurement: All studies used measures which meet rigorous psychometric standards for reliability and validity.

Aggressive Behavior Scale

School Social Behavior Scale

Achenbach Teacher Report Form

Achenbach Child Behavior Checklist

Social Skills Rating System

Endorsement of Aggression Scale

## EBP Panel Review of Second Step Research

2. Comparison Groups: Research has used control groups within schools or used comparable schools as the unit of comparison.

Grossman et al. (1997) used randomized control trial design with schools as unit of randomization

Others used non-randomized control group comparison, randomly assigned control/intervention groups, or repeated measures designs

## EBP Panel Review of Second Step Research

3. Statistical Analysis of Outcome Variables:  
Research used statistical analyses that were thorough, sophisticated, and appropriate for assessing change in the outcome variables.  
ANOVA and MANOVA main and interaction effects, and repeated measure designs  
Generalized estimating equation regression method  
Intra-class correlations to establish inter-observer agreement  
Adequate sample sizes

## EBP Panel Review of Second Step Research

4. Implementation Fidelity: Curriculum was taught by teachers, school counselors or other school personnel who had completed 4-16 hours of training with Committee for Children trainers.  
Intervention occurred over the course of several months  
Intervention was usually presented in 30 minute lesson plans 1-2 times/week  
Treatment fidelity was monitored through the use of logs, support teams, administrator supervision, and /or self-report

## EBP Panel Review of Second Step Research

5. Replication: Consistent evidence that the curriculum increases social skills and prosocial behavior, and decreases anti-social and/or aggressive behaviors. Replication of self-reported attitudes, knowledge, and skills.

## EBP Panel Review of Second Step Research

6. Ecological Validity: All research in public schools, and findings successfully replicated across contexts with:
  - Different ages (PreK-8)
  - A range of racial/ethnic groups
  - Males and Females
  - Economically and socially diverse populations

## EBP Panel Review of Second Step Research

7. Persistence of Effect: Variable findings for persistence of effects at 3 months, 6 months, and one year.

Grossman et al. (1997) found that behavioral effects were still strong at 6 months.

## Second Step EBP Protocol Summary

<b>Protocol Domain</b>	<b>Second Step</b>
1. Measurement	Strong Evidence
2. Comparison Groups	Strong Evidence
3. Statistical Analysis of Outcome Variables	Strong Evidence

## Second Step EBP Protocol Summary

<b>Protocol Domain</b>	<b>Second Step</b>
4. Implementation Fidelity	Strong Evidence
5. Replication	Strong Evidence
6. Ecological Validity	Strong Evidence
7. Persistence of Effect	Strong Evidence

## Student Success Skills Evaluation

- ▶ 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).

## 1. Measurement

- ▶ All three studies used the **Florida Comprehensive Achievement Test (FCAT)** as the primary outcome measure.
  - criteria referenced state achievement test
  - meets standards of psychometric rigor
- ▶ All three studies collected teacher ratings of students using the **School Social Behavior Scale (SSBS)**
  - meets rigorous psychometric standards for reliability and validity.

## 1. Measurement – Strong Evidence

- ▶ The Panel judged in the Measurement domain, SSS research reflects **Strong Evidence** based on the FCAT.
- ▶ FCAT scores were used as pre and post-test measures of achievement for both the experimental and control groups.

## 1. Measurement – Limitation of SSBS

- ▶ Only those in Treatment group were rated in pre-post on the SSBS
- ▶ In spite of the psychometric properties of the SSBS, the use of this instrument added little to the:
  - understanding of the mechanisms by which SSS may have impacted academic achievement or
  - confidence in the impact of SSS.

## 1. Measurement - Linkage Missing

- ▶ Convincing linkage between process and results data was missing
- ▶ 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.
- ▶ Logical links between the nature of the SSS intervention and the constructs measured by the SSBS are *neither obvious nor explicitly linked* to impact on student achievement data, specifically their performance on the FCAT.

## 1. Measurement -Recommendations

- ▶ Future studies of SSS should include measures that reflect the ***specific constructs targeted*** by the SSS interventions (e.g. cognitive/meta-cognitive skills, social skills and self-management skills)
- ▶ Relationships between ***changes on these variables and increases in academic achievement must*** be ascertained.
- ▶ We also recommend that adult raters should be “**blind to treatment**” condition.

## 2. Comparison Groups

- ▶ **All three SSS studies** compared a Treatment Group to an untreated Control Group.
- ▶ 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.

## 2. Comparison Groups - Promising

- ▶ Difficult to remove **suspicion of the impact of attention or placebo** on expectations.
- ▶ Follow-up studies with placebo control groups are needed to **ensure that the effects of the intervention are related to the learning that takes place** in the SSS process rather than to the additional attention students may have received or to the expectations teachers and/or students may have had that these students would improve with intervention.

## 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.
- ▶ Two studies significant effects reported on both FCAT **reading and math** (Brigman and Campbell, 2003; Campbell and Brigman, 2005).
- ▶ Third study found a significant **effect for math but *not* reading** (Webb, Brigman, and Campbell, 2005).
- ▶ Effect sizes for the impact of SSS on FCAT scores *were not reported*
- ▶ Cohen's d effect size statistic was used by Panel to estimate effect sizes from the post-test FCAT data.

### 3. Statistical Analysis of Outcome Variables - Promising

- ▶ Based on the effect size, The Panel concluded that SSS research to date reflect **Promising but not Strong evidence of effectiveness** [ between .176 and .216 for the FCAT Reading Test; and between .142 and .154 for the FCAT Math Test.]
- ▶ 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.]

### 4. Implementation Fidelity - Strong

- ▶ 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.

## 5. Replication - Strong

- ▶ 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**.
- ▶ 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. Independent replication is needed.

## 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 has scored average or below average on the previous year's FCAT (25th-50th percentile or 25th-60th percentile).
- ▶ The relatively small numbers non-White students made it impossible to determine whether SSS is more effective with some groups of students.

## 6. Ecological Validity - Strong

- ▶ The research reports 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 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*.

## 7. Persistence of Effect - Weak

- ▶ None of the studies investigated persistence of effects of SSS on academic achievement beyond the year in which intervention occurred.
- ▶ Research reports are **unclear** about the time period between the last SSS session and FCAT testing.
  - In the Brigman and Cambell (2003) study the 8-weekly-session group intervention was completed at the beginning of December but 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 were followed by with booster sessions in January through April.
- ▶ No study reported FCAT test date but highly likely it was very short time period between the last booster session and the test.
- ▶ Given this short time period, the Panel concludes that the **persistence of effect for SSS is not yet available** and that therefore it qualifies as **Weak Evidence**.

## Summary

- ▶ SSS is developing a strong research base and has some very promising findings.
- ▶ **SSS Strong Evidence** in the domains of:
  - Measurement
  - Implementation Fidelity
  - Ecological Validity
- ▶ **SSS Promising Evidence** in:
  - Comparison Groups
  - Replication
  - Statistical Analysis of Outcome Variables
- ▶ **SSS Weak Evidence:**
  - Persistence of Effects

## Summary

- ▶ ***The Panel strongly*** recommends:
  - additional research targeted at determining whether the beneficial effects of SSS persist.
  - 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).

## Summary – Recommendations

- ▶ **Large and diverse N studies** will be needed if the only outcome variable is academic achievement measured by standardized test scores.
- ▶ 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.
- ▶ Finally, the Panel recommends additional studies that employ “**placebo**” controls and studies that investigate how effective SSS is with different student subgroups

## SSS EBP Protocol Summary

### Protocol Domain

1. Measurement
2. Comparison Groups
3. Statistical Analysis of Outcome Variables

### Second Step

- Strong Evidence  
Promising Evidence  
Strong Evidence

## SSS EBP Protocol Summary

<b>Protocol Domain</b>	<b>Second Step</b>
4. Implementation Fidelity	Strong Evidence
5. Replication	Promising Evidence
6. Ecological Validity	Strong Evidence
7. Persistence of Effect	Weak Evidence

## Issues

- ▶ Give credit for hard work being done on these two developmental programs
- ▶ Not doing nearly enough research like this
- ▶ Caveats for Second Step - every program can get better
  - Not a quick fix or a panacea - "modest positive effects"
  - Absence of comparison treatments
  - Schools need to make a major buy in - not an "a la carte" experience
  - Many outcomes not studied at the level we would like and inferences based on prior research used to make links to valued outcomes

## Issues

- ▶ **Common factors versus specific treatments**
- ▶ **Paying more attention to construct validity issues in our research**
- ▶ **Helping schools to really define what the problem is, so that we can assess whether or not we are making progress towards solving the problem**
- ▶ **Great opportunity for us - form partnerships and engage in the process!**

## Construct Validity - closely match the operations we are using to the constructs we think we are aiming at

The "SIMS" - every school counseling intervention has to deal with the SIMS

- ▶ Students - Are the students really representative of the construct we are after (e.g., focusing on at-risk students)
- ▶ Interventions - Are the interventions closely tied to the construct we are evaluating? (e.g., behavioral therapists used more empathy statements than Rogerian counselors)
- ▶ Measurements - Are the observations/measurements we are making confounded by other constructs? (e.g., when studying the relationship between academic achievement and race, are we really measuring the effects of socioeconomic status on academic achievement?)
- ▶ Settings - In what contexts does this intervention take place? (e.g., in Columbia, 3 elementary schools - 1 urban, 1 suburban, 1 rural. They are very different places even though the state designates all of our town with 1 value)

Construct Validity - closely match the operations we are using to the constructs we think we are aiming at

- ▶ More attention has to be given to how our interventions are delivered, received, and then adhered to.
- ▶ Have to make sure that what we have intended to implement is what is really going on.
- ▶ Use Qualitative research methods to study this!