



UvA-DARE (Digital Academic Repository)

Searching for similarities: transfer-oriented learning in health education at secondary schools

Peters, L.W.H.

Publication date
2012

[Link to publication](#)

Citation for published version (APA):

Peters, L. W. H. (2012). *Searching for similarities: transfer-oriented learning in health education at secondary schools*. [Thesis, fully internal, Universiteit van Amsterdam].

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

Chapter 4

EFFECTIVE ELEMENTS OF SCHOOL HEALTH PROMOTION ACROSS BEHAVIORAL DOMAINS: A SYSTEMATIC REVIEW OF REVIEWS¹

Most school health education programs focus on a single behavioral domain. Integrative programs that address multiple behaviors may be more efficient, but only if the elements of change are similar for these behaviors. The objective of this study was to examine which effective elements of school health education are similar across three particular behavioral domains.

A systematic review of reviews of the effectiveness of school-based health promotion programs was conducted for the domains of substance abuse, sexual behavior, and nutrition. The literature search spanned the time period between 1995 and October 2006 and included three databases, websites of review centers and backward search. Fifty-five reviews and meta-analyses met predetermined relevance and publication criteria and were included. Data was extracted by one reviewer and checked by a second reviewer. A standardized data extraction form was used, with detailed attention to effective elements pertaining to program goals, development, content, methods, facilitator, components and intensity. Two assessors rated the quality of reviews as strong, moderate or weak. We included only strong and moderate reviews in two types of analysis: one based on interpretation of conflicting results, the other on a specific vote-counting rule.

Thirty six reviews were rated strong, 6 moderate, and 13 weak. A multitude of effective elements was identified in the included reviews and many elements were similar for two or more domains. In both types of analysis, five elements with evidence from strong reviews were found to be similar for all three domains: use of theory; addressing social influences, especially social norms; addressing cognitive-behavioral skills; training of facilitators; and multiple components. Two additional elements had positive results in all domains with the rule-based method of analysis, but had inconclusive results in at least one domain with the interpretation-based method of analysis: parent involvement and a larger number of sessions.

Five effective elements of school health promotion were found to be similar across the three behavioral domains examined (substance abuse, sexual behavior, nutrition). An integrative program that addresses the three domains seems feasible. The five elements are primary candidates to include in programs targeting these behaviors.

¹ Peters, L. W. H., Kok, G., Ten Dam, G. T. M., Buijs, G. J., & Paulussen, T. G. W. M. (2009). *Effective elements of school health promotion across behavioral domains: a systematic review of reviews*. BMC Public Health, 9, 182.

1. BACKGROUND

Adolescents are a popular target group for health education and promotion programs because many health-risk behaviors, which contribute to the leading causes of morbidity and mortality among youth and adults, develop or augment during adolescence (Currie et al., 2006; Eaton et al., 2006). These behaviors include use of tobacco, alcohol and other substances, unprotected sexual activity, poor dietary habits, physical inactivity, and behaviors that contribute to unintentional injuries and violence. More and more evidence shows that several of these behaviors tend to co-occur (Basen-Engquist, Edmundson, & Parcel, 1996; Donovan, Jessor, & Costa, 1991; Driskell, Dymont, Mauriello, Castle, & Sherman, 2008; DuRant, Smith, Kreiter, & Krowchuk, 1999; Prochaska, Spring, & Nigg, 2008; Sallis, Prochaska, & Taylor, 2000; Wiefferink et al., 2006) and have similar determinants (Peters, Wiefferink, et al., 2009; Wiefferink et al., 2006), which opens up opportunities for integrative programs that address multiple behaviors (Prochaska, 2008). Yet, most adolescent health promotion programs continue to address only one behavioral domain.

The majority of adolescent health promotion programs are intended for use in schools, often as a supplement to the regular curriculum. In many countries school staff feel overwhelmed by the ever-increasing supply of prevention programs, especially since they are faced with overcrowded curricula and limited opportunities for implementing prevention programs (Lee, Keung, & Tsang, 2004; Leurs, Jansen, Schaalma, Mur-Veeman, & De Vries, 2005). Integrative programs that address multiple risk behaviors effectively and efficiently may reduce the burden on schools and teachers (Ten Dam, 2002). Several authors have suggested that integrative programs can be efficient if the change processes or effective elements for different health behaviors are similar (Paulussen, Panis, Peters, Buijs, & Wijnsma, 1998; Prochaska et al., 2008).

The observation that most programs focus on a single behavior also holds for the review literature that discusses effectiveness and effective elements of school-based health promotion. As Prochaska (2008, p. 283) argues, “science tends to value specificity, and specialists are trained to know what is specific to their disciplines rather than what is common across disciplines”. Although many authors have observed that elements of effective programs appear to be similar across different behaviors (Nation et al., 2003; Schaalma, Abraham, Gillmore, & Kok, 2004; Summerfield, 2002; Thomas, Micucci, Ciliska, & Mirza, 2005), only a few authors have yet examined these commonalities systematically (Nation et al., 2003). Knowledge of the similarities and dissimilarities of effective programs across behavioral domains may not only contribute to the development or elaboration of integrative programs. It may also deepen our understanding of what does and does not work in school health promotion and may contribute to transfer of knowledge and ideas from one domain to another.

The present review focuses on similarities between effective elements of school health education programs across three behavioral domains: substance abuse, sexual behavior and healthy nutrition. It was conducted to inform development of an inte-

grative educational program that addresses all three domains. These domains were selected because they are among the ones most frequently addressed in Dutch secondary schools (Dafesh, 2006).

In light of the task of assessing three domains and the extensive body of literature on effectiveness that already exists in these domains, we opted for a review-of-reviews approach. As Nation and colleagues (2003) stated, prevention now has a sufficient knowledge base to begin a meta-assessment of the characteristics of effective prevention programming. More and more, reviews draw on previous reviews for making statements about effectiveness (e.g., Ellis & Grey, 2004; Micucci, Thomas, & Vohra, 2002; Mulvihill & Quigley, 2003; Nation et al., 2003; Poobalan, Taylor, Clar, Helms, & Smith, 2008; Thomas et al., 2005).

2. METHODS

2.1 *Literature searches and inclusion/exclusion criteria*

Three internet databases (Pubmed, PsycINFO, ERIC) were searched for relevant reviews published between January 1995 and October 2006 by combining groups of keywords pertaining to school health promotion, effectiveness and the three health behavior domains (see Table 1), generating over 1600 papers. The number and types of databases searched can be considered comprehensive (Thomas et al., 2005) and efficient for locating literature about effectiveness of health promotion (Peersman, Harden, Oliver, & Oakley, 1999). Also, the internet sites of six international review initiatives were searched for relevant reviews (see Table 1) and reference lists of already retrieved publications were scanned for additional reviews.

Titles and abstracts of publications were screened for relevance, and in case of doubt, entire publications were checked. Reviews were deemed relevant if they: a) included a review of primary effect studies (reviews of reviews were excluded); b) focused on one or more of the targeted risk behaviors (substance abuse, early or unprotected sexual behavior, dietary behavior); c) focused on regular, secondary-school-age youth or adolescents (12-18 years); d) included school-based programs with an educational approach; and e) discussed programs implemented in western countries. Furthermore, reviews had to be written in English, be published in a peer-reviewed journal listed on the Thomson Scientific master journal list or by an international review initiative, and be available over the Internet or from university libraries in The Netherlands.

Fifty five reviews met these criteria and were included: 5 about multiple domains of our interest (mostly about substance abuse and sexual behavior, see references 26-30 in Table S1 at the end of this chapter), 24 about substance abuse (references 31-54 in Table S1), 17 about sexual behavior (references 55-71 in Table S1) and 9 about nutrition (references 72-80 in Table S1).

Table 1. Databases and keywords used in search strategies

Databases			
Pubmed keywords	PsycINFO keywords	ERIC keywords	Review initiatives websites
School health promotion: Curriculum Health-education Health-promotion School-health-services Health-plan-implementation Effectiveness: Program-evaluation Evaluation-studies Risk-reduction-behavior Behavior focus: Smoking Alcohol-drinking Sex-education Diet Food-habits	School health promotion: Curriculum Curriculum-development Educational-programs Schools School-environment Health-education Health-promotion Effectiveness: Effectiveness Educational-program-evaluation Treatment-effectiveness-evaluation Health-attitudes Health-behavior Health-knowledge Behavior focus: Tobacco-smoking Alcohol-abuse Safe-sex Sex-education Sexuality Sexually-transmitted-diseases Food-intake Nutrition Health-behavior Lifestyle	School health promotion: Curriculum School-health-services Health-programs Health-education Comprehensive-school-health-education Intervention Instruction Effectiveness: Program-effectiveness Program-evaluation Program-implementation Outcomes-of-education Knowledge-level Feedback Learning Behavior focus: Tobacco Smoking Alcohol-education Drinking Substance-abuse Sex-education Sexuality Nutrition Nutrition-instruction Eating-habits	Campbell Collaboration Centre for Reviews and Dissemination, York UK Cochrane Collaboration Effective Public Health Practice Project, Hamilton Canada EPPI-Centre, London UK Guide to Community Preventive Services

Publication year: January 1995 – October 2006. Language: English.

Note: The keywords within one group of keywords (e.g., school health promotion) were combined with 'OR', the groups were combined with 'AND'.

Of the substance abuse reviews, 5 focused specifically on tobacco, 4 on alcohol and the remaining 15 addressed tobacco and/or alcohol, possibly in combination with other substances. All reviews about multiple domains addressed substance abuse and sexuality programs and two also included nutrition programs. As some of these reviews focused on specific types of programs (e.g. peer programs) and not so much on specific behavioral domains, the results were usually not discussed for each specific domain. Some reviews in the multiple behavior and nutrition categories also addressed behaviors outside our focus (e.g. exercise), but results for these additional behaviors were not recorded.

2.2 *Data extraction*

A standardized form (available from the first author) containing 27 categories was used for recording information about the characteristics and results of the 55 included reviews. This form was developed ad hoc for this review, but was based on tools previously used by others. Nine categories, derived from other reviews of reviews (Micucci et al., 2002; Thomas et al., 2005), pertained to characteristics of the focus and methods of each review: general and specific behavior focus, target population, intervention setting, type of review, time span, number of studies included, criteria for study design and outcome measures. One category was used for recording general results with respect to effectiveness, such as overall effect sizes or general statements. The other 17 categories addressed results with respect to effective elements of programs, participants or studies. This level of specificity was chosen to maximize learning about characteristics associated with effectiveness. Seven of these categories, which are all discussed in this review, pertained to elements of programs: focus/goal, development, content, methods, facilitator, components, and intensity. The remaining 10 categories pertained to elements of participants (e.g. gender, pre-test risk behavior) or studies (e.g., type of study design, length of follow-up). The three main categories of effective elements (programs, participants and studies) and specific elements within these categories (e.g., for program characteristics: goal, development, et cetera) are commonly used in data extraction forms of systematic reviews (e.g., see Ammerman, Lindquist, Lohr, & Hersey, 2002; Tobler et al., 2000). Due to the length of this paper we will not discuss the results for elements of participants or studies in full but will only address them when they are relevant to results for program elements.

Results and statements about effectiveness and effective elements were recorded in the appropriate categories as specifically as possible, often by literally quoting the review author. In addition, the results of each review were summarized using the symbols +, -, 0 and ? for respectively a positive, negative, null or unclear contribution of the element to effectiveness. This 'shorthand notation' facilitated tabulation, whereas the underlying extensive information warranted preservation of details. This process resulted in a 195-page summary document and an 80-page document with tables.

The first author extracted all data and conferred with the third author in case of doubt about interpretation or recording of a specific result; this was the case with 20

reviews. The third author also read six reviews (11%) and checked all data extracted from these reviews; only a few disagreements were found and these were discussed until a unanimous decision was reached.

2.3 *Quality rating*

The included reviews were rated for methodological quality using the Quality Assessment Tool for Reviews. This tool was developed by the Effective Public Health Practice Project and has been used in several reviews of reviews (Micucci et al., 2002; Poobalan et al., 2008; Thomas et al., 2005). It comprises the following seven criteria, which are all awarded one point, with a maximum score of 0 to 7: a) statement of the search strategy; b) comprehensiveness of the search; c) description of relevance criteria; d) some quality assessment of primary studies; e) comprehensive quality assessment of primary studies; f) integration of findings; and g) adequacy of the reported data to support the review's conclusions. Quality was rated by two raters in a staged manner. First, the independent ratings of 13 reviews were compared (inter-rater reliability overall: kappa=0.639, $p < .001$), and disagreements were discussed and resolved. Then, the remaining reviews were rated independently, and compared, and any disagreements were discussed until all scores were unanimous. Reviews were rated strong if they met six or seven of the criteria, moderate if they met four or five, and weak if they scored three or less. Strong reviews tend to be systematic, and weak reviews tend to be traditional narrative reviews. In addition to quality criteria d and e, which are quite general and only ask whether reviews assessed the quality of primary studies, we recorded which specific methodological inclusion criteria were applied in reviews (see Table S1).

2.4 *Analysis*

For each program element, the results of included reviews were compared, first within each domain, then across domains. Following procedures used in other reviews of reviews (Ellis & Grey, 2004; Micucci et al., 2002; Poobalan et al., 2008), only the results of strong and moderate reviews were considered for statements about effective elements. We considered a program element to be effective in a particular domain if it was labeled as such in at least one strong or moderate review from that domain and, in case of multiple reviews, if the overall conclusion was positive. If strong and/or moderate reviews in one domain had conflicting results (e.g., positive versus null results), we attempted to reach an overall conclusion by examining the methodology of the reviews (e.g., did follow-up periods or criteria for effectiveness differ between reviews?) and giving the highest weight to the review with the highest quality score, the strictest methodological criteria, and the clearest and most narrowly defined operationalizations; if no overall conclusion could be drawn the evidence was considered to be inconclusive.

Additionally, it was examined whether the results would be the same when using an alternative analytical approach, which was derived from others (Ellis & Grey, 2004). In this second type of analysis, the strength of evidence is rated as sufficient,

tentative or insufficient based on explicit rules. The evidence is sufficient if it is based on conclusions in at least one strong review from that domain and if there are no conflicting conclusions between strong reviews. The evidence is tentative if it is based on at least one moderate review or if the conclusions of strong reviews conflict (e.g., positive versus null results). If moderate reviews have conflicting conclusions, the evidence is considered to be insufficient. The main differences between the two types of analysis are that the second type strictly distinguishes between strong and moderate reviews and relies on a strict rule for handling conflicting results, whereas the first type relies more on interpretation of conflicting results. Hence, the first type is called interpretation-based and the second is called rule-based.

The results of weak reviews were deemed to be too questionable for conclusions about effective elements. However, in light of the focus of this review on similarities across domains, they were included in a supplementary way. Specifically, if a particular element had evidence from strong or moderate reviews in at least one domain, the results of weak reviews in other domains were explored and treated as a suggestion that the element might be effective in these other domains.

3. RESULTS

The results of the literature review are displayed in Tables S1 to S8 at the end of this chapter. In Table S1, references are identified by a reference number, in Tables S2 to S9 only the reference numbers are displayed. In order to facilitate the combined reading of the text and the results tables, we refer to publications in the text of the Results section by means of their reference number.

3.1 *Characteristics, relevance and quality rating of included reviews*

Table S1 gives an overview of characteristics of the 55 reviews. The reviews are categorized by behavior focus, and within these categories, by quality rating and publication year.

In addition to - or instead of - a preset focus on one or more behaviors, some reviews focused on specific populations (e.g., young adolescent girls [39]), intervention types (e.g., peer education [26,37,69]) or even specific programs (e.g., Life Skills Training [47]). Such specific foci are reported in Table S1.

All reviews included school-based programs (not reported in Table S1), and 23 of them entirely focused on programs in this setting, among which 15 in the substance abuse domain. Substance abuse prevention and sex education are usually implemented in secondary schools (junior high and/or senior high) and may also include the upper elementary grades 5-6. This corresponds with the age range most frequently stated in reviews: 11-18 years. Many nutrition reviews also included younger elementary-aged children.

The number of included studies differs widely across the reviews (3-144 studies) and appears to be largely due to differences in review focus (e.g., specific program

type) and strictness of methodological inclusion criteria. For reviews that provided sufficient information about studies, we recorded in Table S1 how many of the included primary studies met our relevance criteria (targeted behaviors, secondary-school-age, school-based educational intervention). In the nutrition domain, some reviews included only one relevant study, as most nutrition programs target elementary students. For these reviews, only the results of this one study were recorded. In the other domains, the number of relevant studies was much larger, and often all studies were relevant.

Except for a review about sexual knowledge [67], all reviews applied behavioral criteria to determine program effectiveness. Many reviews also addressed effects on psychosocial determinants, and in the sexuality domain one third of reviews examined results for biological outcomes such as pregnancy.

As for the quality rating, 36 reviews (65%) were rated strong, 6 moderate (11%), and 13 weak (24%). Weak reviews generally did not report methodological inclusion criteria, whereas strong reviews did. Criteria used most frequently pertained to study design and outcome measure; other criteria were much less frequently applied, e.g. for equivalence of groups, minimal follow-up period, or reporting of all outcomes. The inclusion criteria differed markedly, even between strong reviews. Many strong reviews subjected the included studies to additional quality rating. Fifteen reviews applied meta-analytic techniques (mostly in the substance abuse and sexuality domains, not reported in Table S1) and nearly all of them had a quality score of 7.

3.2 *Effect sizes and general statements about effectiveness*

Qualitative statements about the occurrence or magnitude of behavioral effects were cautiously positive in most reviews. Only very few reviews reported overall absence of effects and none reported overall negative effects. There do not appear to be clear relationships between type of statement and behavioral domain or review quality. The quantitative results of meta-analyses and reviews, expressed in effect sizes (ES), odds ratios (*OR*) or percentage reductions, are in line with the above mentioned qualitative statements in the reviews: in light of Cohen's [81] classification of ES as small (.20), medium (.50) or large (.80), many ESs reported in reviews were statistically significantly different from zero, explaining positive statements, but most can be considered small, explaining reservations.

In the *substance abuse* domain, average ESs reported for tobacco use ranged from -.02 [41: for the total set of non-interactive programs] to .29 [32: for life skills programs evaluated within 12 months after end of the program], with most meta-analyses reporting ESs between .10 and .18 [32,40,41,43,45]. Botvin and colleagues [29,47,53] reported typical reductions of 30-50% for social influence programs and 40-80% for life skills programs. A review of long term (> 2 years) tobacco outcomes reported a mean reduction of 11.4% in the percentage of baseline nonusers who initiated smoking [44]. For alcohol use, meta-analyses [40,41] and reviews [29,44,47,53] have reported ESs and percentage reductions of the same magnitude as for tobacco use.

In the *sexuality* domain the results vary per outcome measure examined and per review. Statistically significant positive effects have been reported for condom use (ES=.07 [56]; OR=.66 [58]). For birth control, one meta-analysis that included non-controlled studies found statistically significant positive effects (ES=.27 [61]) but a meta-analysis with stricter study design criteria did not [57]. Of five reviews that examined sexual activity, frequency or number of partners, two reported statistically significant positive effects (both ES=.05 [56,59]), whereas the other three did not [57,58,61]. No effects were found on diagnosis with STD [56,58]. As for pregnancy, the meta-analysis that included non-controlled studies reported a positive effect (ES=.15 [61]), whereas one with stricter criteria found no effect for females and a negative effect for males (OR=1.54 [57]).

In the *nutrition* domain, statistically significant positive effects have been reported for intake of fat (OR=2.19 [75]) and fruit and vegetables (increase of .30 to .99 servings per day [72]). One intensive high school intervention even increased daily servings of fruit and vegetables by over 2.5 [74,76].

ESs reported for psychosocial determinants are usually larger than those for behavior. In the substance use domain, a meta-analysis [41] reported an average ES of .38 for knowledge, .26 for attitude and .24 for skills for programs with much peer interaction. A tobacco-specific meta-analysis [32] reported comparable ESs for knowledge (.53 to .19, depending on the follow-up interval), attitude (.22 to .10), and skills (.22 to .09). In the sexuality domain, the following ESs have been reported: .41 for knowledge [67], .30 for condom use skills and .50 for condom negotiation skills [56].

3.3 *Effective elements of programs*

The results for the various categories of program elements are presented in Tables S2-S8 and are discussed in separate paragraphs below. As stated in the Methods section, the analysis focused on results of strong and moderate reviews; weak reviews were only used for supplementary purposes in the absence of stronger reviews. The elements are italicized in the text below to enhance combined reading of text and tables, and elements that are considered effective in all three domains are marked bold in the text and tables. In light of the large number of elements that have been examined in the reviews and our focus on similarities across domains, the tables only include aspects that have been examined in at least two domains.

3.4 *Program focus or goal*

As shown in Table S2, several strong reviews in the nutrition and sexuality domains concluded that programs with a *specific behavioral focus* (e.g., fruit consumption, condom use) are more effective than programs that discuss general nutritional or sexuality issues; supplementary, a comparable statement in one weak substance abuse review was that programs should be tailored to specific substances [52].

The issue of *abstinence goals* has been addressed by strong reviews in the sexuality and substance abuse domains. Not one sexuality review stated positive conclusions about the effectiveness of abstinence-only programs, which portray abstinence from sex as the only or very best prevention option and usually do not discuss contraception, and one even reported negative effects [63]. In contrast, one strong sexuality review [61] reported positive effects of programs that do discuss contraception (abstinence-plus or safer sex programs). Comparatively, in the substance abuse domain, one strong review cautioned that the goal of harm reduction or prevention of abuse may be more effective than a goal of abstinence or delayed use, at least for youth who already use [35].

3.5 Program development

In the substance abuse, sexuality and nutrition domains there is broad consensus among strong reviews that *theory-based* programs produce better effects than non-theory-based programs [see Table S3], although some reviews did not find obvious differences [42], only found a contribution of theory in univariate and not multivariate analysis [56] or stated that the exact contribution of using theory is unclear [26]. With respect to specific theories, strong reviews in the substance abuse [36,40] and nutrition [77] domains made special reference to Bandura's *social cognitive theory*; supplementary, a weak review in the sexuality domain stated that the evidence for using this theory is tentative but not yet convincing [70].

Addressing behavioral determinants was reported to be an effective element by a strong nutrition review [77] and a moderate sexuality review [66]; supplementary, weak reviews in the substance abuse domain had the same conclusion [52-54]. Three other characteristics of program development were stated to be important for enhancing effects, but each only in one or two domains: *needs assessment* among the target group, *participant involvement* in program planning and implementation, and *pretesting*. The evidence for the second element involved only a supplementary weak review in the substance abuse domain [54], and the evidence for the third was mixed, as a meta-analysis in the sexuality domain reported that stated use of pretesting was not related to the effect size for condom use [56].

The issue of *tailoring interventions to the culture* of the target group was addressed by several strong or moderate reviews in the substance abuse domain and a moderate review in the sexuality domain. The sexuality review had positive conclusions [66], as did most substance abuse reviews [33,41,47]. However, the substance abuse review with the strictest criteria reported this issue to be unclear because no high-quality study had compared culture-specific interventions with standardized interventions [31]. In the nutrition domain, this issue was only addressed by a supplementary weak review, which stated the issue to be unclear and in need of further research [79]. *Tailoring to cognitive ability or age* has been examined by three strong reviews, which cover all three domains. The sexuality [65] and nutrition [77] reviews reported favorable results, but again, the review in the substance abuse domain applied the strictest criteria and reported inconclusive results because of a lack of high-quality comparison studies [31].

3.6 Program content

Table S4 presents the results for elements of program content. Since many elements were mentioned in the reviews, we included headings to indicate that there may be some similarity between elements.

3.6.1 Knowledge, risk, attitude.

Health education programs in all domains usually include information about health consequences and prevention methods. In all domains a *knowledge-only approach* was reported to have no effect on behavior, but in the sexuality domain this involved only a supplementary weak review [29]. Some authors commented that this approach has hardly been tested rigorously [31] or only with traditional, non-engaging methods [51]. In the sexuality domain, a strong and a moderate review stated that accurate, *factual information* is an element of effective interventions [60,66]; supplementary, this was also reported in a weak substance abuse review [54]. The results of two strong sexuality reviews for *enhancing perceived risk* were mixed [58,65]; in the substance abuse domain, the related issue of fear arousal was reported to be ineffective by a moderate review [47]. Several other elements were each addressed in only one domain and are therefore not included in Table S4 nor further discussed here.

3.6.2 Social influences.

Social influences have been addressed in all domains, especially in the substance abuse domain where the social influences approach has been widely prevalent for decades. In all domains, strong reviews stated that this approach is effective, although reservations were reported in one tobacco review [31] as the largest and most rigorous study found no evidence of a sustained effect on smoking prevalence. While the social influence approach entails several components [see 51], two components have received most attention in the review literature: reinforcing or changing social norms (e.g., correcting overestimations of peer smoking) and training in recognizing and resisting peer, media and other influences (e.g., learning to negotiate safer sex). In all domains, strong reviews reported the first component, addressing **social norms**, as an effective element. In the nutrition domain attention to norms does not seem to take the form of normative feedback but rather of building normative support for desired changes and for creating a more supportive school or community environment [77]. The second component, *resistance skills training*, was not addressed in nutrition reviews and had inconsistent results in other domains. There is some evidence that this element may only be effective in conjunction with normative education or with a rationale or motivation for refusal and may even be counterproductive when used alone [28]. This latter review [28] reported that resistance skills training is only effective if it is behavior-specific.

3.6.3 Skills

In all domains, training of *skills* was generally reported to be effective. Although the types of skills were not always specified or seemed to vary, the following similarities were observed. In the nutrition and sexuality domains, some strong reviews mentioned *domain-bound practical skills*, such as food preparation or condom use skills.

In each domain, *cognitive-behavioral programs* have been found effective in one or two strong reviews. Although not all authors used the same terms or were clear about what this approach entails exactly, we included this element to refer to statements about the importance of addressing both motivations and cognitive and behavioral skills. In the nutrition domain, one strong review stated that effective behaviorally focused curricula address cognitive, affective and behavioral aspects [77]. In their meta-analysis of tobacco outcomes of psychosocial programs, Hwang and colleagues [32] used a narrower definition of cognitive-behavioral programs. They distinguished social influence, cognitive behavioral, and life skills modalities. Cognitive-behavioral programs were those that included the social influence approach “plus at least two cognitive skills such as problem solving, decision making, assertiveness, self-control, and/or other coping skills. Life skills programs included the defined aspects of the social influence and cognitive-behavioral modality programs plus at least one affective skill such as self-confidence, values clarification, and/or generic social skills”.

Life skills training can be regarded as a specific type of cognitive-behavioral program, one that addresses self-management and social skills (decision-making, anxiety management, communication, assertiveness). Strong reviews in the substance abuse domain reported that this training enhances the effects of a social influence approach on tobacco and alcohol use. Life skills training has only been tested in the substance use domain, and only in combination with a social influence approach. However, in the sexuality domain some strong and moderate reviews seem to refer to similar skills when stating the importance of coping, communication, and negotiation skills [58,60,62,65,66, not reported in Table S4].

3.7 Program methods

Statements about effective methods were relatively scarce in the reviews [see Table S5]. In the substance abuse domain four strong reviews consistently reported *interactive methods* to be effective; supplementary, weak reviews in the sexuality and nutrition domains mentioned specific examples of interactive methods (discussion and role-play). Tobler and colleagues [40,41], who provided the strongest evidence for interactive methods in large meta-analyses in the substance abuse domain, stated that interaction should be between students, not so much between student and teacher.

In both the nutrition and sexuality domains, *having students personalize information* was identified as an effective element in one strong or moderate review. Four other elements of program methods had evidence from one or two strong reviews in one domain, but had been examined by only weak reviews in another domain. The re-

sults for these elements were consistent across these domains (the domain named first in parentheses had evidence from a strong review): a *traditional, didactic style* ('lecture') is reported to be ineffective (nutrition, substance abuse), whereas it is effective to use *multiple channels* (sexuality, multiple behaviors), *active, experiential methods* such as experiments and taste testing (nutrition, substance abuse), and *cognitive-behavioral skills training* (sexuality, substance abuse). According to one review [47], the latter training consists of: instruction and demonstration, behavioral rehearsal with role play, feedback on each student's performance, social reinforcement, and extended practice through behavioral 'homework' assignments. Several other methods have only been reported in a single domain and are thus not included in Table S5 nor discussed here (e.g., modeling, goal-setting).

3.8 Program facilitator

The impact of the type of program facilitator on program effectiveness has had most attention in the domains of substance abuse and sexuality [see Table S6]. Especially in the substance abuse domain, many types of facilitators have been examined (not shown in Table S6).

Only *peer leaders* and *teachers* have been examined in more than one domain. The evidence conflicted between the nutrition and sexuality domains, as a strong nutrition review reported favorable results for the use of peer leaders [72], whereas three strong sexuality reviews did not find evidence for a differential impact of the type of facilitator [55,56,59]. In the substance abuse domain, the results of strong and moderate reviews were mixed. Both peer leaders [45,47] and teachers [47] have been involved in effective programs and several meta-analyses and reviews that analyzed the contribution of the type of facilitator to ES did not find overall significant differences between these facilitator types [34,35,41]; however, some reported results favoring peers over teachers, either overall [43] or for a particular intervention type [34,40] or measurement period [36]. A meta-analysis of studies comparing implementation of the same program by *peers versus teachers* reported that peers have shown better effects, but only in the short term and not at 1- or 2-year follow-up [37]. However, in light of variations in effects and lack of high-quality studies, this review did not conclude that implementation by peers is better. Also, a recent tobacco review [31] stated that not one comparison study was of high quality. Our overall conclusion for the substance abuse domain is that there are some indications that peers may have better effects than teachers, but the evidence is yet inconclusive and not one type of facilitator has generally proven to be more effective than another. There was one element of the facilitator that was consistently reported by strong reviews in all domains to have a positive contribution to effectiveness: ***facilitator training***.

3.9 Program components

Table S7 presents the results of reviews with respect to program components. The term ‘component’ is used here to refer to different approaches to behavior change (education, environmental change) or the inclusion of different settings (school, family, community). We paid extra attention to reviews with a specific focus on schools, and we were especially interested in the added value of school-wide, family and community components in addition to the usual classroom education approach.

Strong reviews in all domains were consistently positive about the effectiveness of *programs with multiple components*, except for one sexuality review with null results but unclear operationalization [57] and one tobacco review that reported positive effects only for the long term [32]. The element of multiple components includes statements about the (better) effects of multi-component programs in general, about specific multi-component programs and about combinations of specific components.

Drawing overall conclusions about specific components is more difficult because reviews varied as to the specificity of their statement, the operationalization of components, and the criteria for assessing effectiveness (e.g., are direct comparisons necessary?). For instance, several reviews distinguished family from community components, whereas others included all family, media and community mobilization activities under the heading of community components. In light of these differences between reviews, the conclusions below about specific components should be regarded as tentative.

Programs with *school-wide change and family or community components* have been reported by strong reviews to be effective, but have only been examined in the substance abuse and nutrition domains. Strong reviews in the substance abuse and sexuality domains made positive statements about *community interventions*, and these were supplemented by a weak review in the nutrition domain; however, the strong alcohol review by Foxcroft and colleagues [33] referred more to hypotheses about cost-effectiveness than to actual evidence. The added value of *community adjuncts to classroom interventions* is convincing in the nutrition domain but was not examined in the sexuality domain. In the substance domain, several strong reviews and meta-analyses had positive conclusions, but their operationalizations or statements were general and included also family activities [32,36] or life skills modalities [31].

The evidence for *school-wide activities* is consistently positive in the nutrition domain (foodservice); supplementary weak reviews in the sexuality domain were also consistently positive (school health clinic with family planning services), but weak reviews in the substance abuse domain were not (school drug policies).

There is some evidence from strong reviews in all domains that including *parents or families* is effective; however, in the substance abuse domain this may apply only to high-risk youth, and in the nutrition domain only to elementary-aged children [77].

In the nutrition domain one strong review examined policies that impact on accessibility of products. *Price regulation* has been found effective in this domain

[78]; this was also reported for tobacco and alcohol by two weak substance abuse reviews [52,54].

All in all, there is some evidence in all domains that multi-component programs with school-wide, community and/or family components can be effective or can be more effective than curricular interventions, but the added value of such components is unclear.

3.10 Program intensity

Table S8 lists the review results with respect to program intensity and duration. It should be noted that it is not always clear what authors mean when using these terms. The more narrowly defined term of *number of sessions/hours* was addressed by strong reviews in all domains. Only reviews in the nutrition domain consistently reported a positive association with outcomes ('more is better') [74,76,77]. In the sexuality domain, the results appear to differ per type of review: three narrative reviews reported such an association [55,60,62], whereas two meta-analyses did not [58,59]. In the substance abuse domain, one review and one meta-analysis did not find clear evidence that more is better [35,41], whereas another meta-analysis did, but only for interactive programs and not for non-interactive programs [40].

Several strong or moderate reviews identified a specific minimum number of sessions/hours required for producing effects, and the numbers were comparable across domains: 8 hours for sexuality programs [60] and 10 sessions for substance abuse [48] and nutrition programs [74], although one nutrition review considered 10-15 sessions insufficient [77]. These numbers are in accordance with effects reported in one strong and one moderate substance abuse review about specific programs [38,47], but another review stated that recent substance abuse studies tend to recommend fewer sessions, specifically 4, 5 or 8 [35]. However, in light of the results already discussed, the evidence that a larger number of sessions enhances effects is only consistent in the nutrition domain. The same conclusion can be reached for the less well-described terms of *intensity* and *duration*.

The issue of *booster sessions* has mainly been examined in the substance abuse domain, except for one strong sexuality review with positive results [65]. In the substance abuse domain, the results of strong reviews were mixed. Of two strong tobacco-specific reviews, one concluded that boosters enhance long-term effects [44], but our recalculations of the presented data led us to question this conclusion; the second review had unclear results [43]. One broader substance abuse review reported benefits of boosters for behavior maintenance [35], while another did not find conclusive evidence and stated that boosters may increase effects for some programs but not for others [36]. All in all, this issue remains inconclusive.

4. DISCUSSION

4.1 *Similarities across domains*

This review of reviews examined effective elements of adolescent health promotion programs in three behavioral domains – substance abuse, sexual behavior and nutrition. We specifically focused on similarities across these domains, and indeed, we identified many similarities. The results are discussed here in light of the two types of analysis that have been explained in the Methods section: an interpretation-based method and a rule-based method. Based on our interpretation-based examination of the evidence that is currently available from strong and moderate reviews, five elements were identified to be effective in all domains. These five elements have evidence from at least one strong review in each domain:

- a) use of theory, with specific reference to social cognitive theory
- b) addressing social influences, especially social norms
- c) addressing cognitive-behavioral skills
- d) training of facilitators
- e) including multiple components.

When using the rule-based method of analysis, the results are similar: all five elements have at least tentative evidence in all domains. Elements b, c and d even have sufficient evidence in all domains; elements a and e have tentative evidence in one or two domains due to conflicting results between strong reviews in these domains (positive versus null or unclear results). Using the rule-based method, no other elements were identified as having sufficient evidence in all three domains, but two additional elements had at least tentative evidence for a positive contribution to effectiveness in each domain:

- f) parent involvement
- g) a larger number of sessions.

These two elements were not identified as similar across domains with the interpretation-based method of analysis, since we found the evidence in at least one domain to be inconclusive due to conflicting results between strong reviews; in the rule-based method such conflict leads to the conclusion that the evidence is tentative. The different results of the two methods of analysis for these two elements can thus be explained by the different approaches to handling conflicting results.

In addition to the above elements, which had evidence from strong or moderate reviews in each of the three domains, several other elements also tended to have similar results across the three domains, but their evidence involved only weak reviews in one or two domains. Although weak reviews were not included in the analysis, they were used for exploring whether there is any indication that a particular element might be effective in a particular domain. The following elements had similar results across all domains; domains with strong or moderate reviews are given between parentheses:

- h) a focus on specific behavior (sexuality, nutrition)
- i) addressing behavioral determinants (sexuality, nutrition)
- j) a knowledge-only approach (ineffective element; substance abuse, nutrition)
- k) use of interactive methods (substance abuse).

In addition to the above elements, the results for many other elements were comparable across at least two of the three domains. We did not find one element for which the results indicated opposing directions of influence between domains (e.g., a positive contribution to effectiveness in one domain and a negative contribution in another domain). In cases where the results were not similar across domains, this was usually because results in one or more domains were unclear or indicated null findings, whereas those in other domains indicated a positive contribution to effectiveness.

The results of the present review are fairly similar to those of other systematic reviews of reviews that examined the domains of substance abuse and sexuality separately and that included only high-quality reviews (Lister-Sharp, Chapman, Stewart-Brown, & Sowden, 1999; Thomas et al., 2005), suggesting that the results for these domains are robust. This review adds rigor and specificity to the general observation in several reviews that effective elements in the domains of substance abuse and sexuality appear to be similar (e.g. Botvin et al., 1995; Kirby, 2002b; Schaalma et al., 2004; Summerfield, 2002; Thomas et al., 2005) and extends this observation to also include the nutrition domain. In contrast to the present review, these reviews did not examine the issue of similarity systematically or in detail.

Perhaps more importantly, our results are largely comparable to, and in some cases more specific than, those of a review of reviews that specifically focused on similarities across multiple domains (Nation et al., 2003). That review examined partly different domains (substance abuse, risky sexual behavior, school failure, and juvenile delinquency and violence), included a smaller and different set of reviews (35 narrative reviews that explicitly discussed common features of effective programs) and used a somewhat different review methodology (determining the percentage of reviews that mentioned an element as consistently effective). In that review (Nation et al., 2003), nine elements of effective programs were identified, which were claimed to reflect general principles that transcend specific content areas. Seven of these elements coincide with the ones identified by us, although some tend to be formulated in more general terms than ours. These seven elements and, between brackets, the corresponding letters from our list, are: theory-driven [a]; socio-culturally relevant (address cultural norms and beliefs) [b,i]; varying teaching methods (skills-based component, active and interactive format) [c,k]; providing opportunities for positive relationships (parent-child communication, peer influences) [b,f]; well-trained staff [d]; comprehensive (multi-modal, multiple settings) [e]; and sufficient dosage [g]. Two of the elements they identified are not represented in our own set of eleven elements: appropriate timing and inclusion of outcome evaluation. The issue of outcome evaluation was not considered relevant for the present review, as it is an aspect of studies rather than programs. The issue of appropriate timing has to do with tuning interventions to student characteristics such as age, cognitive and social development and experience with the risk behavior. This issue is generally recommended in health promotion theory (Bartholomew, Parcel, Kok, & Gottlieb, 2006; Green & Kreuter, 1999), and indeed, tailoring to age was reported to be effective by strong sexuality and nutrition reviews in this paper (Contento et al., 1995; Pedlow & Carey, 2004). However, we did not include it in our empirical-

ly-based list of effective elements because in the substance abuse domain it was reported to be unclear due to absence of high-quality comparison studies (Thomas & Perera, 2006). One element from our own list, a focus on specific behavior, is not represented in the list from the other review (Nation et al., 2003). Unfortunately, due to the limited reporting of results in the Nation et al. review, we cannot examine the causes for this difference. Possibly, the issue of behavioral focus may only be relevant for certain domains or may have been overlooked in certain domains.

4.2 *Implications for practice*

Researchers and practitioners in the three domains can use the effective elements identified in this review, and especially the ones that are similar across domains, as guidelines for developing and improving their adolescent health promotion programs. They can also look beyond the boundaries of their own domain to generate ideas for programs or research from results in other domains.

The fact that another multiple-domain review (Nation et al., 2003) found comparable effective elements while examining partly different domains (also school failure, juvenile delinquency and violence) suggests that the effective elements may transcend broadly to other content areas. In fact, the effective elements pertaining to program development (use of theory, addressing determinants) appear to be applicable universally, as they are general recommendations from health promotion planning models and quality assurance procedures such as PRECEDE-PROCEED (Green & Kreuter, 1999), intervention mapping (Bartholomew et al., 2006) and Preffi (Molleman, Peters, Hosman, Kok, & Oosterveld, 2006).

The finding that several effective elements are comparable across the three targeted domains indicates that integrative programs can address these domains with the same program characteristics. This is important in light of the recent interest in multiple health behavior research and its potential implications for integrative interventions (Noar, Chabot, & Zimmerman, 2008; J. O. Prochaska, 2008; J. J. Prochaska et al., 2008). The results will be used for guiding the development of our own integrative program. The effective elements pertaining to program content - address not only information, but also social influences and cognitive-behavioral skills - fit well with those of a previous review that assessed similarities between behavioral determinants across the same three domains (Peters, Wiefferink, et al., 2009). In that review the following determinants were found to be important for all domains: attitudinal beliefs about immediate gratification and social advantages, social norms, modeling behavior and resistance skills. Together, both that review and the present one provide sufficiently valid input for the development of an integrative program that addresses all three domains.

4.3 *Limitations*

Given our broad focus on several health-related behaviors and the already extensive body of knowledge in each domain, we applied a review-of-reviews approach, an approach that has gained acceptance in recent years (e.g., Cuijpers, 2002a; Ellis &

Grey, 2004; Lister-Sharp et al., 1999; McBride, 2003; Micucci et al., 2002; Mulvihill & Quigley, 2003; Nation et al., 2003; Poobalan et al., 2008; Thomas et al., 2005). Although the search strategy was comprehensive, it is possible that we may have missed relevant reviews. However, it is not very likely that these reviews would have discussed different sets of primary studies and would have led to different conclusions.

A limitation of the review-of-reviews approach is that it relies on 'second-hand' information and is potentially vulnerable to the interpretive and conceptual biases of previous reviewers (Nation et al., 2003). We attempted to limit these biases as much as possible by using a systematic review methodology, by assessing the quality and relevance of each review and relying on reviews of high to moderate quality, by carefully categorizing the results without generalizing too much, and, in case reviews had differential results, by attempting to examine the causes of the differences. We also attempted to check the results of reviews if sufficient information was provided.

Perhaps we would have identified more similarities across domains if we had combined aspects and findings into broader categories. We used a conservative categorization process and were reluctant to generalize findings, because the operationalization, interpretation or analysis of aspects seemed to differ between reviews or were sometimes unclear.

Two-thirds of the included reviews had a high quality score of 6 or 7. In line with other reviews of reviews (Ellis & Grey, 2004; Micucci et al., 2002; Poobalan et al., 2008) we included only strong and moderate reviews in the analysis. Furthermore, we used two methods for analyzing the results and especially for dealing with conflicting results between reviews: one method focused on interpretation of differences and the other set a strict rule. The conclusions based on these two methods were fairly similar. Weak reviews were excluded from the analysis but were used in a speculative way: for elements that had evidence from strong or moderate reviews in at least one domain, the results of weak reviews in the other domains were used to give any indication or suggestion of effectiveness in these other domains.

The methodological aspects assessed in reviews most often pertained to study design, appropriateness of allocation procedures, comparability of groups, validity of assessment and attrition, but only a few reviewers examined additional aspects such as quality of implementation. The strictness of inclusion criteria and assessment of methodological quality varied widely, even among high-quality reviews. Although meta-analyses in several domains reported that effect sizes did not vary with the design or quality of studies (DiCenso, Guyatt, Willan, & Griffith, 2002; Gottfredson & Wilson, 2003; Johnson, Carey, Marsh, Levin, & Scott-Sheldon, 2003; Knai, Pomerleau, Lock, & McKee, 2006; Mullen, Ramírez, Strouse, Hedges, & Sogolow, 2002; Tobler et al., 2000), reviews with the strictest methodological criteria (e.g., accepting only high-quality comparison studies) generally appeared to have more cautious conclusions than reviews with less strict criteria. Reporting the specific criteria applied by reviewers appears to be a valuable addition to the Quality Assessment Tool for Reviews. For reviews of primary studies, the Cochrane Collaboration Handbook (Higgins & Green, 2008) and others (Jackson & Waters, 2005) rec-

ommend using the Quality Assessment Tool for Quantitative Studies, which is also developed by the Effective Public Health Practice Project, Canada.

5. CONCLUSIONS

A multitude of effective elements of school health promotion programs has been identified in literature reviews in the domains of substance abuse, sexuality and nutrition. Many effective elements are similar across at least two domains. Based on strong reviews in all three domains, five elements were found to be similar across the three domains: use of theory; addressing social influences, especially social norms; addressing cognitive-behavioral skills; training of facilitators; and including multiple components. Two additional elements had at least tentative evidence of effectiveness in all domains when using a rule-based method of analysis but had inconclusive evidence in at least one domain when using an interpretation-based method of analysis: parent involvement and a larger number of sessions. For four additional elements, the results were comparable across the three domains but they are more speculative, as in one or two domains these elements had only been examined by weak reviews. Three of these elements have a positive contribution to effectiveness (specific behavioral focus; addressing determinants; interactive methods), whereas the fourth (knowledge-only approach) was considered ineffective. The results suggest that an integrative program that addresses the three domains seems feasible and could be efficient. The five elements with evidence from strong reviews in each domain are likely candidates to include in such a program.

Table S1. Characteristics of included reviews, categorized by behavior focus

Reference ^a (quality)	Behavior focus ^b	Time span ^c	# Studies (# relevant) ^d	Age range ^e	Methodological inclusion criteria ^f
Multiple domains					
26 Harden 1999 (7)	Peer programs. Tob, Sex, other	Up to 1998	49, 12 sound (5)	11-24 yrs	Prospective controlled design, equivalent groups, reporting of all data for all groups.
27 Lister-Sharp 1999 (7)	Health promoting school approach. Tob, Sex, Nut, other	No limits (incl =1979-1998)	12 (6)	5-16 yrs	One-group pre-post or controlled design, health-related outcomes
28 Herrmann 1997 (6)	Refusal programs. Tob, Alc, drugs, Sex	1974-1994	33	Child, Adol	None stated
29 Botvin 1995 (2)	Tob, Alc, drugs, Sex (aids, std, pregnancy)	NR	141 r	Child, Adol	None stated
30 Jason 2002 (1)	Tob, Alc, drugs, Sex, Nut, other	NR	50 r	Child, Adol	None stated
Substance abuse					
31 Thomas 2006 (7)	Tob	1966-2005	94 (54 in analysis, 23 high-quality)	Child (5-12 yrs), Adol (13-18 yrs)	RCT, tobacco use measure for baseline nonusers, minimal follow-up of 6 months after program end. Additional quality rating.
32 Hwang 2004 (7)	Psychosocial programs. Tob	1978-1997	75	Grade 6-12	Control group had no psychosocial program, pre-post measures
33 Foxcroft 2003 (7)	Alc	Up to 2001	56	< 25 yrs	RCT, matched pre-post or interrupted time series design; measures of alcohol use or related problems. Additional quality rating.
34 Gottfredson 2003 (7)	Alc, drugs (not Tob-only)	NR	94	School-age	Control group with no or minimal intervention, behavioral outcome

Reference ^a (quality)	Behavior focus ^b	Time span ^c	# Studies (# relevant) ^d	Age range ^e	Methodological inclusion criteria ^f
35 McBride 2003 (7)	Tob, Alc, drugs	1997-2001. Reviews 1990-2001	5 primary studies, 11 reviews	School-age	Primary studies: (quasi-) experimental design, pre-post measures of behavior, positive effects, discussed methodological issues. Reviews: systematic, applied methodological inclusion criteria
36 Cuijpers 2002 (7)	Tob, Alc, drugs	NR, incl = mostly 1990s	27 (27)	School-age	Meta-analyses comparing program types. Mediator studies. Studies comparing program components.
37 Cuijpers 2002 (7)	Peer-led vs adult-led. Tob, Alc, drugs	NR, incl = 1981-1995	12 (12)	School-age 11-18 yrs	Prospective studies comparing peer- and adult-led program.
38 Sussman 2002 (7)	Towards No Drug Abuse. Tob, Alc, drugs, violence	1994-1998	3 (3)	Grade 10-11	None stated, all studies were RCT
39 Blake 2001 (7)	Tob, Alc, drugs	1980-2000	32 (11)	Adol girls (primary & secondary age)	Female-specific results, no design criteria
40 Tobler 2000 (7)	Tob, Alc, drugs	1978-1998	144, 93 high-quality	School-age	Controlled design, pre-post drug-use measures. Additional criteria for high quality subsample.
41 Tobler 1997 (7)	Tob, Alc, drugs	1978-1990	90, 56 high-quality	Grade 5-12	Controlled study with pre-post drug-use measures. Additional criteria for high quality subsample.
42 Foxcroft 1997 (7)	Alc	1966-1995	33	8-25 yrs	(Quasi-)experimental, pre-post measures of alcohol use or related incidents. Additional quality rating.
43 Rooney 1996 (7)	Social or peer-type programs. Tob	1974-1991	90	Grade 6-12	Control group, measure of tobacco use
44 Skara 2003 (6)	Tob, Alc, drugs	1966-2002	25 (25)	< 21 yrs	(Quasi-)experimental design; at least 2-yr follow-up; measure of tobacco use incidence or prevalence

Reference ^a (quality)	Behavior focus ^b	Time span ^c	# Studies (# relevant) ^d	Age range ^e	Methodological inclusion criteria ^f
45 Posovac 1999 (6)	Peer-based programs. Tob, other	1978-1997	22 Tob programs	Average age 12.5 yrs	Control group, enough information to calculate effect sizes
46 Werch 2001 (5)	Stage-based programs. Alc	About 1990-2000	5 (3)	NR. Included = Grades 5-9	None stated, all studies were RCT
47 Botvin 2000 (5)	Life Skills Training. Tob, Alc, drugs	About 1980-2000	About 9 (about 9)	Middle/junior high school	None stated
48 Dusenbury 1997 (5)	Alc, drugs (incl Tob but not Tob-only)	NR	23 (21)	Primary and secondary age	Pre-post control group design, substance use outcome measure, peer reviewed
49 Hittner 1998 (4)	Alc	1980-1996	36 (31)	Child, Young Adol	Alcohol misuse outcome
50 Flay 2000 (3)	Classroom plus additional component. Tob, Alc, drugs	NR, incl = 1980s and 90s	18 (13)	Primary and secondary age	None stated.
51 Donaldson 1996 (3)	Tob, Alc, drugs	About 1976-1996	78 r	School-age	None stated
52 Montoya 2003 (2)	Tob, Alc, marijuana	About 1980-2003	81 r	Adol	None stated
53 Botvin 2000 (2)	Tob, Alc, drugs	About 1980-2000	70 r	Primary & secondary age	None stated
54 Paglia 1997 (2)	Tob, Alc, illicit drugs	NR	176 r	Youth	None stated
Sexuality					
55 Robin 2004 (7)	Sex: hiv, std, pregnancy	1990-1999	20 (about 13)	Youth and Adol	RCT or quasi-experimental; control for pretest differences; cell size at least 16; follow-up of at least 4 weeks; attrition less than 40%; behavioral or biological outcome
56 Johnson 2003 (7)	Programs with HIV content. Sex: hiv	Up to 2000	44	Adol 11-18 yrs	RCT or quasi-experimental with rigorous controls; outcome measures relevant to sexual risk behavior; sufficient information to calculate effect sizes.

Reference ^a (quality)	Behavior focus ^b	Time span ^c	# Studies (# relevant) ^d	Age range ^e	Methodological inclusion criteria ^f
57 DiCenso 2002 (7)	Sex: pregnancy	1970-2000	26 (13)	Adol 11-18 yrs	RCT; measured initiation of intercourse, use of birth control or pregnancy. Additional quality rating.
58 Mullen 2002 (7)	Sex: hiv	1992-1998	20 (6)	Adol 13-19 yrs	RCT or controlled design with pretest inequivalence controlled; measure behavior or biologic indicator
59 Silva 2002 (7)	Sex	1985-2000	12 (12)	Adol	RCT or quasi-experimental; equivalent no-intervention control group; measure of abstinent behavior; peer-reviewed
60 Yamada 1999 (7)	Sex: std	Up to Sep 1998 (incl= 1992-1998)	24 (about 10)	10-19 yrs	RCT or controlled; sample representative of general population; behavioral outcome. Additional quality rating
61 Franklin 1997 (7)	Sex: pregnancy	Up to 1995	32 (about 14)	Adol 11-20 yrs	Behavioral measure, peer-reviewed. No design criteria for inclusion. Additional quality rating.
62 Kim 1997 (7)	Sex: hiv	1983-1995	40, 4 in meta (19)	Adol 10-18 yrs	No design criteria for inclusion in review. Additional quality rating. RCT criterion for inclusion in meta-analysis.
63 Oakley 1995 (7)	Sex: sexual health	1982-1994	12 sound-65	0-19 yrs	Sound study: RCT or equivalent control group; pre- and post data; reporting of all outcomes
64 Bennett 2005 (6)	Sex: pregnancy	1980-2002	16 (16)	Secondary school	RCT; outcomes sexual behavior, contraceptive knowledge or use, or pregnancy
65 Pedlow 2004 (6)	Sex	Up to Feb 2003	24 (10)	11-18 yrs	RCT; behavioral outcome measure
66 Kirby 2002 (5)	Sex: hiv, std, pregnancy, abstinence	1980 – 2001	73	12-18 yrs	(Quasi-)experimental; sample size at least 100 in groups combined; measure of behavior or behavioral outcome.
67 Song 2000 (5)	Sex	1960-1997	67 (67)	Adol Grade 5-12	Outcome measure of knowledge about sexuality

Reference ^a (quality)	Behavior focus ^b	Time span ^c	# Studies (# relevant) ^d	Age range ^e	Methodological inclusion criteria ^f
68 Thomas 2000 (3)	Programs with focus on abstinence. Sex: std, pregnancy	NR, incl = 1990-1995	9 (8)	Adol	None stated
69 Milburn 1995 (3)	Peer education programs. Sex.	NR	51 r	Young people	None stated
70 Christopher 1995 (2)	Sex: pregnancy	NR, incl = 1981-1994	About 13	Adol	None stated.
71 Jacobs 1995 (2)	Sex: pregnancy, std	NR, incl = 1982-1991	6 (5)	Adol	None stated.
Nutrition					
72 Knai 2006 (7)	Nut: fruit, vegetables	Up to Apr 2004	15 (4)	Child, Adol 5-18 yrs	Controlled; follow-up at least 3 months; behavioral outcome. Additional quality rating.
73 Shilts 2004 (7)	Programs including goal-setting. Nut, exercise	1977- Dec 2003	28 (1)	All ages; sub sample Adol 12-19 yrs	Experimental, quasi-experimental or pre-experimental (no cross-sectional); cell size greater than 5. Additional quality rating.
74 Thomas 2004 (7)	Nut, exercise. Subset of dietary programs.	1985-Aug 2003	57 (4)	6-18 yrs	Prospective controlled studies; behavioral outcome. Additional quality rating.
75 Ammerman 2002 (7)	Nut: fat, fruit, vegetables	1975-Aug 1999	92 (1)	All ages; sub sample school-based	RCT or quasi-experimental; behavioral outcome; sample size at least 40; diet freely chosen by participants. Additional quality rating.
76 Ciliska 1999 (7)	Nut: fruit, vegetables	Up to Aug 1998	15 (1)	> 4 yrs	Prospective study with comparison group; information on process or outcome evaluation. Additional quality rating.

Reference ^a (quality)	Behavior focus ^b	Time span ^c	# Studies (# relevant) ^d	Age range ^e	Methodological inclusion criteria ^f
77 Contento 1995 (7)	Nut	1980-1995	43 (about 20)	School-age	Random or quasi-experimental design; evidence of instrument reliability and validity
78 French 2003 (6)	Environmental programs. Nut: fruit, vegetables	NR, incl = 1993-2003	11 (2)	Primary and second- ary age	No criteria stated. All included studies are controlled, about half randomized
79 Hoelscher 2002 (3)	Nut	1994-2000	17 (about 5)	Adol 11-18 yrs	No criteria stated
80 Lytle 1995 (3)	Nut	1980-1995?	85 r	Child, Adol	Controlled; behavioral outcome; peer-reviewed

Note. Within behavior categories, reviews are ordered by quality rating and publication year.

^a Reference number, first author, publication year, quality score in parentheses (0-3=weak, 4-5=moderate, 6-7=strong). ^b Behavior focus: Alc=alcohol, Nut=nutrition, Sex=sexuality, Tob=tobacco, std=sexually transmitted disease. If applicable, a more specific focus is recorded for sexuality and nutrition reviews (e.g., pregnancy, fat consumption). ^c Time span used in search strategy. NR=not reported, Incl=actual time span of included studies, recorded for reviews that did not report time span of search strategy. ^d For reviews that did not report the number of included primary studies the total number of references is given, indicated by r. Numbers between parentheses indicate the number of primary studies meeting relevance criteria of the present review (target behaviors, secondary school-age, school-based educational intervention). This number was only examined for reviews that provided sufficient information about primary studies. ^e Child=children, Adol=adolescents, yrs=years. ^f RCT=randomized controlled trial.

Table S2. Elements of program focus related to program effectiveness: results of reviews by domain

Element	Multiple domains	Substance abuse	Sexuality	Nutrition
Focus on specific behavior		52 (2): +.	56 (7): +. 66 (5): +.	72 (7): +. 74 (7): +. 76 (7): +. 77 (7): +. 79 (3): +.
Abstinence-plus (AP) vs abstinence-only (AO)		35 (7): +/? 54 (2): +.	59 (7): AP vs AO 0. 61 (7): AP +. 63 (7): AO -. 64 (6): AP vs AO ?. 66 (5): AO ?. 68 (3): AO 0, AP +. 70 (2): AO 0.	

Note. Reviews are indicated by the reference number and, between parentheses, the quality rating (0-3=weak, 4-5=moderate, 6-7=strong). Results are indicated by the following characters: +=positive, -=negative, 0=null, ?=unclear contribution to program effectiveness

Table S3. Elements of program development related to program effectiveness: results of reviews by domain

Element ^a	Multiple domains	Substance abuse	Sexuality	Nutrition
Use of theory	26 (7): ?. 29 (2): +.	36 (7): +. 38 (7): +. 40 (7): +. 42 (7): 0. 47 (5): +. 53 (2): +.	56 (7): +/0. 60 (7): +. 62 (7): +. 66 (5): +. 68 (3): +. 70 (2): +.	76 (7): +. 77 (7): +. 79 (3): +. 80 (3): +.
Social cognitive theory	29 (2): +.	36 (7): +. 40 (7): +. 47 (5): +. 53 (2): +.	70 (2): +/?.	77 (7): +. 79 (3): +. 80 (3): +.
Addressing determinants	29 (2): +.	52 (2): +. 53 (2): +. 54 (2): +.	66 (5): +.	77 (7): +.
Conducting needs assessment	26 (7): +	35 (7): +.		
Students involved in planning, implementation	26 (7): +.	54 (2): +.		
Formative phase, interviews, pretesting		35 (7): +.	56 (7): 0.	
Tailoring to culture, ethnicity		31 (7): ?. 33 (7): +. 41 (7): +. 47 (5): +. 53 (2): +.	66 (5): +. 29 (2): +.	79 (3): ?.
Tailoring to cognitive ability or age		31 (7): ?.	65 (6): +.	77 (7): +. 79 (3): +.

Note. Reviews are indicated by the reference number and, between parentheses, the quality rating (0-3=weak, 4-5=moderate, 6-7=strong). Results are indicated by the following characters: +=positive, -=negative, 0=null, ?=unclear contribution to program effectiveness.^a Elements marked bold are considered by us to have similar results in the domains of substance abuse, sexuality and nutrition.

Table S4. Elements of program content related to program effectiveness: results of reviews by domain

Element ^a	Multiple domains	Substance abuse	Sexuality	Nutrition
Knowledge, risk, attitudes Knowledge only approach	26 (7): 0. 29 (2): 0.	31 (7): 0. 40 (7): 0. 41 (7): 0. 47 (5): 0. 50 (3): 0. 51 (3): 0. 53 (2): 0. 54 (2): 0.	29 (2): 0.	74 (7): 0. 76 (7): 0. 77 (7): 0.
Factual information		54 (2): +.	60 (7): +. 66 (5): +. 70 (2): +.	
Short-term consequences Enhancing perceived risk	29 (2): +.	54 (2): +. Fear arousal: 47 (5): 0. 51 (3): 0. 54 (2): 0.	58 (7): 0. 65 (6): +.	
Social influences Social influences	29 (2): +.	31 (7): 0/+ . 32 (7): +. 35 (7): +. 39 (7): + especially for girls. 40 (7): +. 41 (7): +. 43 (7): +. 50 (3): +. 51 (3): +. 52 (2): +. 53 (2): +.	65 (6): +. 66 (5): +. 68 (3): +.	77 (7): +. 79 (3): + with younger adolescents.
Social norms	29 (2): +.	35 (7): +. 36 (7): +. 50 (3): +. 51 (3): +. 52 (2): +. 54 (2): +.	65 (6): +. 68 (3): +.	77 (7): +.
Resistance skills	26 (7): +. 28 (6): behavior specific +, general 0. 29 (2): +.	35 (7): ?. 36 (7): 0. 38 (7): not suited for alternative high school students. 28 (6): for tob + when embedded, for alc 0/+ when embedded. 47 (5): +. 51 (3): 0/+ . 50 (3): 0/+ . 52 (2): +. 54 (2): 0.	28 (6): 0/-. 66 (5): +. 68 (3): +.	
Skills Skills (unspecified)			61 (7): ?.	77 (7): +. 79 (3): +. 80 (2): +.
Practical domain-bound skills (condom use, food preparation) Cognitive-behavioral program / skills training		32 (7): +. 38 (7): TND program +. 53 (2): +.	55 (7): +. 56 (7): +. 65 (6): ?. 65 (6): +.	72 (7): + food preparation skills. 77 (7): +. 79 (3): +.

Element^a	Multiple domains	Substance abuse	Sexuality	Nutrition
Life skills (self-management, decision-making, social and assertive skills, anxiety manag)	29 (2): +.	31 (7): 0/+; 32 (7): +; 33 (7): 0/+; 36 (7): +; 40 (7): +; 41 (7): +; 47 (5): +; 51 (3): +; 52 (2): +; 53 (2): +, without drug focus; 54 (2): +.		

Note. Reviews are indicated by the reference number and, between parentheses, the quality rating (0-3=weak, 4-5=moderate, 6-7=strong). Results are indicated by the following characters: +=positive, -=negative, 0=null, ?=unclear contribution to program effectiveness.

^a Elements marked bold are considered by us to have similar results in the domains of substance abuse, sexuality and nutrition.

Table S5. Elements of program methods related to program effectiveness: results of reviews by domain

Element	Multiple domains	Substance abuse	Sexuality	Nutrition
Multiple channels or strategies	29 (2): +.		65 (6): +. 66 (5): +.	
Didactic style (lecture)		54 (2): 0.		72 (7): 0.
Interactive (incl. discussion, role play)	28 (6): + discussion. 29 (2): + discussion, role play.	35 (7): +. 36 (7): +. 40 (7): +. 41 (7): +.	68 (3): + role play.	79 (3): + discussion.
(Cognitive)behavioral skills training		51 (3): +. 54 (2): +.	65 (6): +.	
Active, experiential learning		53 (2): +.		72 (7): +. 77 (7): +.
Having students personalize info (e.g., diet self-assessment)		54 (2): +.	66 (5): +.	80 (3): +.
				77 (7): +. 79 (3): +. 80 (3): +.

Note. Reviews are indicated by the reference number and, between parentheses, the quality rating (0-3=weak, 4-5=moderate, 6-7=strong). Results are indicated by the following characters: +=positive, -=negative, 0=null, ?=unclear contribution to program effectiveness.

Table S6. Elements of program facilitator related to program effectiveness: results of reviews by domain

Element ^a	Multiple domains	Substance abuse	Sexuality	Nutrition
Peer leader	26 (7): +/? 28 (6): +. 29 (2): +.	34 (7): +/0. 35 (7): 0. 36 (7): +/0. 40 (7): +/0. 41 (7): 0. 43 (7): +. 45: (6) +/0. 47 (5): +.	56 (7): 0. 59 (7): 0. 68 (3): +. 69 (3): ?.	72 (7): +.
Teacher	28 (6): +.	34 (7): 0. 35 (7): 0. 41 (7): 0. 47 (5): +.	59 (7): 0.	
Peer vs teacher same program	26 (7): +/0.	31 (7): ?. 37 (7): +/0.	55 (7): 0/?.	
Trained facilitator		35 (7): +. 47 (5): +.	55 (7): +. 60 (7): +. 66 (5): +.	72 (7): +. 77 (7): +.

Note. Reviews are indicated by the reference number and, between parentheses, the quality rating (0-3=weak, 4-5=moderate, 6-7=strong). Results are indicated by the following characters: +=positive, -=negative, 0=null, ?=unclear contribution to program effectiveness.

^a Elements marked bold are considered by us to have similar results in the domains of substance abuse, sexuality and nutrition.

Table S7. Elements of program components related to program effectiveness: results of reviews by domain

Element ^a	Multiple domains	Substance abuse	Sexuality	Nutrition
Multi-component programs	27 (7 S): +. 29 (2 S): +.	31 (7 S): 0/+. 32 (7 S): ?. 36 (7 S): +. 40 (7 S): +. 50 (3 S): 0/?. 51 (3): +. 54 (2): +.	56 (7): +. 57 (7): 0. 59 (7 S): +. 61 (7): +. 65 (6): +. 66 (5): +. 68 (3): +. 70 (2): +/?. 71 (2 S): +.	72 (7): +. 74 (7): +. 76 (7): +. 77 (7): +. 78 (6 S): +/0. 79 (3): +. 80 (3): +.
School wide change and family or community component		40 (7 S): +.		77 (7): +. 78 (6 S): +/0. 79 (3): +. 80 (3): +.
Community interventions		33 (7): +/?. 54 (2): +.	61 (7): +. 66 (5): +.	80 (3): +.
Community component additional to school	29 (2 S): +.	31 (7 S): 0/+. 32 (7 S): ?. 36 (7 S): +. 50 (3 S): 0/?. 51 (3): +. 54 (2): +.		72 (7): +. 77 (7): +.
School-wide activities		50 (3 S): 0/?. 52 (2): +. 54 (2): ?.	70 (2): +/0. 71 (2 S): +.	72 (7): +. 74 (7): +. 78 (6 S): +/0. 79 (3): +. 80 (3): +.
Parent / family involvement		31 (7 S): 0/+. 33 (7): high-risk +. 36 (7 S): +. 40 (7 S): +. 50 (3 S): general population 0/?, high-risk +. 52 (2): +. 54 (2): general population 0, high-risk 0/+.	59 (7 S): +. 65 (6): +. 68 (3): +.	72 (7): +. 76 (7): +. 74 (7): +. 77 (7): elementary school +, middle/high school 0. 78 (6 S): +/0. 79 (3): +. 80 (3): elementary age +.
Policy: price regulation		52 (2): +. 54 (2): +.		78 (6 S): +.

Note. Reviews are indicated by the reference number and, between parentheses, the quality rating (0-3=weak, 4-5=moderate, 6-7=strong); reviews with a specific focus on schools are indicated by S between parentheses. Results are indicated by the following characters: +=positive, -=negative, 0=null, ?=unclear contribution to program effectiveness.

^a Elements marked bold are considered by us to have similar results in the domains of substance abuse, sexuality and nutrition.

Table S8. Elements of program intensity related to program effectiveness: results of reviews by domain

Element	Multiple domains	Substance abuse	Sexuality	Nutrition
Intensity		31 (7): ?. 36 (7): ?.	55 (7): +/?.	
Duration		34 (7): 0. 54 (2): ?/+.	55 (7): +/? 66 (5): +.	72 (7): + (at least 12 months).
Number of sessions or hours		35 (7): ?. 38 (7): TND 9 sessions 0 for tob and + for alc (only baseline nonusers alc); 12 sessions + for tob and alc (only baseline nonusers alc). 40 (7): +/0. 41 (7): 0. 47 (5): LST 15 sessions +. 48 (5): 10 sessions in first yr and 5 sessions in second yr +.	58 (7): 0. 59 (7): 0. 60 (7): minimum of 8 hours +. 62 (7): +. 55 (7): +/?.	74 (7): + (at least 10 sessions). 76 (7): +. 77 (7): + (10-15 hrs is too little). 79 (3): +. 80 (3): +.
Boosters, continued education	29 (2): +.	35 (7): +. 36 (7): ?. 43 (7): ?. 44 (6): +/0. 47 (5): 15 booster sessions +. 48 (5): +. 51 (3): +. 53 (2): +. 54 (2): +.	65 (6): +.	

Note. Reviews are indicated by the reference number and, between parentheses, the quality rating (0-3=weak, 4-5=moderate, 6-7=strong). Results are indicated by the following characters: +=positive, -=negative, 0=null, ?=unclear contribution to program effectiveness.