Journal of Clinical Child & Adolescent Psychology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hcap20

Novel Programs, International Adoptions, or Contextual Adaptations? Meta-Analytical Results From German and Swedish Intervention Research

Knut Sundell[a], Andreas Beelmann[b], Henna Hasson[cd] & Ulrica von Thiele Schwarz[c]

[a] Department of Knowledge Based Policy and Guidance, National Board of Health and Welfare

[b] Department of Research Synthesis, Intervention, and Evaluation, Institute of Psychology, Friedrich-Schiller-University

[c] Medical Management Centre, Department of Learning, Informatics, Management, & Ethics, Karolinska Institutet

[d] Center for Epidemiology and Community Medicine, Stockholm County Council

Published online: 11 Apr 2015.

To cite this article: Knut Sundell, Andreas Beelmann, Henna Hasson & Ulrica von Thiele Schwarz (2015): Novel Programs, International Adoptions, or Contextual Adaptations? Meta-Analytical Results From German and Swedish Intervention Research, Journal of Clinical Child & Adolescent Psychology, DOI: 10.1080/15374416.2015.1020540

To link to this article: http://dx.doi.org/10.1080/15374416.2015.1020540

Journal of Clinical Child & Adolescent Psychology, 0(0), 1–13, 2015

ISSN: 1537-4416 print/1537-4424 online
DOI: 10.1080/15374416.2015.1020540

Routledge
Taylor & Francis Group

Novel Programs, International Adoptions, or Contextual Adaptations? Meta-Analytical Results From German and Swedish Intervention Research

Knut Sundell

Department of Knowledge Based Policy and Guidance, National Board of Health and Welfare

Andreas Beelmann

Department of Research Synthesis, Intervention, and Evaluation, Institute of Psychology, Friedrich-Schiller-University

Henna Hasson

Medical Management Centre, Department of Learning, Informatics, Management, & Ethics, Karolinska Institutet and Center for Epidemiology and Community Medicine, Stockholm County Council

Ulrica von Thiele Schwarz

Medical Management Centre, Department of Learning, Informatics, Management, & Ethics, Karolinska Institutet

One of the major dilemmas in intervention and implementation research is adaptation versus adherence. High fidelity to an intervention protocol is essential for internal validity. At the same time, it has been argued that adaptation is necessary for improving the adoption and use of interventions by, for example, improving the match between an intervention and its cultural context, thus improving external validity. This study explores the origins of intervention programs (i.e., novel programs, programs adopted from other contexts with or without adaptation) in two meta-analytic intervention data sets from two European countries and compares the effect sizes of the outcomes of the interventions evaluated. Results are based on two samples of studies evaluating German child and youth preventative interventions ($k = 158$), and Swedish evaluations of a variety of psychological and social interventions ($k = 139$). The studies were categorized as novel programs, international adoption and contextual adaptation, with a total of six subcategories. In the German sample, after statistically controlling for some crucial methodological aspects, novel programs were significantly more effective than adopted programs. In the Swedish sample, a trend was found suggesting that adopted programs were less effective than adapted and novel programs. If these results are generalizable and unbiased, they favor novel and adapted programs over adopted programs with no adaptation and indicate that adoption of transported programs should not be done without considering adaptation.

Correspondence should be addressed to Andreas Beelmann, Department of Research Synthesis, Intervention, and Evaluation, Institute of Psychology, Friedrich-Schiller-University, Jena, Germany. E-mail: andreas.beelmann@uni-jena.de

The international spread of evidence-based programs has resulted in an increase in controlled trials of psychological and social interventions around the world. Some of these trials are true replications that test a fully adopted intervention that is transported from one social context to another. Other studies evaluate adapted programs, which have been changed in one or more aspects for cultural or other reasons. Results from these trials are mixed and include both successful and unsuccessful replications of previous results (e.g., Sundell, Ferrer-Wreder, & Fraser, 2014). One explanation for the unsuccessful replications that has been put forward is that the intervention was culturally or in other ways adapted, thus compromising the desired benefits of the intervention by eroding treatment integrity (Sundell et al., 2014). However, as adaptations of empirically supported interventions are poorly investigated, there is a lack of data to support this claim. In fact, the opposite may likewise apply. This would suggest that unsuccessful replications result from a lack of adaptation to the new context, thus failing to create a fit between the intervention and the target population or situational and cultural context (Kilbourne, Neumann, Pincus, Bauer, & Stall, 2007). In addition, a program developer from context A has the opportunity not only to adopt or adapt programs from context B (e.g., using a program designed and evaluated in the United States in a European country) but also to develop novel programs that are assumed to be even better adjusted to the specific context of application than an imported program.

Up to now, little has been known about the most promising way toward effective program construction. This study reports the extent to which intervention research in two European countries (Germany and Sweden) involved adopted, adapted, and novel programs and compares the effectiveness of these different types of intervention programs.

ADOPTION OF INTERVENTIONS

Adoption of empirically supported programs deals with programs being replicated in a new context with an ambition to achieve high fidelity with the original program. This has long been the foundation of the research-to-practice pathway. First, the efficacy of a treatment is established using research designs allowing for a high degree of control over exogenous variables, primarily using a randomized controlled trial. The next step is to replicate that exact treatment in real-world settings, thereby testing the effectiveness of the treatment. Thereafter, the treatment can be made available to the public through the adoption of the treatment by organizations that will deliver them to their clients. This view implies that any deviations from the original treatment might threaten the integrity of the intervention, and claims of efficacy of the intervention may no longer be valid. Several empirical studies showing that high fidelity is related to better client outcomes support this view (e.g., Becker, Smith, Tanzman, Drake, & Tremblay, 2001; Bellg et al., 2004; Dane & Schneider, 1998; Keith, Hopp, Subramanian, Wiitala, & Lowery, 2010).

ADAPTATIONS OF INTERVENTIONS

The translation of interventions from research to clinical settings is not straightforward, as indicated by the substantial gap between the efficacy of interventions provided in research settings and the effectiveness of interventions delivered in clinical service settings (e.g., Curtis, Ronan, & Borduin, 2004; Emshoff, 2008; Petrosino & Soydan, 2005). Efficacy may drop as a result of the added complexity that comes with heterogeneity among patients, providers, and settings as interventions move from the research setting to the service delivery setting (Chambers, Glasgow, & Stange, 2013; Gottfredson et al., 2006; Weisz, Donenberg, Han, & Weiss, 1995). Of importance, adaptations in the delivery or content of interventions in empirical investigations have been the rule rather than the exception when interventions are spread to clinical settings. Several studies have reported that a large proportion of program users (e.g., 44–88%) adapt evidence-based programs (Hill, Maucione, & Hood, 2007; Kaunitz & Strandberg, 2009; Moore, Bumbarger, & Cooper 2013; Ringwalt, Vincus, Ennett, Johnson, & Rohrbach, 2004). In line with this, it has been argued that in complex social settings, replication of a previously tested program is neither possible nor necessarily warranted (Chambers et al., 2013). Bauman, Stein, and Ireys's (1991) "principle of program uniqueness" stresses that most programs are both developed and tested under more or less unique circumstances (e.g., funding, charismatic leader, and tested in settings unlike that in which the program will later be implemented). This makes attempts of perfect replications unlikely. From this follows suggestions that replications should be considered as reinventions, with focus on how and what to change so that the basic integrity of the program model is maintained. This basic integrity, which has been termed the intervention's core components (Bauman et al., 1991), theory of change, program theory, internal logic (McKleroy et al., 2006) or deep structure (Resnicow, Soler, Braithwaite, Ahluwalia, & Butler, 2000) is according to this line of thinking, where fidelity attempts should focus leaving other intervention aspects open for adaptation and adjustment. This assumes that the core components are known, which seldom is the case (Elliott & Mihalic, 2004).

Another argument for adaptation is to create ownership for and involvement in the program in the local setting as well as creating alignment with local needs (Arthur & Blitz, 2000; Backer, 2000; Durlak & DuPre, 2008). This may be necessary to make the program appear relevant in relation to the local population and services and promote uptake (Lee, Altschul, & Mowbray, 2008). Adaptation has also been argued to support sustainability of the program (Backer, 2000; Durlak & DuPre, 2008), although it has been noted that there is little point in sustaining an intervention if it isn't shown to be effective in the first place (Elliot & Mihalic, 2004).

TYPES OF ADAPTATION

One of the challenges in adaptation research is the lack of consensus on how to operationalize adaptation (Stirman, Miller, Toder, & Calloway, 2013). Although there is a general agreement that adaptations entail changes from the original program, there is a lack of agreement as to whether the term adaptation should be confined to referring to planned and purposeful program changes or if the term also refers to unintentional deviations, also referred to as drift.

Some descriptions of intervention modification and adaptation have been published (e.g., Hill et al., 2007; Moore et al., 2013; Stirman et al., 2013), but there are few efforts to systematically categorize them. For example, in a literature review of evidence-based programs, Stirman et al. (2013) reported that 25% of the 32 published studies had modified the targeted population, 25% the format, and 19% the setting. The content of the program was also altered: 34% tailored, tweaked, or refined elements; 29% added elements; and 15% removed elements. In a study investigating adaptations of evidence-based programs in the prevention of violence and delinquency in Pennsylvania, 43% of those reporting adaptations changed the procedures, 42% changed the dosage, 38% changed the content, 22% made cultural adaptations, and 12% made adaptations to the target populations (Moore et al., 2013). Hill et al. (2007) identified similar types of modifications made to a single evidence-based prevention program in a statewide implementation effort (i.e., modifications of content, dosage, and procedure). Thus, adaptation may involve anything from small changes in terminology or language to larger changes such as removal of program components or integration with other interventions. For the sake of the current study, adaptation refers to planned or purposeful changes to the design or delivery of an intervention.

EFFECTS OF ADAPTATION

The impact of adaptation on treatment outcomes has been studied primarily in the United States and discussed within the cultural adaptation literature. The primary focus has been concerns regarding how to better account for the ethnical and cultural context and values of minority groups. The evidence produced by the body of research on cultural tailoring for ethnic minority groups is mixed. Results from meta-analyses and literature reviews of cultural adaptations from various fields including psychotherapy research, substance abuse, and family interventions ranges from finding no added effects of ethnic responsive elements (Huey & Polo, 2008) or increased retention but reduced positive outcomes (Kumpfer, Alvarado, Smith, & Bellamy, 2002) to studies indicating moderate to strong benefit of culturally adapted interventions (Benish, Quintana, & Wampold, 2011; Griner & Smith, 2006). In addition, several reviews and meta-analyses concluded that the evidence is somewhere in between small effects (Hodge, Jackson, & Vaughn, 2010) and promising but inconclusive effects (Hodge, Jackson, & Vaughn, 2012; Jackson & Hodge, 2010).

There are studies that have assessed the effects of various types of adaptation such as altered setting, procedure, population, and format. For instance, research on the transportation of face-to-face cognitive behavior therapy (CBT) for psychiatric and somatic disorders to Internet-delivered CBT, indicates that the two formats for delivery produced equivalent overall effects (Andersson, Cuijpers, Carlbring, Riper, & Hedman, 2014). Parent management training has also been shown to be equally effective regardless of whether it is delivered by nonspecialist or specialist therapists, directed at a clinic-referred or self-referred population, or provided as part of routine care (Michelson, Davenport, Dretzke, Barlow, & Day, 2013). Similarly, cognitive processing therapy for posttraumatic stress disorder (Galovski, Blain, Mott, Elwood, & Houle, 2012) modified to allow for flexibility in delivery (determination of termination after patient progress rather than fixed number of sessions, possibility to add stress management sessions, use of less experienced therapists, and less restricted inclusion criteria) was found to produce effect sizes similar to those found in previous research and suggests that the adaptations made to the cognitive processing therapy protocol did not diminish the efficacy of the intervention. Stanton et al. (2005) tested an adapted and nonadapted version of a sexual risk reduction intervention designed for urban youths on rural youths. Overall, neither version was as effective when used with rural youths as the original intervention used in the urban setting. In other words, the less altered version was more effective. This was especially true in regard

to youth perceptions of condom use. There were, however, no significant behavioral effects.

There are also a few studies investigating the effect of adaptation when transferring interventions from one nation to another. One example is Strengthening Families Program, which was adapted to better match the values and culture in Sweden. The adaptation involved modification of the setting, dosage, group format, and content (Skärstrand, Larsson, & Andreasson, 2008) of the original program. In a cluster randomized trial with multiple assessments, the adapted Swedish Strengthening Families Program showed no intervention related benefits compared to a treatment as usual control group (Skärstrand, Sundell, & Andreasson, 2013).

NOVEL INTERVENTIONS

Mixed outcomes for adopted and adapted interventions and several possible sources of program mismatch to the routine settings (Castro, Barrera, & Martinez, 2004) are reasons why scholars decide to construct novel or innovative programs. Programs constructed for a specific social, contextual, and cultural setting are assumed to provide a better fit by being culturally grounded and community resonant (Gone, 2007). In addition, scientific progress and new scientific results on, for example, risk and protective factors, should lead to novel and innovative intervention programs.

However, the terms ''innovation'' and ''new'' are not easy to define in social science (Rogers, 2003) because of several distinct features including the cumulative nature of scientific progress. According to social and clinical interventions, a pragmatic definition could be the reference to a new label or brand name. However, within this approach, we have to differentiate between an absolutely innovative program that relies on a new intervention theory and a conceptually new program that refers to a known program type (e.g., social skills training) but with a new kind of application or changes in intervention material, dosage, or target group. Notwithstanding how we define innovative or new programs, there is essentially no empirical data on whether such programs are more effective than adopting or adapting international interventions that are already evidence based.

AIMS

The primary goals of this study were to (a) explore the origins of intervention programs (i.e., novel programs, programs adopted from other contexts with adaptation, and programs adopted from other contexts without adaptation) in two different meta-analytic data sets from two European countries (Germany and Sweden), and (b) to compare the effectiveness of these different types of program constructions.

METHOD

German Study Sample

The German sample data came from a meta-analysis of studies in German-speaking countries on prevention and health promotion programs for children and adolescents (Beelmann, Pfost, & Schmitt, 2014). This review summarized results of published and unpublished studies that (a) evaluated a psychological or educational prevention or health promotion measure and were conducted in a German-speaking area (Germany, Austria, Switzerland) up to 2010; (b) included a randomized or nonrandomized control group with a pre–post or pre–follow-up design; (c) had a target group up to 18 years of age; and (d) presented empirical data that allowed a reliable effect size calculation. We excluded only prevention programs for the medical field (e.g., caries prophylaxis) and cognitive programs (e.g., promotion of inductive thinking).

Comprehensive literature searches in relevant databases (e.g., psyndex), journals, existing reviews, and reference lists of already selected reports revealed 146 research reports with 173 comparisons between an intervention and a control group. These studies and comparisons were coded according to a detailed manual that included characteristics of the report (e.g., publication year and type), intervention (e.g., type, intensity, adviser), field of intervention (e.g., health promotion, drug prevention, prevention of antisocial behavior, etc.), characteristics of the target group (e.g., age, gender, type of prevention), and results (effect sizes, types of outcome measurement, etc.).

For the current analysis, we selected only those studies that evaluated an identified intervention or program with a description sufficient to allow reliable coding according to the type of construction (e.g., novel programs, adoptions, adaptations). We further restricted our analysis to posttest data (i.e., outcome measurements up to three months after the termination of the intervention), whereas all follow-up assessments (3 months or later) were excluded. This resulted in a final study set of 134 reports published between 1969 and 2010 with 158 intervention–control comparisons (two studies had to be excluded because an open intervention concept with no further description was applied). Most of the studies were conducted in Germany (87.3%) and were published in scientific journals (68.7%). Comparisons mainly evaluated universal prevention measures (79.7%) and were implemented primarily in school

(68.4%) and kindergarten/preschool settings (13.9%) using a broad variety of different programs in various fields of prevention and health promotion. A total of 1.276 effect sizes were calculated, measuring different constructs like behavioral problems, prosocial behavior, social-cognitive abilities, and health knowledge and attitudes (for a detailed description, see Beelmann et al., 2014). For each comparison, all single effect sizes (ranging 1–34 within comparisons) were averaged within comparisons, first according to the measurement construct (e.g., behavioral problems, social-cognitive abilities) and second across measurement constructs to calculate the mean comparison effect size. Integration of effect sizes across studies was done using methods proposed by Hedges and Olkin (1985).

For more information on the German study sample, see Table 1 and Appendix S1 (supplementary material with flowcharts on study selection and references of all integrated studies).

Swedish Study Sample

The Swedish study sample consists of studies that (a) were published in peer-reviewed journals between 1990 and April 2012; (b) evaluated a psychological (treatments that use verbal or nonverbal communication) or social (social or material support) intervention in single or in combination with medical or pharmacological treatments; (c) included a randomized or nonrandomized control group with a pre–post or pre–follow-up design; (d) that targeted individuals in order to prevent or rehabilitate from physical, psychological, or social problems; (e) with the primary investigator affiliated with a Swedish university, college, or other organization; and (f) presented information that allowed the calculation of effect sizes according to Lipsey and Wilson (2000). Studies were not included if they investigated the effects of training professionals or organizational changes, unless they were related to outcome measures on the target group (i.e., clients, etc.).

TABLE 1
Description of the Studies Included in the Two Samples

Study Characteristics	*German Sample*[a]	*Swedish Sample*[b]
Randomized Trial (%) vs. Nonrandomized	34	88.5
Efficacy Trial (%) vs. Effectiveness	Not coded	60.4
Universal Prevention (%) vs. Selected and Indicated	80	12.2
Passive Control (e.g., Waitlist) (%) vs. Active	90	29.5
Time Between Pre- and Follow-Up Measure (Months)	0.3 (0.6)	17.7 (18.0)
1–3 Months (%)	100	14.4
4 Months (%)	0	85.6
Total Sample for Analyses (*M*, *SD*)	313.3 (410.6)	186.9 (552.8)
16–50 (%)	16.5	22.3
51– (%)	83.5	77.7
Attrition (in %)	16.2 (13.9)	10.0 (14.9)
0%	9.6	47.5
1–10%	31.7	18.7
11-%	34.1	33.8
Information Missing %	34.2	0

[a] $n = 158$.
[b] $n = 139$.

All Swedish universities and colleges were screened for reports, articles, or unpublished theses that used the following terms in Swedish and English: randomized, controlled, effect, evaluation, outcome study, implementation, randomized controlled trial, and evidence. Furthermore, the main Swedish research councils (The Swedish Research Council, The Swedish Research Council for Health, Working Life and Welfare, The Swedish Research Council Formas, and Vinnova) were scanned, as were the European Research Council (during 2005–2012), and the Swedish Crime Victim Compensation and Support Authority (during 2003–2011). A number of other web sites were scanned, both Swedish (e.g., www.researchweb.org, www.diva-portal.org/smash/search.jsf) and international (e.g., www.clinicaltrials.gov, www.ncbi.nlm.nih.gov/pubmed). Studies were also identified using a snowball strategy, in which we searched the reference sections of identified articles and personally contacted the primary investigators of already identified articles (see also Sundell, 2012).

For this study, only articles published in peer-reviewed journals were included, a total of 145 trials. In addition, 59 technical reports and dissertations were identified but not included in the analyses. Of the 145 peer-reviewed articles, 139 included enough information to retrieve effect sizes. The types of interventions included prevention and treatment of mental health problems ($k = 63$), substance abuse ($k = 33$), child and youth conduct disorders ($k = 20$), behavioral medicine ($k = 15$), and care of the elderly ($k = 8$). Settings in which the interventions were delivered included mental health clinics ($k = 58$), substance abuse treatment ($k = 25$), hospitals and medical clinics ($k = 21$), social welfare agencies ($k = 16$), schools ($k = 9$), elderly care ($k = 8$), employment settings ($k = 1$), and correctional facilities ($k = 1$).

Information was retrieved on the primary outcome measure from the 139 peer-reviewed articles. If primary outcome measure was not specified, or if there were more measures, the first one mentioned in the method section was chosen. If more than one follow-up measurement was reported, the one with longest time was coded. If the study included more than one control

group, the one with the least intervening control (e.g., wait-list) was chosen.

For more information on the Swedish study sample, see Table 1 and Appendix S2 (supplementary material with flowcharts on study selection and references of all integrated studies).

Coding for Program Type

Intervention studies from both data sets were categorized according to a predefined coding scheme for the type of program, with three broadband categories (novel programs, international adoption, and adaptation) and six subcategories (innovation, conceptually new, adoption, cultural adaptation, pragmatic adaptation, and eclectic adaptation). The definitions for each type and subcategory are provided in Table 2. The focus of the coding was on planned adaptations, as opposed to unplanned (i.e., poor program adherence that evolves over time) adaptations. Adopted program refers to those programs that originate in another country and that according to the article have not been altered, modified, or adapted. Programs that are coded as adapted were identified in the original study as having gone through modification (cf. Moore et al., 2013; Stirman et al., 2013) in procedure (e.g., a modified recruitment process), dosage (e.g., that more time was devoted to homework assignments and daily practice between sessions), content (e.g., a shorter exercise book and more emphases on youth-related issues), setting (e.g., a treatment designed to be used in a mental health clinic setting that is now delivered in primary care), format (e.g., a treatment designed to be used one-on-one that is now delivered in a group format), target population (e.g., modifications to adjust for child age and developmental level) or combination of these characteristics. A program is coded (a) as culturally adapted if the modification is explicitly motivated for with cultural reasons (e.g., adjustments to the content of some items to be consistent with Swedish cultural norms), (b) as pragmatic if it is motivated for with practical reasons as time or cost restraints, and (c) eclectic if two or more international programs are mixed. Novel programs are those programs that lack international references. However, they may include replications of broadband intervention approaches (e.g., CBT). Novel programs are coded as innovations when there is no reference to an already existing international program or program type. A conceptually new program is developed completely or almost completely new for the social context where it is applied but with reference to a specific international program type (e.g., a new life skill program but with reference to international life skill programs).

Two researchers reviewed the full articles to determine their eligibility for this study and when necessary discussed differences of opinion regarding inclusion with the rest of the study team. All studies of the two data sets were coded by a trained coder according to the coding scheme (Table 2), followed by a second independent coding by one of the authors (AB, German sample; KS, Swedish sample). Intercoder agreement for the ratings of the six subcategories reached 88.0% in the German sample and 85.6% in the Swedish sample. When the three broad categories (novel programs, adoptions, and adaptations) were used, agreement was 96.2% in the German sample and 92.8% in the Swedish sample. All individual study codings according to the program types are listed in Appendix S1 (German studies) and S2 (Swedish studies).

TABLE 2
The Coding Scheme for the Type of Program Construction With Definitions for Each Subcategory

Program Type	*Subcategory*	*Definition*
Novel	Innovations	The program is developed completely or almost completely new for the social context (i.e., country) where it is applied without any reference to an already existing international program or program type (e.g., social skills training).
	Conceptually new	Program is developed completely or almost completely new for the social context (i.e., country) where it is applied but with reference to a specific international program type (e.g., a new life skill program with reference to international life skill programs). Programs referred only to broadband intervention approaches (e.g., CBT) were coded as innovations.
Adopted		Program is completely or almost completely adopted from an international empirically supported program (i.e., direct translation of a manual with only marginal changes in content, delivery, intensity, and material or target group).
Adapted	Cultural adaptation	An international empirically supported program or treatment is modified for cultural reasons (e.g., dropping or adding material or content, etc.).
	Pragmatic adaptation	An international empirically supported program or treatment is modified for practical reasons (e.g., dropping or adding sessions for time reasons; availability of materials, etc.).
	Eclectic adaptation	Adoption or adaptation of parts of international programs mixed or combined with recently developed parts.

Note: CBT = cognitive behavioral therapy.

RESULTS

The most frequent type of program construction was novel (73.4%, $k = 116$ in the German sample; 69.1%, $k = 96$ in the Swedish sample). Of these, 62.0 and 69.8%, respectively, were innovations and the rest were conceptually new programs. In the German study sample, 10.8% ($k = 17$) were adoptions of international programs, and 15.8% ($k = 25$) were some kind of adaptation (Table 3). In the Swedish sample, 23.0% ($k = 32$) were adoptions and 7.9% ($k = 11$) involved some kind of adaptation (Table 4). The effect sizes for different types of program construction are presented in Table 3 (German study sample) and Table 4 (Swedish study sample). The analyses were conducted in the same way for both samples.

German Study Sample

The comparison effect sizes within the total study sample (Table 3, sample A) range from $d = -0.20$ to 1.98 with an unweighted mean of $d = 0.32$. According to Hedges and Olkin's (1985) model, the mean weighted effect size $d+$ was 0.15 according to the fixed effect model, and 0.24 according to the random effect model (REM). However, as the sample showed significant heterogeneity with a large proportion of random variance, $Q(157) = 622.15$, $p < .001$, $I^2 = 74.8\%$, all further calculations were made using the REM.

Table 3 shows all analyses according to the type of program origin. For the total study sample (Table 3, Sample A), the highest mean effect size was shown for the subcategories innovations ($d+ = 0.31$), cultural adaptations ($d+ = 0.25$), and conceptually new programs ($d+ = 0.20$). Adoptions ($d+ = 0.16$) and pragmatic adaptations as well as eclectic adaptation ($d+ = 0.06$, 0.14) showed lower effect sizes. A comparison between the three main types of program construction (novel program, adoption, and adaptation) confirm a statistical trend, $Q(2) = 5.22$, $p < .08$. However, comparison of the four most frequent subcategories—innovation, conceptually new, cultural adaptations, and adoptions—are significant, $Q(3) = 8.82$, $p < .03$, favoring innovations. A significant difference was also found when comparing all novel programs with all adoptions and adaptations, $Q(1) = 4.18$, $p < .05$.

Because study design and sample size have been shown to have a significant effect on outcomes (Beelmann, 2014; Beelmann et al., 2014), we reconducted the calculations, first excluding comparisons with low design ratings (i.e., designs with nonequivalent control conditions; Table 3, Study Sample B; $k = 141$), second with sample size lower than 50 (Table 3, Study Sample C, $k = 132$), and third with both low design ratings and sample sizes (Table 3, Study Sample D, $k = 119$). However, despite excluding the studies that were most problematic from a methodological viewpoint, results were quite stable across different kinds of study samples, although the mean effect

TABLE 3
Results for the German Study Sample

	A			*B*		*C*		*D*	
	Total Study Sample[a]			*Study Sample Selected by Design*[b,c]		*Study Sample Selected by Sample Size*[d,e]		*Study Sample Selected by Design and Sample Size*[b,d,f]	
Type of Program Construction	*d*	*d+*	*K*	*d+*	*k*	*d+*	*k*	*d+*	*k*
Novel Programs	.36	.27 ± .05	116	.24 ± .05	102	.24 ± .05	98	.22 ± .04	88
Innovations	.40	.31 ± .06	72	.28 ± .07	60	.28 ± .06	62	.25 ± .06	53
Conceptually New	.28	.20 ± .08	44	.19 ± .08	42	.17 ± .07	36	.17 ± .07	35
Adoptions	.23	.16 ± .13	17	.16 ± .14	15	.11 ± .13	13	.10 ± .14	11
Adaptations	.23	.17 ± .10	25	.18 ± .11	24	.15 ± .10	21	.16 ± .11	20
Cultural Adaptations	.27	.25 ± .16	13	.25 ± .16	13	.23 ± .16	10	.23 ± .16	10
Pragmatic Adaptations	.26	.06 ± .23	5	.06 ± .23	5	.01 ± .23	4	.01 ± .23	4
Program Combinations	.12	.14 ± .19	7	.16 ± .19	6	.14 ± .19	7	.16 ± .18	6
All Adoptions and Adaptations	.23	.17 ± .08	42	.17 ± .08	39	.13 ± .08	34	.14 ± .09	31
Total	.32	.24 ± .04	158	.22 ± .04	141	.22 ± .05	132	.20 ± .04	119

Note: d = unweighted mean effect size; $d+$ = weighted effect-size according to the random effects model with 95% confidence interval; k = number of studies.

[a] $k = 158$.

[b] Sample restricted to studies with randomized controlled trial (see text).

[c] $k = 141$.

[d] Sample restricted to studies with sample sizes higher than 50 (see text).

[e] $k = 132$.

[f] $k = 119$.

TABLE 4
Results for the Swedish Study Sample

	A			B		C		D	
	Total Study Sample[a]			Study Sample Selected by Design[b,c]		Study Sample Selected by Sample Size[d,e]		Study Sample Selected by Design and Sample Size[b,d,f]	
Type of Program Construction	*d*	*d*+	*k*	*d*+	*k*	*d*+	*K*	*d*+	*k*
Novel Programs	.48	.43 ± .08	96	.47 ± .09	87	.40 ± .09	76	.45 ± .10	68
Innovations	.52	.47 ± .11	67	.51 ± .11	61	.42 ± .12	52	.47 ± .12	47
Conceptually New	.35	.36 ± .16	29	.38 ± .17	26	.38 ± .16	24	.41 ± .18	21
Adoptions	.27	.26 ± .16	32	.24 ± .17	26	.27 ± .16	25	.27 ± .18	20
Adaptations	.68	.54 ± .26	11	.61 ± .29	10	.37 ± .28	7	.44 ± .33	6
Cultural Adaptations	.54	.45 ± .38	5	.58 ± .45	4	.43 ± .38	4	.60 ± .47	3
Pragmatic Adaptations	.63	.73 ± .65	2	.73 ± .68	2	1.35 ± .82	1	1.35 ± .82	1
Program Combinations	.90	.60 ± .47	4	.61 ± .48	4	−.17 ± .53	2	−.17 ± .53	2
All Adoptions and Adaptations	.38	.33 ± .13	43	.34 ± .14	36	.29 ± .13	32	.31 ± .14	26
Total	.45	.40 ± .07	139	.43 ± .08	123	.37 ± .07	108	.41 ± .08	94

Note: *d* = unweighted mean effect size; *d*+ = weighted effect-size according to the random effects model with 95% confidence interval; *k* = number of studies.

[a] $k = 139$.

[b] Sample restricted to studies with randomized controlled trial (see text).

[c] $k = 123$.

[d] Sample restricted to studies with sample sizes higher than 50 (see text).

[e] $k = 108$.

[f] $k = 94$.

sizes decreased in most cases. For example, comparisons between innovation, conceptually new, adoptions, and adapted interventions yielded significant differences within Sample C, $Q(3) = 9.90$, $p < .02$, as in Study Sample A. In addition, differences between the main categories, novel programs, international adoptions, and adaptations were significant in Study Sample C, $Q(1) = 4.99$, $p < .03$, and marginally significant in Study Sample D, $Q(1) = 2.95$, $p < .09$.

In sum, innovations and cultural adaptations had the highest effect sizes in the German sample, whereas international adopted interventions and pragmatically adapted interventions led to lower and in most cases non-significant results.

Swedish Study Sample

The effect sizes (d) within the total study sample range from −1.37 to 2.33 with an unweighted mean of .45 (Table 4, column A). According to Hedges and Olkin's (1985) model, the mean weighted effect size *d*+ was .21 according to the fixed effect model, and 0.40 according to the REM. However, as the sample showed significant heterogeneity, $Q(138) = 857.68$, $p < .001$, $I^2 = 83.9\%$, all further calculations were made with the REM.

For the total sample (Table 4, Sample A), the highest mean effect size was for pragmatic adaptations (*d*+ = 0.73), eclectic adaptations (*d*+ = 0.59), innovations (*d*+ = 0.47), and cultural adaptations (*d*+ = 0.45). Adoptions (*d*+ = 0.26) showed the lowest effect sizes, albeit effect sizes all being significantly different from zero ($p < .05$), showing that all types of program origins were effective. Comparisons between the three main categories (novel program, adoption, and adaptation) show a trend, $Q(2) = 5.18$, $p < .08$, indicating lower effect sizes for adoptions. A similar trend was found when comparing the four most frequent categories (innovations, conceptually new programs, adoptions, and adaptions), $Q(3) = 6.40$, $p = .09$, favoring adaptations. In contrast to the results in the German sample, no difference was found when we compared all novel programs with all adoptions and adaptations, $Q(1) = 1.79$, *ns*.

Repeating the analysis while excluding low design ratings (Table 4, Study Sample B, $k = 123$) showed similar albeit significant results: three main categories $(2) = 6.97$, $p < .05$; four main categories $(3) = 8.27$, $p < .05$. The results for the additional analysis, taking sample size (Table 4, Study Sample C) and the combination of design ratings and sample size (Table 4, Study Sample D) into account, indicated a significant difference between all subcategories. This resulted from a single, very high effect size for one of the categories (pragmatic adaptations), which makes Study Samples C and D less reliable for comparisons.

Because the Swedish data set also included coding of other effect size moderators (e.g., type of control group, efficacy vs. effectiveness study), which accounted significantly for effect size variance, we tested if these

moderators are confounded with our program construction measure. However, there was no apparent systematic interrelations between these variables (i.e., they did not account for effect size differences between categories of program construction).

In sum, adaptations (including cultural, pragmatic, and eclectic adaptations) and innovations had the highest effect size, whereas adoptions tended to be less efficient in the Swedish sample.

DISCUSSION

The adaptation or adherence issue is one of the major unresolved dilemmas in the evidence-based intervention movement. To our knowledge, this is the first attempt to statistically investigate the effects of program origin (i.e., novel programs vs. international adopted and adapted programs) on program outcomes in two international meta-analytic data sets. Given the methodological limitations of the study, the conclusions are to be considered as hypothesis generating, indicating the need for the results to be replicated and further expanded by looking at different types of adaptation in more detail.

The results show that novel programs (i.e., completely or conceptually new national programs) were both the most effective and the most frequent (constituting approximately 70% of the programs) approach. In addition, compared to most of the other categories, international programs adopted without any adaptations had the lowest average effect sizes in both samples, even after some crucial methodological aspects were controlled statistically. Although adoptions proved to be effective (effect sizes were significantly different from zero), they were not as effective as most of the adapted programs or the novel programs in general. In light of the adaptation or adherence dilemma, the results suggest that adaptation may not always compromise observed program effects but instead increase them. In this respect, the results confirm those of previous studies based on more homogeneous samples that report on successful adaptations of settings, populations, and format (e.g., Andersson et al., 2014; Galovski et al., 2012; Michelson et al., 2013; Stanton et al., 2005). These results may indicate that transported programs should not be adopted without considering adaptation to the local setting.

This conclusion might be considered controversial, and it needs to be discussed in relation to the methodology of the study. For instance, the results are partially based on small study samples, especially when it comes to subtypes of adapted programs, and include several interventions types, outcomes, and target populations. Although comparisons according to simple categorization always entails the risk of being confounded with further study characteristics in meta-analysis (Lipsey, 2003), the problem of confounded moderators may present a particular danger with such a heterogeneous set of interventions as in this study.

Novel programs, particularly innovations, were the construction type with the highest effect size in the German study sample and among the highest in the Swedish sample. One reason for the high effect size may be that a novel program encompasses the greatest possible fit to the population and context where it takes place. Novel programs may even involve tailoring the program to the needs of the specific setting, compared to internationally transported programs that are originally developed for another context. In addition, a novel program may also be more likely to involve the program developers in the supervision of the trial, which has been shown to increase the efficacy of the program because of a greater engagement (e.g., Curtis et al., 2004; Emshoff, 2008; Petrosino & Soydan, 2005). However, these results may also be a consequence of conflict of interests, which should not be underestimated in evaluation research (Eisner & Humphreys, 2011). We were not able to directly test these alternative interpretations, simply because program development and type of program construction are interrelated (studies on adapted programs have, to our knowledge, never been done by the original program developers). Therefore, future research should initiate studies that compare novel, adopted, and adapted programs conducted by independent evaluators for an unbiased estimation of effects.

In this study, adaptation justified explicitly for cultural reasons was significantly more effective than international adopted programs without any adaptation. This adds to the prior literature, which has shown contradictory results regarding the effects of cultural adaptation (e.g., Dushay, Singer, Weeks, Rohena, & Gruber, 2001; Huey & Polo, 2008; Kumpfer et al., 2002; Stanton et al., 2005). One challenge in determining the influence of cultural adaptation on outcome is that the cultural context can be conceptualized in a variety of ways (Hofstede, 2001; Super & Harkness, 1999; Sussman, Unger, & Palinkas, 2008). Examples of contextual factors include local and national policies about the system and the services that are provided, a provider's perception of the evidence supporting the use of an evidence-based program, and characteristics of the individuals involved in the implementation effort. The importance of these cultural differences is widely acknowledged but has rarely been the focus of systematic research in translating an evidence-based program to new cultures (Castro, Barrera, & Holleran Steiker, 2010; Ferrer-Wreder, Sundell, & Mansoory, 2012). Similarly, there is no consensus about the criteria for determining when cultural adaptation is needed (Cardemil, 2010;

Ferrer-Wreder et al., 2012; Huey & Polo, 2008). Thus, although our study supports cultural adaptation, the specificities remain to be untangled. Our suggestion for future research is to focus on different types of cultural adaptation and their possible effects on program outcomes.

Methodological Considerations

This study involves several methodological challenges. One is that the coding of programs is based on the information available in the articles. Overall, adaptations are poorly described in intervention research (Hodge et al., 2010; Huey & Polo, 2008), thus preventing us from considering all the aspects of adaptation (e.g., dosage, adherence) or levels of modification (e.g., those related to surface vs. deep structure) that have previously been suggested (e.g., Moore et al., 2013; Stirman et al., 2013). The lack of detailed information on adaptation implies a risk of merging both substantial modifications (e.g., removal of core components) and less significant modifications (e.g., replacing material with more culturally relevant), thus risking confusion in the concept of adaptation. This issue is closely related to the lack of solid definitions of adaptation, adoption, and novel interventions. Program developers may help in this regard by clearly defining what are considered to be core components of programs, and identifying modifiable elements.

Another challenge is that there are several possible confounders related to method and design. For instance, the review included different subsets of psychosocial interventions, where some aspects of the intervention itself might be related to the degree and types of adaptation. In the present analysis, data were controlled for low design rating and sample sizes below 50, without altering the main results. This indicates the robustness of the findings. However, the impact of other confounders, which have been shown to influence effects, remains to be investigated in future research.

A third challenge is publication bias. The Swedish sample provides some information on this issue as the Swedish screening for intervention research identified another 59 studies that were not published in peer-reviewed articles. According to preliminary results by the first author (KS), the latter sample was less often randomized, more often targeted universal populations, and based on smaller samples than those published in peer-reviewed journals. However, there was no clear difference between the two samples in the types of novel, adopted, and adapted programs.

Because the two samples used in this study were initially gathered for other purposes (i.e., meta-analysis of efficacy of interventions), some differences between them may have influenced the results. First, the German sample is based on the average means of all outcomes in the articles, whereas the Swedish sample includes only the primary outcome. This may be one reason for the overall larger effect sizes in the results based on the Swedish study sample. A second reason may be that the majority of the German studies include universal prevention, which normally produces weaker effect sizes (e.g., Beelmann & Raabe, 2009; Stice, Shaw, Bohon, Marti, & Rohde, 2009), whereas these studies are a minority in the Swedish sample. Another difference is the use of passive controls (e.g., waitlist), which was the most frequent control condition in the German sample but a less common control condition in the Swedish sample. Because studies with passive control conditions have been related to larger effects compared to active control conditions (Baldwin, Christian, Berkeljon, & Shadish, 2012; Magill & Ray, 2009; Shadish, 2011), this would work in the opposite direction, producing larger effect sizes in the German study than the Swedish. Finally, the German sample includes only child and youth interventions, whereas the Swedish sample includes a wide variety of interventions. The exact implication of this is unknown. In sum, although the differences between the two samples affects the sizes of the effects within the samples, the differences also strengthen the validity of the results, as roughly the same results were produced using the two samples.

Implications for Intervention Research with Children and Adolescents

The results have implications for child and adolescent intervention research. The first is an urgent need for additional information on adaptation in coming efficacy and effectiveness trials in order to enable extended analyses of adaptation. Some steps have recently been taken to improve the accuracy, comprehensiveness, and transparency of study reports and to provide a framework to help reviewers assess implementation data across trials (Montgomery, Grant, et al., 2013; Montgomery, Underhill, Gardner, Operario, & Mayo-Wilson, 2013).

The second implication is that trials of transported empirically supported interventions should incorporate research designs that allow for a differential examination of various aspects of adaptation. A number of models for cultural adaptation are beginning to emerge from attempts to replicate interventions across and within countries (Ferrer-Wreder et al., 2012). Typically, these models prescribe a series of steps or decision-making guidelines for adapting, implementing, and evaluating an intervention for a new context.

Finally, we need a better understanding of what constitutes a novel program and how to differentiate it from an adapted program. Models for how to develop a new intervention program and for dimensions of

programs (e.g., Beelmann, 2011; Fraser, Richman, Galinsky, & Day, 2009; Roth & Brooks-Gunn, 2003) are helpful, but most studies lack detailed information on the program construction, making it generally difficult to clearly evaluate its foundation.

CONCLUSIONS

This study adds to the adaptation or adherence debate by describing the effects of novel programs, international adoptions, and contextual adaptations. It questions the general assumption that interventions are optimized prior to implementation in a new setting, particularly for complex social interventions (Chambers et al., 2013). Although the results need to be replicated and further expanded by looking at different type of adaptation in more detail, the results indicate that adoption without careful consideration of the need for adaptation should be discouraged. For this, we urge journal editors to request, and intervention researches to include, detailed information about adaptation in reports on interventions in order to allow more detailed analysis of when and how adaptation may improve outcomes.

SUPPLEMENTAL DATA

Supplemental data for this article can be accessed on the publisher's website (Appendix S1 and S2: Flowchart of study selection, additional information on codings of the program type and the references of integrated studies [German and Swedish study samples]).

REFERENCES

Andersson, G., Cuijpers, P., Carlbring, P., Riper, H., & Hedman, E. (2014). Guided internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: A systematic review and meta-analysis. *World Psychiatry*, *13*, 288–295. doi:10.1002/wps.20151

Arthur, M. W., & Blitz, C. (2000). Bridging the gap between research and practice in the drug abuse prevention through needs assessment and strategic community planning. *Journal of Community Psychology*, *28*, 241–256. doi:10.1002/(SICI)1520-6629(200005)28:3<241::AID-JCOP2>3.0.CO;2-X

Backer, T. E. (2000). The failure of success: Challenges of disseminating effective substance absue prevention programs. *Journal of Community Psychology*, *28*, 363–373. doi:10.1002/(SICI)1520-6629(200005)28:3<363::AID-JCOP10>3.0.CO;2-T

Baldwin, S. A., Christian, S., Berkeljon, A., & Shadish, W. R. (2012). The effects of family therapies for adolescent delinquency and substance abuse: A meta-analysis. *Journal of Marital and Family Therapy*, *38*, 281–304. doi:10.1111/j.1752-0606.2011.00248.x

Bauman, L. J., Stein, R. E. K., & Ireys, H. T. (1991). Reinventing fidelity: The transfer of social technology among settings. *American Journal of Community Psychology*, *19*, 619–639. doi:10.1007/BF00937995

Becker, D. R., Smith, J., Tanzman, B., Drake, R. E., & Tremblay, T. (2001). Fidelity of supported employment programs and employment outcomes. *Psychiatric Services*, *52*, 834–836. doi:10.1176/appi.ps.52.6.834

Beelmann, A. (2011). The scientific foundation of prevention: The status quo and future challenges for developmental crime prevention. In T. Bliesener, A. Beelmann, & M. Stemmler (Eds.), *Antisocial behavior and crime: Contributions of developmental and evaluation research to prevention and intervention* (pp. 137–164). Cambridge, MA: Hogrefe.

Beelmann, A. (2014). Möglichkeit und Grenzen systematischer Evidenzkumulation durch Forschungssynthesen in der Bildungsforschung [Potentials and limits of systematic evidence accumulation via meta-analysis within the educational research]. *Zeitschrift für Erziehungswissenschaft*, *17*(Suppl. 4), 55–78. doi:10.1007/s11618-014-0509-2

Beelmann, A., Pfost, M., & Schmitt, C. (2014). Prävention und Gesundheitsförderung bei Kindern und Jugendlichen. Eine Meta-Analyse der deutschsprachigen Wirksamkeitsforschung [Prevention and health promotion in children and adolescents. A meta-analysis of the German language outcome research]. *Zeitschrift für Gesundheitspsychologie*, *22*, 1–14. doi:10.1026/0943-8149/a000104

Beelmann, A., & Raabe, T. (2009). The effects of preventing antisocial behavior and crime in childhood and adolescence: Results and implications of research reviews and meta-analyses. *European Journal of Developmental Science*, *3*, 260–281.

Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Sharp Minicucci, D., Ory, M., ... Treatment Fidelity Workgroup of the NIH Behavior Change Consortium. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH behavior change consortium. *Health Psychology*, *23*, 443–451. doi:10.1037/0278-6133.23.5.443

Benish, S. G., Quintana, S., & Wampold, B. E. (2011). Culturally adapted psychotherapy and the legitimacy of myth: A direct-comparison meta-analysis. *Journal of Counseling Psychology*, *58*, 279–289. doi:10.1037/a0023626

Cardemil, E. V. (2010). Cultural adaptations to empirically supported treatments: A research agenda. *The Scientific Review of Mental Health Practice*, *7*, 8–21.

Castro, F. G., Barrera, M., & Holleran Steiker, L. K. (2010). Issues and challenges in the design of culturally adapted evidence-based interventions. *Annual Review of Clinical Psychology*, *6*, 213–239. doi:10.1146/annurev-clinpsy-033109-132032

Castro, F. G., Barrera, M., & Martinez, C. R. (2004). The cultural adaptation of prevention interventions: Resolving between fidelity and fit. *Prevention Science*, *5*, 41–45. doi:10.1023/B:PREV.0000013980.12412.cd

Chambers, D. A., Glasgow, R. E., & Stange, K. C. (2013). The dynamic sustainability framework: Addressing the paradox of sustainment amid ongoing change. *Implementation Science*, *8*, 117. doi:10.1186/1748-5908-8-117

Curtis, N. M., Ronan, K. R., & Borduin, C. M. (2004). Multisystemic treatment: A meta-analysis of outcome studies. *Journal of Family Psychology*, *18*, 411–419. doi:10.1037/0893-3200.18.3.411

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, *18*, 23–45. doi:10.1016/S0272-7358(97)00043-3

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American*

Journal of Community Psychology, *41*, 327–350. doi:10.1007/s10464-008-9165-0

Dushay, R. A., Singer, M., Weeks, M. R., Rohena, L., & Gruber, R. (2001). Lowering HIV risk among ethnic minority drug users: Comparing culturally targeted intervention to a standard intervention. *The American Journal of Drug and Alcohol Abuse*, *27*, 501–524. doi:10.1081/ADA-100104515

Eisner, M., & Humphreys, D. (2011). Measuring conflict of interest in prevention and intervention research. A feasibility study. In T. Bliesener, A. Beelmann, & M. Stemmler (Eds.), *Antisocial behavior and crime. Contributions of developmental and evaluation research to prevention and intervention* (pp. 165–180). Cambridge, MA: Hogrefe.

Elliott, D., & Mihalic, S. (2004). Issues in disseminating and replicating effective prevention programs. *Prevention Science*, *5*, 47–53. doi:10.1023/b:prev.0000013981.28071.52

Emshoff, J. G. (2008). Researchers, practitioners, and funders: Using the framework to get us on the same page. *American Journal of Community Psychology*, *41*, 393–403. doi:10.1007/s10464-008-9168-x

Ferrer-Wreder, L., Sundell, K., & Mansoory, S. (2012). Tinkering with perfection: Theory development in the intervention cultural adaptation field. *Child & Youth Care Forum*, *41*, 149–171. doi:10.1007/s10566-011-9162-6

Fraser, M. W., Richman, J. M., Galinsky, M. J., & Day, S. H. (2009). *Intervention research. Developing social programs*. New York: Oxford University Press.

Galovski, T. E., Blain, L. M., Mott, J. M., Elwood, L., & Houle, T. (2012). Manualized therapy for PTSD: Flexing the structure of cognitive processing therapy. *Journal of Consulting and Clinical Psychology*, *80*, 968–981. doi:10.1037/a0030600

Gone, J. P. (2007). ''We never was happy living like a Whiteman'': Mental health disparities and the postcolonial predicament in American Indian communities. *American Journal of Community Psychology*, *40*, 290–300. doi:10.1007/s10464-007-9136-x

Gottfredson, D., Kumpfer, K., Polizzi-Fox, D., Wilson, D., Puryear, V., Beatty, P., & Vilmenay, M. (2006). The strengthening Washington D.C. families project: A randomized effectiveness trial of family-based prevention. *Prevention Science*, *7*, 57–74. doi:10.1007/s11121-005-0017-y

Griner, D., & Smith, T. B. (2006). Culturally adapted mental health interventions: A meta-analytic review. *Psychotherapy: Theory, Research, Practice, Training*, *43*, 531–548. doi:10.1037/0033-3204.43.4.531

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.

Hill, L. G., Maucione, K., & Hood, B. K. (2007). A focused approach to assessing program fidelity. *Prevention Science*, *8*, 25–34. doi:10.1007/s11121-006-0051-4

Hodge, D. R., Jackson, K. F., & Vaughn, M. G. (2010). Culturally sensitive interventions for health related behaviors among Latino youth: A meta-analytic review. *Children and Youth Services Review*, *32*, 1331–1337. doi:10.1016/j.childyouth.2010.05.002

Hodge, D. R., Jackson, K. F., & Vaughn, M. G. (2012). Culturally sensitive interventions and substance use: A meta-analytic review of outcomes among minority youths. *Social Work Research*, *36*, 11–19. doi:10.1093/swr/svs008

Hofstede, G. (2001). *Culture's consequences. Comparing values, behaviors, institutions and organizations across nations*. Thousand Oaks, CA: Sage.

Huey, S. J., & Polo, A. J. (2008). Evidence-based psychosocial treatments for ethnic minority youth. *Journal of Clinical Child & Adolescent Psychology*, *37*, 262–301. doi:10.1080/15374410701820174

Jackson, K. F., & Hodge, D. R. (2010). Native American youth and culturally sensitive interventions: A systematic review. *Research on Social Work Practice*, *20*, 260–270. doi:10.1177/1049731509347862

Kaunitz, C., & Strandberg, A. (2009). Aggression Replacement Training (ART) i Sverige - evidensbaserad socialtjänst i praktiken [Aggression Replacement Training (ART) in Sweden-evidence-based social work practice]. *Socionomens forskningssupplement*, *26*, 36–52.

Keith, R. E., Hopp, F. P., Subramanian, U., Wiitala, W., & Lowery, J. C. (2010). Fidelity of implementation: Development and testing of a measure. *Implementation Science*, *5*, 99. doi:10.1186/1748-5908-5-99

Kilbourne, A. M., Neumann, M. S., Pincus, H. A., Bauer, M. S., & Stall, R. (2007). Implementing evidence-based interventions in health care: Application of the replicating effective programs framework. *Implementation Science*, *2*, 42. doi:10.1186/1748-5908-2-42

Kumpfer, K. L, Alvarado, R., Smith, P., & Bellamy, N. (2002). Cultural sensitivity and adaptation in family-based prevention interventions. *Prevention Science*, *3*, 241–246. doi:10.1023/A:1019902902119

Lee, S. J., Altschul, I., & Mowbray, C. T. (2008). Using planned adaptation to implement evidence-based programs with new populations. *American Journal of Community Psychology*, *41*, 290–303. doi:10.1007/s10464-008-9160-5

Lipsey, M. W. (2003). Those confounded moderators in meta-analysis: Good, bad, and ugly. *The Annals of the American Academy of Political and Social Science*, *587*, 69–81. doi:10.1177/0002716202250791

Lipsey, M. W., & Wilson, D. B. (2000). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Magill, M., & Ray, L. A. (2009). Cognitive-behavioral treatment with adult alcohol and illicit drug users: A meta-analysis of randomized controlled trials. *Journal of Studies on Alcohol and Drugs*, *70*, 516–527.

McKleroy, V. S., Galbraith, J. S., Cummings, B., Jones, P., Harshbarger, C., Collins, C., ... ADAPT Team. (2006). Adapting evidence-based behavioral interventions for new settings and target populations. *AIDS Education and Prevention*, *18*(Suppl.), 59–73. doi:10.1521/aeap.2006.18.supp.59

Michelson, D., Davenport, C., Dretzke, J., Barlow, J., & Day, C. (2013). Do evidence-based interventions work when tested in the ''real world?'' A systematic review and meta-analysis of parent management training for the treatment of child disruptive behavior. *Clinical Child and Family Psychology Review*, *16*, 18–34. doi:10.1007/s10567-013-0128-0

Montgomery, P., Grant, S., Hopewell, S., Macdonald, G., Moher, D., Michie, S., ... Mayo-Wilson, E. (2013). Protocol for CONSORT-SPI: An extension for social and psychological interventions. *Implementation Science*, *8*, 99. doi:10.1186/1748-5908-8-99

Montgomery, P., Underhill, K., Gardner, F., Operario, D., & Mayo-Wilson, E. (2013). The Oxford implementation index: A new tool for incorporating implementation data into systematic reviews and meta-analyses. *Journal of Clinical Epidemiology*, *66*, 874–882. doi:10.1016/j.jclinepi.2013.03.006

Moore, J. E., Bumbarger, B. K., & Cooper, B. R. (2013). Examining adaptations of evidence-based programs in natural contexts. *The Journal of Primary Prevention*, *34*, 147–161. doi:10.1007/s10935-013-0303-6

Petrosino, A., & Soydan, H. (2005). The impact of program developers as evaluators on criminal recidivism: Results from a meta-analysis of experimental and quasi-experimental research. *Journal of Experimental Criminology*, *1*, 435–450. doi:10.1007/s11292-005-3540-8

Resnicow, K., Soler, R., Braithwaite, R. L., Ahluwalia, J. S., & Butler, J. (2000). Cultural sensitivity in substance use prevention. *American Journal of Community Psychology*, *28*, 271–290. doi:10.1002/(SICI)1520-6629(200005)28:3<271::AID-JCOP4>3.0.CO;2-I

Ringwalt, C. L., Vincus, A., Ennett, S., Johnson, R., & Rohrbach, L. A. (2004). Reasons for teachers' adaptation of substance use prevention curricula in schools with non-white student populations. *Prevention Science*, *5*, 61–67. doi:10.1023/B:PREV.0000013983.87069.a0

Rogers, E. M. (2003). *Diffusion on innovations* (5th ed.). New York, NY: Free Press.

Roth, J. L., & Brooks-Gunn, J. (2003). What exactly is a youth development program? Answers from research and practice. *Applied Developmental Science*, *7*, 94–111. doi:10.1207/S1532480XADS0702_6

Shadish, W. R. (2011). Randomized controlled studies and alternative designs in outcome studies: Challenges and opportunities. *Research on Social Work Practice*, *21*, 636–643. doi:10.1177/1049731511403324

Skärstrand, E., Larsson, J., & Andreasson, S. (2008). Cultural adaptation of the strengthening families programme to a Swedish setting. *Health Education*, *108*, 287–300. doi:10.1108/09654280810884179

Skärstrand, E., Sundell, K., & Andreasson, S. (2013). Evaluation of a Swedish version of strengthening families program. *The European Journal of Public Health*, *24*, 578–584. doi:10.1093/eurpub/ckt146

Stanton, B., Guo, J., Cottrell, L., Galbraith, J., Li, X., Gibson, C., … Harris, C. (2005). The complex business of adapting effective interventions to new populations: An urban to rural transfer. *Journal of Adolescent Health*, *37*, 163. doi:10.1016/j.jadohealth.2004.10.005

Stice, E., Shaw, H., Bohon, C., Marti, C. N., & Rohde, P. (2009). A meta-analytic review of depression prevention programs for children and adolescents: Factors that predict magnitude of intervention effects. *Journal of Consulting and Clinical Psychology*, *77*, 486–503. doi:10.1037/a0015168

Stirman, S. W., Miller, C. J., Toder, K., & Calloway, A. (2013). Development of a framework and coding system for modifications and adaptations of evidence-based interventions. *Implementation Science*, *8*, 65. doi:10.1186/1748-5908-8-65

Sundell, K. (2012). *Svenska effektutvärderingar av psykosociala interventioner [Swedish outcome research of psychological and social interventions]*. Stockholm, Sweden: Socialstyrelsen.

Sundell, K., Ferrer-Wreder, L., & Fraser, M. W. (2014). Going global: A model for evaluating empirically-supported family-based interventions in new contexts. *Evaluation & the Health Professions*, *37*, 203–230. doi:10.1177/0163278712469813

Super, C. M., & Harkness, S. (1999). The environment as culture in developmental research. In S. L. Friedman & T. D. Wachs (Eds.), *Measuring environment across the life span* (pp. 279–323). Washington, DC: American Psychological Association.

Sussman, S., Unger, J. B., & Palinkas, L. A. (2008). Country prototypes and translation of health programs. *Evaluation & the Health Professions*, *31*, 110–123. doi:10.1177/0163278708315918

Weisz, J. R., Donenberg, G. R., Han, S. S., & Weiss, B. (1995). Bridging the gap between laboratory and clinic in child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology*, *63*, 688–701. doi:10.1037/0022-006X.63.5.688