Outcome Evaluation of Arizona's

School-Based Smoking Prevention Program –

A Multilevel Study

Inauguraldissertation zur Erlangung des akademischen Grades eines Doktors der Sozialwissenschaften der Universität Mannheim

A Doctoral Dissertation presented to the Department of Psychology, University of Mannheim, Germany

In Fulfillment of Requirements for the Degree
Doctor of Philosophy

vorgelegt von / prepared by
Frederic Malter, Diplom-Psychologe (M.S.)

Universität Mannheim (Germany)
Fakultät für Sozialwissenschaften

In collaboration with the University of Arizona, Tucson, Arizona, USA

Dekan der Fakultät für Sozialwissenschaften der Uni Mannheim (Dean):

Prof. Dr. Berthold Rittberger

1. Gutachter (First Advisor):

Prof. Dr. Emeritus Werner Wittmann, Universität Mannheim

2. Gutachter (Second Advisor):

Prof. Dr. Emeritus Lee Sechrest, University of Arizona, USA

Disputationstermin (Date of Doctoral Defense):

05. Oktober 2009

Acknowledgements

I would like to thank a number of people for their supporting me during the process of writing this dissertation. First off, I would like to thank my advisors Prof. Dr. Werner Wittmann and Prof. Dr. Lee Sechrest who have provided invaluable input and without whom I could not have completed this work. Both have shaped the way I think and feel about social science and its methodology and their open-mindedness will always be something for me to aspire for myself. Likewise, I am grateful to my wife Susanne Auer for her patience and constant willingness to listen to my explications. I am glad she has been with me. A number of people provided input to my thinking at various stages of this endeavor. For this, I would like to thank Dr. Patricia Herman, Dr. Kathy McKnight, Dr. Patrick McKnight, Dr. Manuel Voelkle, and Dr. Michele Walsh.

Finally, being a member of EGAD has given me countless opportunities to broaden my horizon through thoughtful discussions and constant exchange with interesting and smart people. This has influenced my thinking in subtle yet distinctive ways.

Tucson, Arizona, USA, September 2009

Table of Contents

List of Figures

List of Tables

Abbreviations

ADE (Arizona Department of Education), ACJC (Arizona Criminal Justice Commission), ANOVA (Analysis of Variance), AYS (Arizona Youth Survey), BTEP (Bureau of Tobacco Education and Prevention), CDC (Centers for Disease Control and Prevention), ERDU (Evaluation, Research, and Development Unit), FY (Fiscal Year), GRAT (Get Real About Tobacco), HLM (Hierarchical Linear Modeling), ICC (Intra-class correlation), MACTUPP (Maricopa County Tobacco Use Prevention Program), MAR (Missing At Random), MC (Maricopa County), MCAR (Missing Completely At Random), ML (Maximum Likelihood), OLS (Ordinary Least Squares Regression), MNAR (Missing Not At Random), PCA (Principal Component Analysis), REML (Restricted Maximum Likelihood), SAI (Smoking Acquisition Index), SAMHSA (Substance Abuse and Mental Health Service Administration), SES (Socio-economic Status).

Note on language

The stylistically unfortunate linking of multiple nouns, such as 'school-based tobacco prevention program' was unavoidable. Otherwise, unduly convoluted constructions of relative sentences and or chains of multiple genitives would have resulted. Many terms consisting of two nouns were hyphenated, such as 'student-level'. Synonymous to student-level was the term 'level-1', which was used to ease reading and make a quick reference to the statistical multilevel concept. The generic use of masculine nouns was replaced by random substitution of male and female versions.

Reporting of statistical significance

Levels of statistical significance are indicated as follows: '*' refers to $p <= 0.1$, '**' indicates $p <= 0.01$, '***' indicates $p <= 0.001$.

Abstract

Objective. The purpose of this study was to conduct a quantitative outcome evaluation of Arizona's school-based tobacco use prevention efforts and identify school-level factors indicative of a high smoking prevalence. School-based prevention services reach a sizeable share of Arizona's in-school youth every year and cost about $4,000,000 annually.

Method. Data were obtained from two sources, the Arizona Youth Survey for outcome data and process implementation records for intervention data. Because intervention data were only available on the school level, a series of hierarchical linear models was constructed to link outcome data to intervention data, based on Wittmann's general evaluation framework of the five-data-box conceptualization. This study used a quasi-experimental design to evaluate implementation efficacy. Missing data were addressed with a number of methods to minimize bias.

Results. Parental smoking, parental approval of smoking, age and American Indian ethnicity were individual-level risk factors for self-reported smoking and favorable attitudes towards smoking. A school-level intervention index showed a small relationship to school-level means in ecological and HLM models on reported number of friends who smoked and perceived normativity of smoking, but nil effects on self-reported behavior. School-level academic achievement was a strong predictor for self-reported smoking, attitudes and number of friends who smoked. School-level poverty was a redundant predictor once academic achievement was taken into account.

Conclusions. Smoking prevention programming showed very weak effects on outcomes of eighth grade students. If programming is supposed to reach schools with the highest smoking issue, program administrators should target schools with low ratings on academic achievement. More rigorous study designs could shed a light on exact magnitude of effects.

1. Introduction & Rationale

1.1. Purpose of study

Every fiscal year (FY, July-June), the Bureau of Tobacco Education and Prevention (BTEP) within the Arizona Department of Health Services reaches a large number of elementary and middle school students with curriculum-based smoking prevention programs. In FY 2004-05, 42,000 students in 433 schools were reached. In FY 2005-06, this number increased to 46,000 in 498 schools and in FY 2006-07 rose to over 60,000 students in 653 schools. The number of students in FY 2006-07 represented 15 percent of the entire student population of the state in the target grades four through eight. Yet, the effect of these activities is unknown, because no quantitative (or any other) outcome evaluation of these prevention services has been conducted to date (May 2009). In general, the impact of an intervention at the population level can be thought of as reach x efficacy (Reid, McNeill, & Glynn, 1995). Reach of BTEP's prevention service was outlined above, whereas effectiveness (i.e. efficacy under ecological (so-called 'real-world' conditions) was never investigated. The purpose of this study was to fill that gap and conduct a quantitative outcome study of BTEP's curriculum-based, in-school prevention activities with available data sources. The goal was to provide quantitative evidence of potential effectiveness of such prevention programming. A second purpose emerged during the planning of this study and was integrated as another research question to be investigated in this research. As a "byproduct" of statistical models built to investigate BTEP's prevention efforts, school-level characteristics were identified that can serve as a proxy for the severity of the smoking epidemic in a given school. Knowledge of such proxy measures that would be visible to policy makers (e.g. the school's poverty rate) – as opposed to immediate indicators of the smoking epidemic that are usually invisible to policy maker (e.g. smoking prevalence) – could enable program planers to target schools with an aggravated smoking problem, as has been advocated by a number of leading prevention researchers (Leatherdale, Cameron, Brown, Jolin, & Kroeker, 2006; Leatherdale, Cameron, Brown, & McDonald, 2005).

1.2. Background of tobacco control activities in Arizona

In 1994, Arizona voters approved Proposition 200 by a margin of 50.7 percent. The proposition raised the tobacco tax by 40 cents per pack of cigarettes. Earmarks governed that 23% of the generated cigarette tax revenue, in 1995 about 27 million dollars, ought to be reserved for tobacco control activities. These include efforts to get current smokers away from their habit ("cessation"), prevent young people from becoming smokers in the first place ("prevention") and activities related to protecting all Arizona residents from second-hand smoke (Arizona Department for Health Services, 2007; Bialous & Glantz, 1997). In 1995, moneys transferred into the Health Education Fund led eventually to the creation of the Tobacco Education and Prevention Program (TEPP, now renamed BTEP) under the auspice of the Arizona Department for Health Services (ADHS). A complex network of stakeholders was established through grant-like funding mechanisms in order to advance the three mission goals of cessation, prevention and reduction in secondhand-smoke exposure. Most notably, county health departments, non-profit organizations not associated with lobbying for the 1994 proposition, and American Indian tribes were eligible for funding to conduct tobacco control activities. All 15 county health departments established regional tobacco control offices that engaged in the coordination of various activities related to reducing the noxious effects of tobacco on their local populations. These offices mainly coordinated local, decentralized efforts by organizing community classes to help smokers quit and collaborating with school districts or individual schools to administer short (i.e. one-time, "brief") and multi-sessions, curriculum-based prevention activities ("intensive"). Every local health department operates largely on its on terms with respect to implementing school-based prevention programs.

Evaluating effects of these intensive, curriculum-based, in-school prevention activities is the purpose of this study.

1.3. Evaluation of tobacco control activities in Arizona

All activities sponsored by TEPP were and still are subjected to evaluation requirements. The Evaluation, Research and Development Unit (ERDU) at the University of Arizona won the evaluation contract in 2002 and has since then been one of the major evaluation evidence providers for the program leadership and grantees.

In October 2007, after many changes in the program's leadership and staffing over the course of its existence, the new leadership changed the name of the division into 'Bureau of Tobacco Education and Prevention' and embarked on an extensive remodeling process of all aspects of the tobacco control program. A strategic re-alignment of priorities was sought after with input from all stakeholders across the state. It led to the release of a strategic plan for the next five years. As for activities targeted at youth prevention, all county health departments and other funded entities were advised to focus these efforts on 'higher risk' youth, such as students in Title One schools and out-of-school youth. The evaluation study that this work represents, however, deals with data fathered during fiscal year 2004-05 and 2005-06.

Many challenges to evaluating TEPP/BTEP's efforts emerged over time and were met with multi-pronged evaluation strategies. Until 2004, every county provider of in-school prevention services used its own pre-post test evaluation tool, asking individual students about their experience with the classes. To get a more comprehensive, comparable way of evaluating prevention efforts, student-based evaluation was standardized across the state, with all administering staff completing class evaluation forms. Additionally, students who received prevention services completed post-class questionnaires. ERDU spent the years from 2002 to 2005 with building data collection systems to capture school-level and student-level information about recipients of prevention interventions. On the school level, databases assembled information such as the number of students served on a certain grade in a certain fiscal year, the curriculum used, number of sessions, duration of the lesson etc. This information was merged with structural information on schools provided by the Arizona Department of Education, namely average academic achievement (AZLEARNS score) and school poverty rates (percentage of students on free/reduced lunches). The purpose was to provide both BTEP leadership and prevention providers (i.e. county health departments) with knowledge about how the schools they served compared to the entire universe of Arizona district and charter schools of roughly 1,500 schools. However, all evaluation activities around prevention services could be deemed 'process evaluation', because the nature of program implementation and data collection tailored to it prohibited linking of intervention data to any relevant outcome data. More specifically, "benchmark designs", such as longitudinal control group designs (see e.g. Peterson, Kealey, Mann, Marek, & Sarason, 2000), are very costly and have to be planned early on, before implementation begins. They require substantial investments in data collection and tracking of individuals over time with associated costs and commitment to spend such monies over extended periods. Such benchmark designs have never been instituted in the case of Arizona's school-based smoking prevention programming. As a consequence, no formal, quantitative outcome evaluation of BTEP-sponsored, curriculum-based tobacco prevention activities has been conducted as of now (May 2009).

Regardless of feasibility, such designs may be simply impractical in a public health system devoted to delivering services rather than realizing high-profile research designs. However, the lack of a 'straightforward', evaluation design planned early on to examine effects of prevention services does not obliterate the necessity for quantitative evaluation evidence. This study attempts to address that need. Due to the lack of 'clean-cut' quasi-experimental data, the author attempted to address research questions by linking different data sources already available and by building statistical models that helped ruling out alternative explanations of effects as much as possible. Data on prevention activities gathered by ERDU were linked with available outcome data from a large-scale youth surveillance system, the Arizona Youth Survey (AYS) from the school year 2006 (data sources are described in detail in chapter 3.2 (p.43). The fact that many schools captured by this surveillance system were also ‘intervention schools’ with BTEP-sponsored prevention facilitated an outcome evaluation of BTEP’s school-based tobacco prevention services in the first place.

Evidence on effects of school-based smoking prevention appeared especially desirable as BTEP spent $4,400,000 in the fiscal year 2006-2007 alone on these efforts, which makes up about 20% of the program’s entire budget.

2. Literature review

2.1. The problem of tobacco use among adolescents and school-based prevention efforts

Tobacco use is the single-most preventable cause of premature mortality worldwide (World Health Organization, 2008), including the United States (CDC, 2007). Although youth cigarette smoking declined in the past (CDC, 2006) it remains a pressing public health concern because most adult smokers have initiated smoking in their adolescent years (Wills, Pierce, & Evans, 1996). Preventing the initiation of tobacco use throughout late childhood and adolescence would greatly contribute to future public health as adverse health effects of smoking (and associated costs) typically set in only in the future, after many years of continued smoking (Wiehe, Garrison, Christakis, Ebel, & Rivara, 2005). School-based tobacco prevention programs are an efficient way to reach large number of youngsters with important interventions. The Centers for Disease Control and Prevention (CDC) considered them a crucial part of a comprehensive tobacco control program (Centers for Disease Control and Prevention, 2007), if combined with more comprehensive measures such as large-scale media campaigns and strong anti-smoking policies. School-based approaches to prevention of tobacco use have started in the 1970s and have undergone various conceptual changes in the light of new evidence about effectiveness of interventions and new insights regarding the validity of underlying theories of smoking initiation. The initial 'information deficit' model - which assumed that young people picked up smoking because of a lack of adequate information about the harmful effects of smoking – and related interventions were replaced by the 'social influence' model and interventions were based on that rationale (Bruvold, 1993). This model holds that initiation of smoking is a function of various influences from peers and the greater social environment. Related interventions were based on teaching interpersonal resistance skills. The most recent concepts are based on 'comprehensive' approaches that emphasize the importance of influences of the distal and proximal social environment and stipulate the implementation of 'comprehensive' interventions that incorporate the family, school and community (Thomas, 2002). More details of findings related to the effectiveness of anti-tobacco interventions follows in the next section.

2.2. Empirical evidence on school-based smoking prevention programs

A great deal of research has been devoted to evaluating school-based tobacco prevention programs. The largest, methodologically most rigorous individual study, the Hutchinson Smoking Prevention Project (Peterson, Kealey, Mann, Marek, & Sarason, 2000), yielded no significant differences on measured outcomes between the intervention group over the control group. However, even this large-scale study with a benchmark design drew considerable criticism. Especially its wide-ranging conclusions questioning the overall effectiveness of smoking prevention programs in general was attacked in later editorials (e.g. Sussman, Hansen, Flay, & Botvin, 2001), accompanied by a call to move away from the main effects question ("what works") to the moderated effects questions ("what works why under what conditions for whom"). Another important implication in the debate over the fairness of conclusions of program effectiveness was to base policy recommendations on meta-analytical reviews rather than individual studies. To date, a large number of reviews and meta-analyses have been performed on drug prevention studies including tobacco prevention. These meta-analytical findings vary in their conclusions, depending on inclusion criteria and general methodology. Table 1Table 1 below gives an overview of findings from several meta-analytical studies on school-based tobacco prevention programs.

Table 1. Overview of meta-analytical findings on school-based tobacco prevention programs

Study	Major findings
Bruvold, W., 1993	Social norm-based or social pressure resistance curricula are superior to information-deficit model based curricula in changing behavioral outcome measures.
Hwang et al., 2004	"The grand mean of effect size was .36 for knowledge, .16 for attitude, .16 for skill, and .15 for smoking behavior. Therefore, the psychosocial smoking programs were effective in about a 10% relative reduction in smoking behavior. " (p. 707).
LaTorre et al., 2005	Effect sizes range from 5% to 60% in favor of intervention groups. "In conclusion, in order to achieve a higher level of effectiveness it is widely recognized that smoking prevention programs should have the following components: sustained application, booster sessions over several years; reinforcement in the community; involvement of parents and the mass media; programming smoking prevention activities within a more comprehensive school health promotion programme." (p.285)
Rooney, B. & Murray, D.; 1996	Effects of smoking prevention programs may be limited in magnitude. "Even under optimal conditions, the reduction in smoking may be only 0.50 to 0.75 standard deviation units, or perhaps 20%-30%." (p. 48)
Rundall, T. & Bruvold, W., 1988	Smoking prevention programs are more effective than alcohol prevention programs
Thomas, R. & Perera,R., 2002 (Cochrane Review)	"Three of the four high quality multi-modal interventions showed a positive significant effect. It is possible that combining social influences models with other components, such as community interventions and generic social competence training may improve effectiveness."
Tobler, N., 1997	Interactive programs (those including refusal skill training etc.)are much more effective in preventing substance abuse (incl. cigarettes) than non-interactive (such as knowledge-based or emotional self-regulation)
Wiehe et al., 2005	Very little evidence for long-term effects of school-based prevention programs

Although somewhat equivocal results seemed to emerge from various meta-analyses, components of effective programs can be summarized as follows. They are interactive (rather than the teacher-centered lecturing) with focus on general skill and social skill development, embedded in broader community interventions (such as media campaigns, cessation for adults), they provide booster sessions to increase temporal duration and they are ideally tailored to the specific needs of the target audience. One recent study by Pizacani et al. (2008) warrants special mentioning, as their methods resemble somewhat the methods used in this study. They found that increases in smoking prevalence in cohorts after three years were significantly greater for cohorts in the defunded period than for cohorts in the funded period and not different in districts that were never funded. It can be concluded that smoking prevention programs can have an impact on self-reported smoking and cognitive precursors of smoking such as favorable attitudes towards smoking.

2.3. Risk factors for smoking & the importance of school environments

This section provides an overview of factors that were identified as relevant for adolescent smoking. This study focused on possible effects of school-based prevention efforts on smoking-related outcomes, with the interventions conceptualized as a school characteristic. It is therefore critical to illustrate the importance of the school environment on children's health behavior and school-level factors that increase students risk to use tobacco. On the other hand, individual-level risk factors for smoking are not a primary concern of this research. However, because of their potentially confounding influence on conclusions regarding higher-order (i.e. school) effects, they have to be included as 'control variables' when running multilevel models. That is, as West (2006) has pointed out, school effects research has to invalidate threats to conclusions on school-level factors by ruling out that differences in student composition between schools can sufficiently explain variation between schools. For example, if some schools contain more impoverished students than other schools, conclusions about school-level factors such as provision of prevention services may be confounded. An even more obvious example would be comparing schools serving only high achievers with schools serving primarily students who had dropped out of district schools for misbehavior. Without taking into account the vastly different make-up of the student population (realized as level-1 predictors), conclusions regarding the influence of school-level factors would be severely compromised. The presentation of previous evidence in the following sections is structured based on their utility to identify relevant variables for the final evaluation model.

2.3.1. Person-level factors influencing smoking behaviors

Numerous individual-level factors been identified that increase the likelihood for adolescents to pick up smoking, such as socioeconomic status, peer and family bonding, and many others (Conrad, Flay, & Hill, 1992; Hawkins, Catalano, & Miller, 1992; Flay, 1999; Tyas & Pederson, 1998). As these risk factors are not the primary concern of this study, only factors with relevance to the pertaining research question had to be taken into account. The main concern of this study is the influence of school-level characteristics on students' and schools' tobacco-related outcomes. Aveyard et al. (2004) have pointed out that failure to adjust for confounding individual-level factors can lead to flawed conclusions in multilevel studies where the focus is on associations between school-level characteristics and student-level outcomes. For example, if schools differ in their students' composition of personal background (e.g. if school A has many students from smoker households and school B has only few students with smoking parents), the school environment -including BTEP's efforts- may exert little influence on a student's tobacco-related attitudes and behaviors. To the extent that schools differ in their student composition relevant for tobacco-related behaviors (such as accumulating students from higher-risk backgrounds), relationships at the ecological level are biased upwards. This would be a case of under-controlling confounding factors and a threat to internal validity of findings (Aveyard, Markham, & Cheng, 2004). However, over-adjusting can be as misleading, i.e. yield deflated level-2 coefficients by reducing the variability of outcome measures with little power to detect effects of level-2 factors. This is the case if a study's focal interest – such as in this work - is identifying the influence of school-level factors, because school-level factors can themselves modify individual risk factors that are related to the outcome under study (such as tobacco-related knowledge & attitudes or school bonding). Therefore, Aveyard et al.'s (2004) recommendations were put to action by including those individual-level risk factors presumably unaffected by school characteristics. This strategy was employed to rule out that the student composition of the school accounts for observed differences between schools. Some of these factors may not have been included because available outcome data (AYS 2006) did not contain them.

The table below outlines possible individual-level confounders as identified by the general framework of Aveyard et al. (2004) that helps decide which level-1 variables have to be considered when the focus is on school-level predictors of individual-level smoking.

Table 2. Individual characteristics unaffected by school environment (level-1 control variables)

Family characteristics	**Inclusion in the model**	**Reason for (non-)inclusion**
Acculturation	No	No measure available
Socio-economic status of the family	No	No measure available
Parental smoking	Yes	Powerful confounder
Parental attitudes	Yes	Parents' attitudes towards smoking available
Attachment to family	Yes	Scale score of emotional attachment to parents available (Cronbach's alpha =.76)
Personal characteristic	Inclusion in the model	Reason for (non-)inclusion
Age	Yes	Strong relationship with smoking
Gender	Yes	Needs to be included because of different student composition between schools providing outcome data
Ethnicity	Yes	May constitute a proxy for unmeasured risk and/or protective factors.
Personal income	No	No measure available

A careful analysis has to reveal to what extent these factors provide redundant information, i.e. are inter-correlated. Based on the results, factors will be excluded

2.3.2. School factors influencing smoking behaviors

It is well established that school-level characteristics do exert an influence on a large number of developmental outcomes, such as academic achievement and health behavior (Sellstrom & Bremberg, 2006; West, 2006; Aveyard, Markham, & Cheng, 2004). In the case of smoking, a strict non-smoking school policy (Aveyard, Markham, & Cheng, 2004), smoking prevalence in higher grades (Leatherdale, Cameron, Brown, Jolin, & Kroeker, 2006) among many other factors are associated with increased likelihood to use tobacco over and above individual level risk factors (Kairouz & Adlaf, 2003). Likewise, a school norm less disapproving of drug use increases the likelihood of never-users to become users (Kumar, O'Malley, Johnston, Schulenberg, & Bachman, 2002). Although there seems to be little doubt that school level factors influence health-related behaviors and attitudes of children, it remains unclear through what mechanisms these influences work. As Aveyard et al. (2004) have pointed out, school-level measures that are supposed to exert an influence at the individual and/or school level have to be derived from substantial theoretical considerations. There should be buttressing previous evidence explaining **why** there ought to be an influence. For example, in a study of school level norms and smoking, Kumar et al. (2002) didn't find evidence for the three school-level factors school size, urbanicity and type of school (public vs. private). They did not offer any theoretically derived reasons why those factors should be important ecological factors for increased risk. The following paragraphs will outline school level variables that could show a significant relationship with susceptibility to smoking & pro-smoking attitudes. Theoretical explanation and brief empirical evidence as to why those higher-order indicators could affect students is given.

2.3.3. School-level poverty

On the individual level, poverty was identified as an important risk factor for smoking among adults and youth (Conrad, Flay, & Hill, 1992). The CDC reported for 2006 (Centers for Disease Control and Prevention, 2007):

> *The prevalence of current smoking was higher among adults living below the federal poverty level (30.6%) than among those at or above this level (20.4%).*

Individual level poverty might affect risk for substance use through different mechanisms than an environment of pervasive poverty, such as a high rate of impoverished students. To avoid an 'atomistic fallacy' (Diez-Roux, 1998) by assuming individual-level relationships would apply to higher order units of analysis (such as schools), evidence is presented that links school level poverty to unfavorable student outcomes, such as cigarette smoking.

In theoretical terms, it could be argued that in schools recruiting students from low SES catchment areas a majority of pupils would be used to smoking at home, because smoking tends to concentrate in socio-economically disadvantaged populations (CDC, 2007). This relationship could lead children from lower socioeconomic strata to perceive smoking as more normative than their peers from less disadvantaged backgrounds. In their review, Sellstroem and Bremberg (2006) suggest that high socioeconomic status is among those school factors that are conducive to favorable student outcomes. Several multi-level studies found that neighborhood poverty was associated with higher smoking prevalence after potentially confounding person-level factors were controlled for (Datta et al., 2006; Shohaimi et al., 2003). Therefore, impoverished environments appear to constitute a risk factor for problem behaviors in general and smoking in particular. In this sense, the concentration of students from impoverished backgrounds could constitute a "breading ground" for a school norm less disapproving of tobacco use (Conrad, Flay, & Hill, 1992). Consequentially, a measure of school-level poverty was included in the evaluation model for two reasons. First, poverty of the student population could be weakening the influence of curriculum-based prevention services by indicating a 'higher risk' population less sensitive to such interventions. Secondly, this research tried to identify school-level proxy measures that are indicative of a more pronounced smoking epidemic in a school and readily visible to administrators.

2.3.4. School-level academic achievement

On the individual level, an extensive body of evidence confirms the connection between academic failure and students' adverse behavioral outcomes (Hawkins, Catalano, & Miller, 1992; Bryant, Schulenberg, O'Malley, Bachman, & Johnston, 2003). A study by Aveyard et al. (2004) could not find an association between school-level achievement scores and smoking rates after educational style of the school (authoritative vs. laissez faire) was included in the analysis. It remains unclear, however, how much that finding resulted from specific features of the UK school system or the choice of school-level predictor variables. Although educational style of a school may be more important than academic achievement, the goal of this research was finding out if school-level academic failure constitutes a proxy that helps identify schools with an aggravated smoking problem. As school-level academic failure is readily visible to educational administrators or anyone with an internet connection (per ADE website), program planners could then use this knowledge to target schools with a presumed high-risk profile for more intense or different health promotion services.

It could be argued that a climate of academic underachievement might be a 'breeding ground' for various other problematic behaviors, as disaffection from school was identified as an important individual-level risk factor for smoking (Conrad, Flay, & Hill, 1992). If a large proportion of a school's student population fails academically, then this indicates an environment that may also provide more exposure to other undesired behaviors (such as drinking, gang involvement etc.) that are related to academic underachievement. A high level of school-level academic achievement may itself be a proxy for other protective school-level factors, such as high commitment to the school by a large part of the student body (Resnick et al., 1997) and ability of the school to efficiently educate their students, an indicator of "school ethos" (West, Sweeting, & Leyland, 2004). For these reasons, a measure of school-level academic achievement was included in the evaluation model.

It is important to consider that school-level indicators may be inter-related. Battistich et al. (1995) showed that low levels of school poverty were associated with student's achievement in mathematics. In a very impressive study, Harris (2007) concluded after examining a near-census sample of US public schools, that

> *"Low-poverty schools are 22 times more likely to reach consistently high academic achievement compared with high-poverty schools. Schools serving student populations that are both low poverty and low minority are 89 times more likely to be consistently high performing compared with high-poverty, high-minority schools." (p 367)*

These findings indicate that school-level factors need to be examined for their correlation amongst each other in order to avoid multicollinearity.

2.3.5. Curriculum-based tobacco prevention

Most evaluation studies examined the effect of school-based prevention activities on the person level, i.e. tried to find out if students in experimental groups had more favorable tobacco-related outcomes than those in control groups. In this study, curriculum-based prevention activities were conceptualized as a single school factor by integrating data from administrative records. The lack of individual-level intervention data necessitated a different approach to the assessment of possible intervention impacts. Based on evaluation findings on school-based tobacco prevention activities (chapter 2.2, p.17), it had to be assumed that conceptualizing treatment as a school characteristic would not yield high effect sizes in favor of the intervention, as even studies with a more rigorous design have failed to find even moderate effect sizes (Peterson, Kealey, Mann, Marek, & Sarason, 2000). However, in the case of some prevention providers in Arizona, curriculum-based prevention activities may be indicative of activities that extend beyond the immediate class rooms that received the curricula to the school as a whole. The conceptual logic behind the school-level prevention index is as follows. Data on number of students receiving programming is combined with number and duration of lessons, adjusted for enrollment size. This index increases as the relative number of students receiving programming increases and to a lesser degree if amount of services was higher per students, i.e. more and/or longer lessons. Students providing outcome data in schools with a relatively high prevention index score had a higher likelihood of being recipients of actual programming than students in schools with lower prevention index score. In addition, a high prevention index indicates that many students had received prevention programming. This may lead to a school climate less accepting of smoking and pro-smoking attitudes. Finally, a high prevention index could be indicative of school administrators that are more concerned about smoking and represent an already less permissive environment for smoking. Details on the computation of a school-level prevention index can be found in chapter 3.4.7, p.56.

2.3.6. School-level factors with potential relevance not included in this research

Several school-level characteristics shown to influence students' smoking behavior were not included in the evaluation model of this study. Variations in the school's smoking policy have been found to influence students smoking behavior (e.g. Pentz, 1989; Sellstrom & Bremberg, 2006). However, very little variability between Arizona schools for tobacco-related policies was found by a 2002 survey among Arizona school principals and health education teachers. For example, 99 percent of all schools had policies prohibiting smoking for students and 96 percent prohibited smoking for school staff (Arizona Department of Education, 2002). With almost no variability between schools with respect to smoking policies, it is highly unlikely that differences in smoking-related outcomes could be attributed to differences between smoking policies of schools.

Smoking prevalence in higher grades was consistently identified to increase students' risk to become susceptible to smoking (e.g. Leatherdale, Cameron, Brown, Jolin, & Kroeker, 2006). Most schools in this study (85%), however, served only students up to grade eight so that the study subjects constituted the highest grade in most examined schools. Smoking 'role models' in higher grades could therefore not be a possible explanation of between-school differences in smoking-related outcomes.

Finally, dropout rates were not included for the same reason as prevalence in higher grades. In most schools, the target population of this study (students in grade eight) represented the highest grade and most students in eighth grade were within the age range for which compulsory schooling applies in Arizona (>=16 years, Arizona Department of Education, http://www.azed.gov/asd/dropout/azlaw.asp). Almost no variability could have been expected for school-level dropout rates.

2.4. Outcome measures for school-based tobacco prevention programs

Ultimately, school-based programs to prevent substance use (including tobacco) were created to prevent undesired outcomes from occurring in the first place, before their first manifestation. As Kaplan (1990) pointed out, behavior is the crucial outcome of preventive and health promotion interventions. Many evaluation studies of substance prevention programs, however, have also included cognitive measures as program outcomes, such as attitudes towards substance use, perceived ‘coolness’ of use, future intentions to use, and others (Dielman, 1994; MacKinnon, 2008; Botvin et al., 1992). These cognitive factors may represent mediators of program effects (MacKinnon, 2008; Botvin et al., 1992) and should ideally be analyzed in such a way. Although testing of mediational hypotheses is beyond the scope of this study, effects of BTEP prevention activity on these mediating variables, i.e. attitudes and norms, were examined.

Because most tobacco prevention curricula – including the ones used by most Arizona prevention providers (see chapter 3.1.2, p.41) - also try to ‘denormalize’ smoking (i.e. debunk the myth held by many youth that smoking is highly prevalent among their peers and that smoking is an expression of ‘coolness’), attitudinal and norm-related measures will be treated as outcome measures in this study. If an effect could not be found for self-reported smoking but for attitudinal measures, this would be an encouraging finding in that at least precursors of actual behavior have been influenced by the intervention.

Therefore, the main outcomes for this evaluation are self-reported *ever smoking* (with never smoking defined as having never smoked - or only once or twice in the past - but not at all in the past 30 days), c*urrent smoking* (defined as any amount of smoking during the past 30 days), and an ordinal measure of smoking acquisition. This latter measure of self-reported use was chosen in accordance with the methodology used by Peterson et al. (2000) to examine if schools with high prevention activity led to a lower “commitment” to smoking among those who already initiated. This outcome also represents a less rigid outcome than the rather strict outcome of *ever smoking* and c*urrent smoking* which are all-or-nothing concepts that may oversimplify actual smoking behavior of middle school students. If amount of program activity and amount of smoking among those who already smoke would be inversely related, this would indicate that BTEP prevention activity in a school might create an atmosphere that leads smoking youth to smoke less. Details on this quantitative outcome measure can be found in chapter 3.4.3, p.52.

The second set of outcomes was measures of smoking-related attitudes and norms. Details on actual measures and item content will be provided in section 3.4, p. 51. Again, possible mechanisms of mediation are beyond the scope of this study and will not be considered.

Finally, number of smoking friends was analyzed as outcome variable as it can be conceived of as a “radar” into the school environment. If amount of prevention activity were associated with students reporting fewer friends who smoke, this would be an indicator of effectiveness of prevention activities beyond those who actually contributed data to AYS.

Because this study deals with two conceptual levels – the school-level and the person-level - all outcomes specified in chapter 3.4 were aggregated to the school level for ecological-level-only analyses.

2.5. The Five-Data-Box Conceptualization – A General Evaluation Framework

Wittmann (Wittmann & Klumb, 2006; Wittmann, Nuebling, & Schmidt, 2002; Wittmann, 1990) proposed a general framework for program evaluation, the five-data-box conceptualization (see Figure 1 below) . The boxes refer to different pieces of information that have to be considered when conducting social science research and program evaluation.

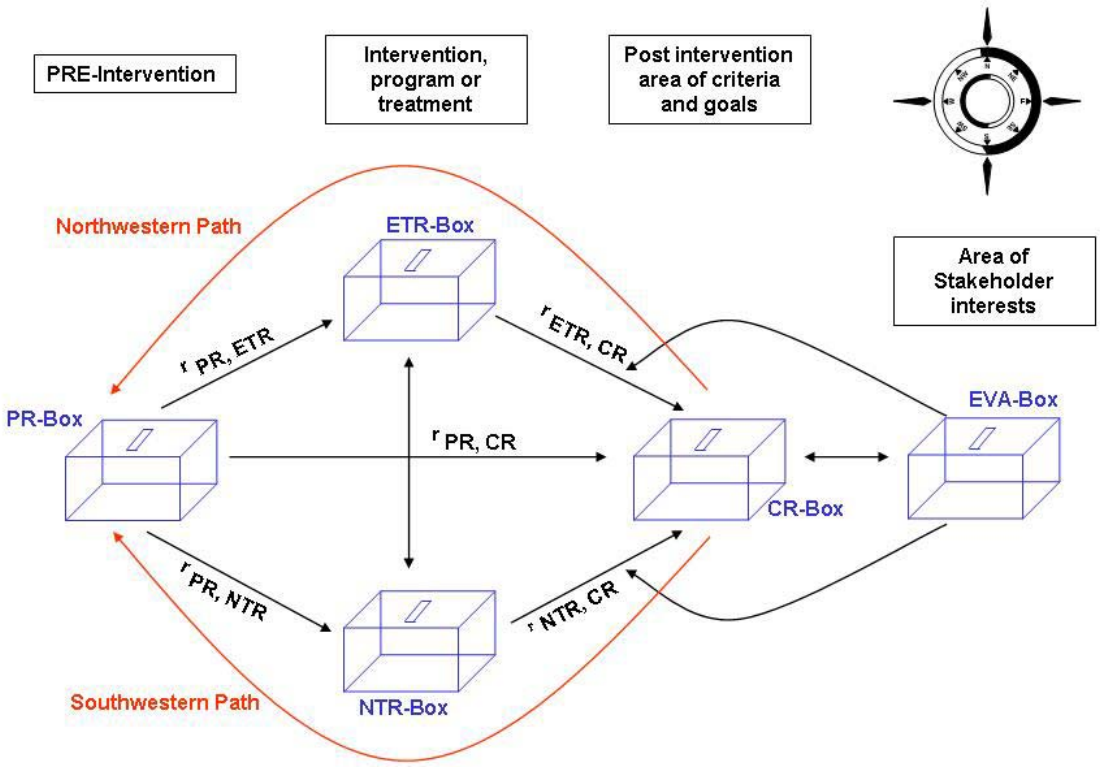

Figure 1. Five-Data-Box Conceptualization (adapted from Wittmann & Klumb, 2006; in: Bootzin & McKnight, 2006, p. 186)

The EVA box (for evaluation) contains expectations and input from stakeholders interested in a program's performance and outcomes, such as administrators, funders, and policy makers. These expectations and inputs are (and should be) reflected in selected outcomes measures, which are contained in the CR box (for criterion measure or outcome). In the case of BTEP's school-based prevention programming, these include the variables spelled out in the previous section (2.4). The ETR box (for experimental treatment) contains variables that capture experimentally manipulated variables. In the approach taken for this research there were no truly randomized treatments, especially not on the student level, because schools were selected for treatment and not individual students. The NTR box contains all variables that could not or were not manipulated randomly, such as dosage or strength of an intervention. The measure of this research that could be located in the NTR box is the school-level index of BTEP prevention activity, as schools were not assigned in a perfectly random manner. This index represents a dosage measure that expresses ho much prevention services a school received over an aggregated amount of time (two years) and aggregated amount of students. The PR box (for predictor) entails all variables that may have a potentially confounding impact on the relationship between outcomes and intervention. Variables in this box are also used to assess potential selection-into-the-treatment effect. In my research, all level-1 predictor variables belong into this box, because the focus of this research is on effectiveness of BTEP intervention activities and these level-1 predictors have to be controlled for in multi-level modeling in order to rule out that the differences in student demographic composition between schools explains differences in outcomes, rather than BTEP activities.

The different paths between the boxes help identifying important relationships to be examined and serve as a heuristic guide to data analysis. In this study, the intervention was neither perfectly randomized nor completely systematic, i.e. when assignment of treatment to units (i.e. schools) is driven by some pre-intervention cut-off score or other deliberations such as political fairness. This is indicated by the double-headed arrow between NTR box and ETR box to indicate that experimental treatments and non-experimental interventions are not dichotomies but rather constitute poles of a continuum.

The $r_{PR,NTR}$ path represents the relationship between covariates or pre-intervention measures and the treatment measure. This path symbolizes possible 'selection-into-the-treatment' effects and stands for all analyses & steps performed to ensure the equivalence of comparison groups. These steps include inspection of missing data patterns by group status (control vs. intervention) to ensure no differential attrition has occurred. Such differential attrition is a major threat to internal validity of a study (Shadish, Cook, & Campbell, 2002). Another step to ensure adequate treatment of possible confounding is to include level-1 predictors into the multi-level model building. This ensures that variations in outcomes are not due to differences in students populations between schools.

The $r_{PR,CR}$ path denotes relationships between predictor variables and outcomes. Wittmann (2006) points out that there should be symmetry between criteria and the set of predictor levels in order to avoid validity issues, such as assuming a zero correlation when in fact there was a mismatch in symmetry (e.g. a mismatch between levels of generality or aggregation) between operationalization of predictor and criterion variables. No detailed discussions of all features of symmetry can be given here. Readers are referred to Wittmann et al. (2006, 2002, 1990).

In Wittmann's (2006) terminology, this research follows the "Southwestern Path" by regressing the CR box on the NTR box and by incorporating variables contained in the PR box. This renders this evaluation of BTEP's prevention programming as a quasi-experimental approach, which is fully justified by the fact that students receiving the treatment were not assigned perfectly randomly to treatment conditions. The $r_{NTR,CR}$ path, which symbolizes the crucial relationship to be examined here, will be represented in statistical analyses on two levels: the school-level (level-2) and the individual level (level-1) combined with the school level. The former represents an "ecological" approach as only relationships among level-2 measures were examined, with level-1 outcomes aggregated to the school level (mostly as school means or prevalence rates). An OLS regression model was built to examine these relationships. This ecological analysis could also be classified as "sociological perspective" because classic sociological questions will be addressed, such as how the poverty level of a school is related to the smoking epidemic. Answers to these questions helped shed light on the second major goal of this study, i.e. identification of 'proxy variables' readily accessible to administrators that indicate a high tobacco problem in a given school. The ecological level also addressed the BTEP effectiveness question, but cannot rule out possible confounding influences due to differences in student demographics. To this end, multi-level regression models were developed and - based on theoretical considerations and statistical fit indices (such as difference scores in deviance chi-squares) - refined in order to arrive at a maximally parsimonious yet accurate representation of results. Both models are outlined in more detail in the next section.

2.6. Theoretical model for evaluating outcomes of Arizona's school-based tobacco prevention programs

Figure 2 below shows graphically what relationships were examined in this study. Aggregated student-level outcomes were regressed as school-level outcomes on school-level predictors (model above dotted line with dark-green block arrow). Findings are reported in chapter 4.2.2 (p.95).

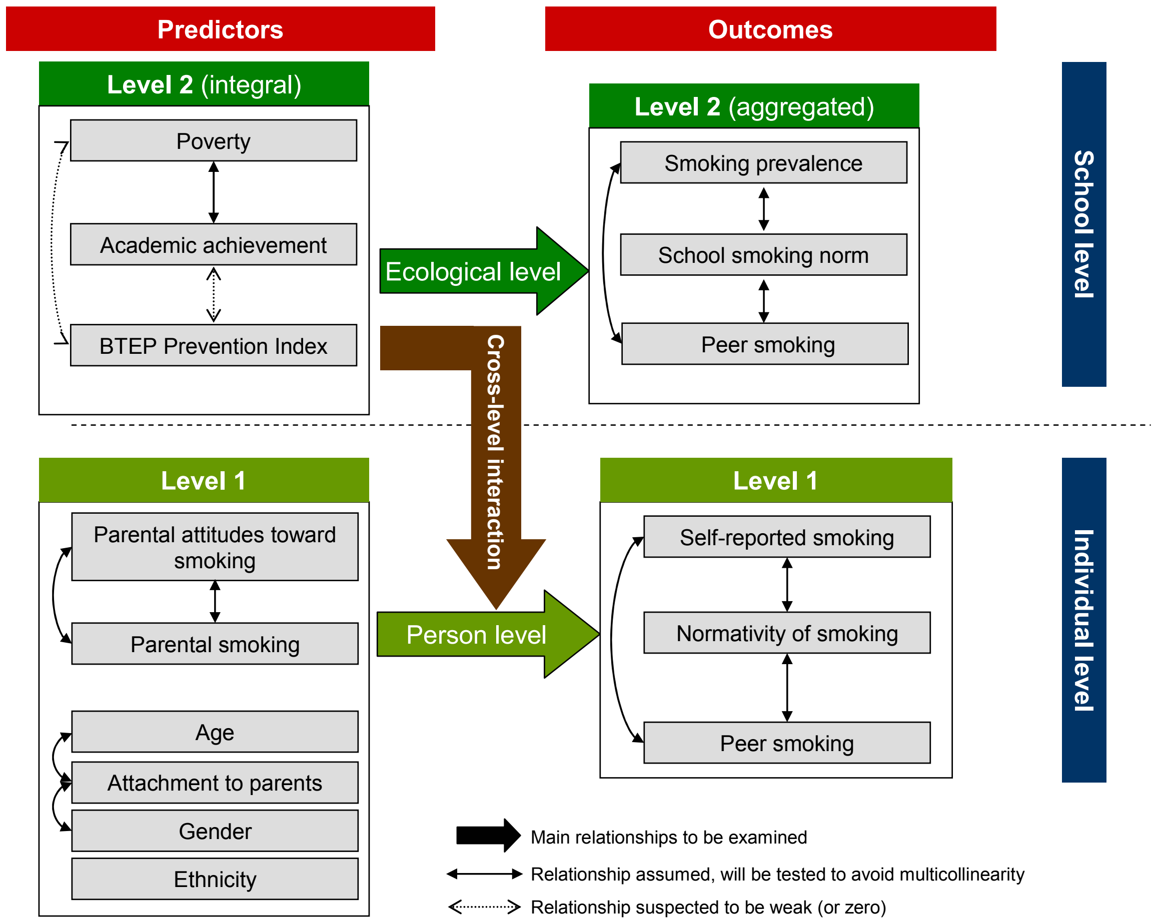

Figure 2. Conceptual model to evaluate BTEP prevention effectiveness

The model below the dotted line indicates the structural level-1 model that was built as a basis for consecutive multilevel models. In this model, level-1 predictors were regressed on level-1outcomes (light-green block arrow, chapter 4.1.4, p.75). This level-1 only model represents an instantiation of Wittmann's $r_{PR,CR}$ path. The brown block arrow indicates that relationships between level-1 predictors and level-1 outcomes were modeled as outcomes in the hierarchical linear models. Not visualized in Figure 2 is the examination of school averages of outcome variables with level-2 predictors once level-1 predictors have been controlled. Results of these models are presented in chapter 4.3.2 (p.105) The dark-green block arrow for level-2 models and the brown block arrow for cross-level interaction terms are realizations of Wittmann's $r_{NTR,CR}$ path.

3. Methodology

As pointed out in chapter one, this research tries to address a pragmatic need for evidence on the effectiveness of BTEP's curriculum-based prevention services. In the terminology established by Flay (1986), this study constitutes an "implementation effectiveness study". Flay (1986) distinguished between classes of efficacy and effectiveness studies. An implementation effectiveness study examines an effective intervention (in the case of this research mainly two prevention curricula, see below) that reached a variable amount of the target population with unclear acceptance of the treatment by the target audience. Further, interventions happen under ecological (so-called "real-world") conditions and implementation parameters (such as implementation fidelity) vary widely, as is the case with county health departments across the state of Arizona. This research also qualifies as quasi-experimental because some level-2 units (schools) received treatment and some others did not. Implementation under ecological circumstances created a 'natural experiment' situation where schools ended up in the control vs. treatment group neither totally at random nor totally systematic. This study also qualifies as an "ecological" approach because of the combination of group-level and individual-level variables and because of the unknown extent to which individuals received treatment (Friis & Sellers, 2004).

A number of reasons led to choosing a methodology with 'archival' data for this evaluation study. First, implementation of curriculum-based prevention services is subcontracted to county health departments, who in turn collaborate with schools and school districts to deliver the curricula. Different counties implement prevention services in varying ways. In fact, it would not be very unreasonable to demand 15 separate outcome evaluations, because county health departments are left to their own digression for delivering prevention services. However, no systematic data collection or evaluation design that would have allowed for separate outcome evaluations was established when prevention service delivery first started. Rather, systematic and standardized data collection on prevention services did not start until 2004 (Arizona Department for Health Services, 2007). Accordingly, the pure lack of systematic, long-term, rich outcome data together with dispersed implementation necessitated tailoring an outcome evaluation around available data. That data were available from an enormous statewide youth survey, the Arizona Youth Survey 2006 (Arizona Criminal Justice Commission, 2006). Details of these data are given further down (see chapter 3.2.1, p.43). Two more reason speaks against 'benchmark' research designs such as randomized, longitudinal control group experiments - such as the one used in the Hutchinson Smoking Prevention Project (Peterson, Kealey, Mann, Marek, & Sarason, 2000). Although they may have a high degree of internal validity, such designs are neither permissive nor desirable in an applied setting such as Arizona's statewide prevention program. They are not permissive because program implementation would then follow research design considerations. Program implementation, however, may be guided and informed by other priorities such as maximum spread across the targeted population and the providers' desire to connect to certain schools. Randomized control experiments may not even be desirable because program provision ought to maximize accessibility and availability of services to a maximum number of clients (students) or may target populations with the highest needs. Both of these goals render randomized control-group experiments inappropriate for applied settings where the focus is on program delivery, not realization of research principles.

Various data sources had to be integrated to arrive at data sets suited to address the crucial research questions. The following sections give details about implementation of tobacco prevention programming in Arizona, the data sources as well as decisions made regarding statistical manipulations, processing and aggregating.

3.1. School-based tobacco prevention programming in Arizona

In 1996, BTEP started funding intensive curriculum-based prevention education in public and charter schools throughout the state (Arizona Department for Health Services, 2007). Only very coarse directions regarding implementation were given to contracting entities, such as to target students in grades four through eight, choose an "evidence-based" curriculum ("evidence-based" meaning endorsed by SMHSA or the CDC) that addresses knowledge and affective components of tobacco use, social influences or personal and social skills training. A minimum of eight sessions plus no less than two booster sessions in subsequent grades was required from implementing agencies. However, some curricula chosen by AZ contractors were not designed to provide booster sessions in higher grades. Further, implementation of curricula is often not happening in consecutive cohorts so that students who received a curriculum in grade X in year Y are not necessarily targeted again with a booster session in grade X+1 in year Y+1. Finally, administering staff was advised to be trained and not deviate from the curriculum content to assure maximum treatment fidelity (Arizona Department of Health Services, 2003). This high degree of freedom of choice produced a hodgepodge of different program implementations in terms of targeted audience, selected curricula, implementing staff and treatment fidelity. This variation directly relates to the evaluation strategy used in this research. That is, technically, every prevention provider would have needed to gather his or her own pre-intervention data, post-intervention and follow-up data in a prospective (within-subject) design. Ideally, every prevention provider would have also needed a no-treatment control group to adequately address the effectiveness question. Apparently, such commitments in terms of money, infrastructure, personnel and long-term planning could not be asked from local agencies dealing with instable commitment from schools and a focus on delivering interventions rather than conducting methodologically high-end, costly evaluations. Additionally, high turnover rates among BTEP administration staff left many local health departments at their own terms for making program-related strategic decisions.

3.1.1. Implementation of curriculum-based prevention programming

In essence, the situation of school-based tobacco prevention in Arizona can be described as a multi-site, multi-curriculum, minimally standardized endeavor that poses considerable challenges to evaluation efforts. BTEP is contracting with county health departments to deliver prevention services to Arizona schools. Figure 3 below shows the location of schools with grades four through eight by their intervention status in the fiscal year 2006-07.

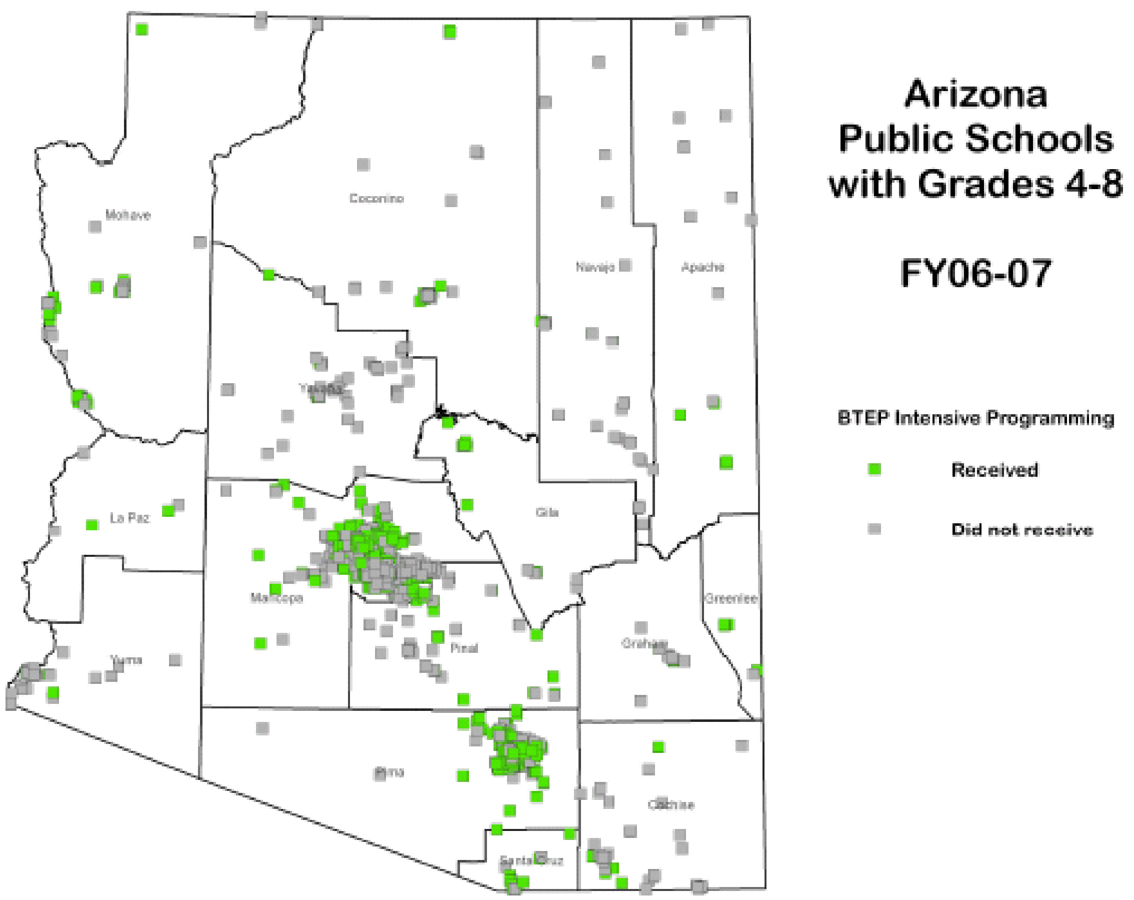

Figure 3. Location of Arizona schools by BTEP intervention status in FY 2006-07 (Taken from Evaluation Research and Development Unit, 2008, p.16)

One county health department may directly provide services by sending health educators into schools; others may choose to subcontract with schools or school districts to teach students. County health departments select their intervention targets (school districts, schools and grades therein) by largely unknown criteria but are advised to reach students at 'higher risk' for tobacco use. Schools are deemed to put students at higher risk for using tobacco if one or more of the following criteria are fulfilled: Title I schools (federal program for Improving the Academic Achievement of the Disadvantaged), schools with high percentages of children who are enrolled in the federal free or reduced price lunch program (National School Lunch Program), schools with high proportions of minority populations, including American Indian and Hispanic/Latino students, and schools that are underperforming academically. Contractors were also encouraged to serve public schools located on sovereign American Indian tribal lands. As pointed out above, one major aim of this research is to validate or falsify several proxy measures with respect to their ecological validity of indicating a "high-smoking-risk environment". As pointed out above, BTEP requires its prevention contractors to choose curricula from a list of evidence-based programs that were reviewed and supported by federal public health agencies such as SAMSHA and the CDC (Coffman, 2007). Because of the contractors' freedom to choose curricula according to their preferences, a multitude of curricula is in use throughout the state, creating great heterogeneity between and even within contractors. Furthermore, not all curricula in use qualify as 'evidence-based'. As can be seen in

Figure 4, most of the students in grades six through eight (around 80%) receive either one of two curricula "Project Alert" or "Get Real About Tobacco", both of which were deemed 'evidence-based' by SAMSHA. A more detailed review of curricula in use can be found further down.

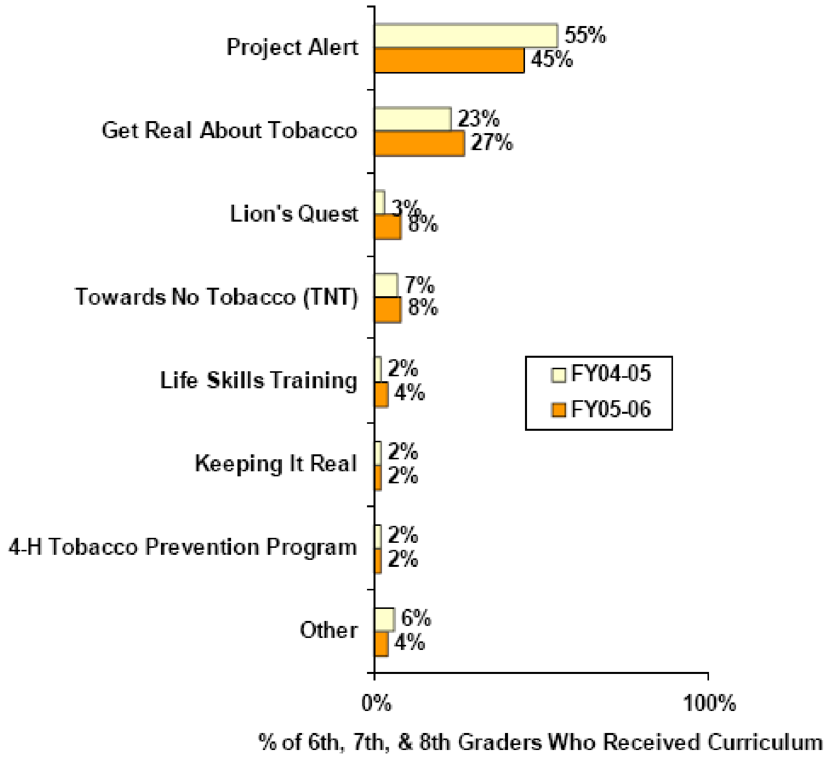

Figure 4. Tobacco Curricula used for grades 6-8 (taken from Coffman, 2007, p.21)

As pointed out above, even within contractors, there may be considerable heterogeneity of implementation because it remains unclear to what extent curricula are used consistently over time or consistently within the same schools. Examining the effects of variation in these implementation parameters is, however, beyond the scope of this research. The evaluation approach taken in this research aggregates various implementation measures without considering their variation over time within the same unit of analysis (i.e. school). As can be seen in Figure 5, classroom teachers administered the majority of classes in grades six to eight.

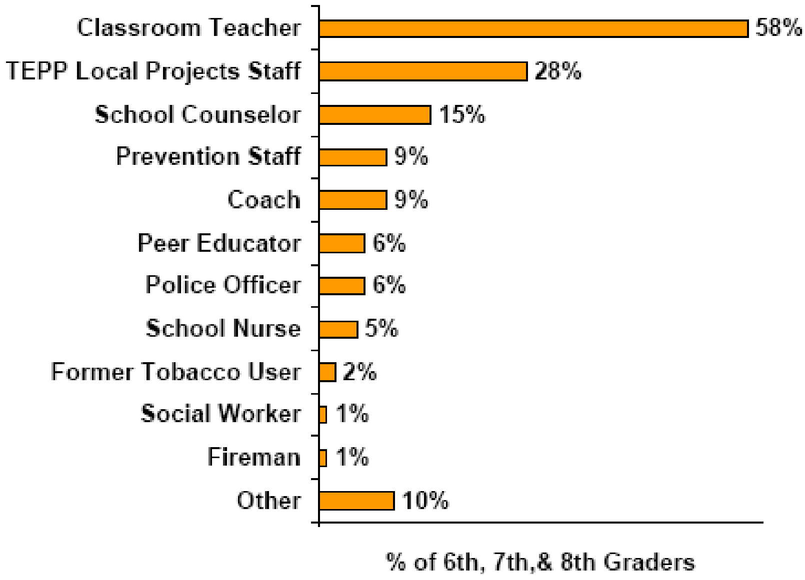

Figure 5. Personnel administering intensive prevention programs, grades 6 to 8 (taken from Coffman, 2007, p.27)

Possible differences of effectiveness by administering personnel are also beyond the scope of this research. Although influences of administering staff on effectiveness of prevention programs cannot be excluded per se, Cameron et al. (1999) found no differences between teacher-administered or nurse-administered interventions on smoking related outcomes of eighth-grade students.

3.1.2. Content and evaluation evidence of most widely used prevention curricula

The two most frequently used curricula in middle schools grades are 'Project Alert' and 'Get Real About Tobacco' (see Figure 4 above). Project Alert is built on the following basic assumptions (www.projectalert.com). 1) Adolescents use drugs primarily for social reasons and because of forces in their social environment (smoking parents, siblings or peers). They want to act in a socially desirable way. The curriculum teaches that drug use is not as normative as they may think, provides role models that successfully resist (through videos), and teaches specific resistance skills. 2) Prevention programs ought to encourage the development of motivation to refuse drugs. The curriculum provides videos of older teens who explain why they do not use drugs, as they were shown to be more credible than adult persons (such as teachers, Ellickson et al., 1993). The interactive nature of many lessons adds to immediate learning success and increases the appeal of being part of the program. 3) Drug prevention should target substances that most widespread in use and are used first in young people's lives (i.e. tobacco, alcohol and marijuana). Project Alert puts an emphasis on these substances. Ellickson et al. (1993) conducted a randomized control group experiment with middle schools students. They found that Project Alert lowered cognitive precursors of actual drug use (i.e. perceived consequences of drug use, normative beliefs, and expectations of future use) significantly among students who received the program, as well as self-reported actual smoking behavior. Because cognitive factors such as attitudes are known risk factors for the eventual uptake of smoking, it was important for this study to examine effects on measures of intentional and cognitive correlates of actual tobacco use, i.e. perceived coolness, future intentions to smoke, and normativity of smoking.

No empirical evidence could be found for the second most prevalent curriculum in Arizona middle schools, 'Get Real About Tobacco' (GRAT). According to an internet resource[i], GRAT's goals are to

> *"Reduce the likelihood that students will start using tobacco products, encourage students who do use tobacco to quit, and to help students promote anti-tobacco messages"*

According to the Rocky Mountain Center for Health Promotion and Education[ii], a not-for-profit organization funding school-based tobacco prevention in Colorado, GRAT helps students identify situations of high susceptibility to tobacco use, teaches skills to successfully master them, describes the effects of tobacco use and second hand smoke on the body, and allows students to share their thoughts about tobacco use. All tobacco curricula share the goal of preventing uptake of smoking and ‘denormalizing’ tobacco use. Outcome measures that capture aspects of these goals are suited to assess the impact of prevention curricula.

3.2. Data sources

The following sections describe the sources that provided data for this research.

3.2.1. Student-level data (and aggregated school-level outcomes)

Self-report data on tobacco-related behaviors and attitudes came from the Arizona Youth Survey (AYS). This student-level data represents the “backbone” of this entire research. The number of schools participating in AYS 2006, the large student sample size, and the wealth of psychosocial data make this database a unique opportunity for exploring possible prevention effects of BTEP’s programming. The survey used here was administered by the Arizona Criminal Justice Commission (ACJC) from January to April 2006 to 8^{th}, 10^{th} and 12^{th} graders in 362 Arizona public and charter schools (Arizona Criminal Justice Commission, 2006). The 2006 fielding wave yielded almost 72,000 completed surveys from students in 362 schools throughout Arizona. Selection of participating schools comprised several steps. Arizona public and charter schools were randomly chosen from lists of small, medium and large schools. These lists were stratified to represent student population by county. In addition, all schools implementing AYS in 2004 were contacted again and invited to participate again. Finally, additional schools could choose to opt in. Active and passive consent was obtained from students’ legal guardians, based on the districts’ policies. Passive consent was more frequently used. The protocol for administering staff contained an introduction to be read aloud to participating students assuring confidentiality and voluntariness of participation.

The final dataset contained responses of 60,401 students because pupils who failed to validate their honesty through reasonable responses to 'honesty indicators' (e.g. by reporting use of fictional drugs or impossible high combinations of drug use) or who admitted dishonesty were excluded from further analysis. The resulting distribution of students by grade and county revealed very good fit with enrollment statistics, making a weighting procedure to reflect county enrollment unnecessary.

As BTEP requires contractors to deliver curriculum-based services to students in grades four to eight exclusively, analyses of this study will focus on responses of eighth graders in AYS 2006 only. Of all students providing data per AYS 2006, students in 8^{th} grade are most proximal in time to possible BTEP interventions. 10^{th} and 12^{th} graders are at least two and four years, respectively, removed from services. Detecting an effect after such a long period of time was shown to be highly unlikely with even the most sensitive and rigorous designs in the context of a strong intervention (Peterson, Kealey, Mann, Marek, & Sarason, 2000).

The AYS survey questionnaire is build on a risk and protective factor model, based on work of J. David Hawkins and Richard F. Catalano (e.g. Hawkins, Catalano, & Miller, 1992). AYS measures sixteen risk and protective factors with an average of four items. No further explanation of AYS content will be given here as this research will only utilize selected measures from the instrument and is not concerned with psychometric underpinnings of the suggested factors. A multitude of demographic items and detailed information about substance use, including tobacco, were also collected by the survey. Some demographic items were used as student-level predictor variables. Details of items and measures used can be found in chapter 3.4 (p.51).

3.2.2. Intervention data (school level)

AYS contained no information about if (or how much) individual students received BTEP-sponsored prevention services, or education on preventing risk behaviors at all. Instead, intervention data came from databases maintained by ERDU. Databases were created from responses to paper-pencil forms that teachers & program staff administering the curricula had to complete and submit to ERDU for further processing. Forms contain information about the instructor, the curriculum used, number of students who received prevention programming, number of sessions taught and average number of minutes per session, and further items (see Appendix A). Items and algorithm used to create the BTEP Prevention Index can be found in chapter 3.4.7 (p.56).

3.2.3. Structural school data (academic achievement & poverty)

Information about school-level measures, 'poverty' and 'academic achievement', was gathered from Arizona Department of Education (ADE) records. ADE provides detailed public records (on their website http://www.ade.az.gov/edd/) about structural school characteristics such as enrollment figures, ethnic composition, rate of students on free or reduced lunch as well as a six-step scoring system that integrates various academic achievement measures and academic progress indicators into one coding system (so-called AZ LEARNS system). All these measures are captured in School Report Cards.

3.3. Data structure, sample size and power analysis

Student-level data from AYS was nested in multiple ways. Students were nested within classrooms, classrooms within grades, and grades within schools. Strictly, there are two more levels of nesting. Schools are nested within districts and districts are nested within counties. Nesting within school districts could not be addressed with available data. In addition, implementation parameters did not suggest any influence of district-level factors on outcomes. County nesting was addressed by running all analyses for all available schools and Maricopa County schools only (explanations for this decision are given further down).

Nested data structures render conventional OLS methods inappropriate because of the potential violation of independence of observations and correlated error terms (Palmer, Graham, White, & Hansen, 1998). An expression of homogeneity of elements of the same group (within-variability) in relation to variability between groups is given by the intraclass correlation ICC (Snijders & Bosker, 1999).

In clustered or nested data structures, statistical coefficients such as slopes, intercepts and their random counterparts as well as statistical power issues are embedded in a context of at least two sample sizes: the sample size at level-1 (individual) and level-2 (number of groups or clusters, in this research schools). All constraints for reliability of coefficients known from OLS methods pertain even more to multi-level modeling (Bickel, 2007). A recent simulation study on sample size in multi-level modelling confirmed that a maximum level-2 sample size together with a maximally large level-1 sample size is desirable. Level-2 sample size should ideally exceed N=50 (Maas & Hox, 2005), i.e. more than 50 schools for this study. In order to obtain an approximation of the statistical power of this study, the free-of-charge tool "Optimal Design" (Liu, Spybrook, Congdon, Martinez, & Raudenbush, 2008) helped the author estimating statistical power under the given parameters of this study. Optimal Design was developed to calculate optimal sample sizes at various levels when intact social groups (such as classrooms, clinics etc.) are randomly assigned to interventions instead of individuals. This was clearly the case in Arizona's school-based tobacco prevention interventions where entire classrooms in a selected school were given anti-smoking curricula. Table 3 displays possible data structures based on minimum quota of observations within a given school. Computations are based on cleaned AYS 2006 dataset, 8th grade students only. Optimal Design produced estimations of statistical power under the following parameters: significance level alpha=0.10, effect size ES (difference in 30-day smoking prevalence between control and intervention schools) = 2%, lower boundary of smoking prevalence

=1%, upper boundary of smoking prevalence = 54%. These parameters were based on preliminary analysis of pertaining data (more details can be found in sections 4.1.1 and 4.2.1, p.89). Results are given in Table 3 for all possible data structures and in Figure 6 below.

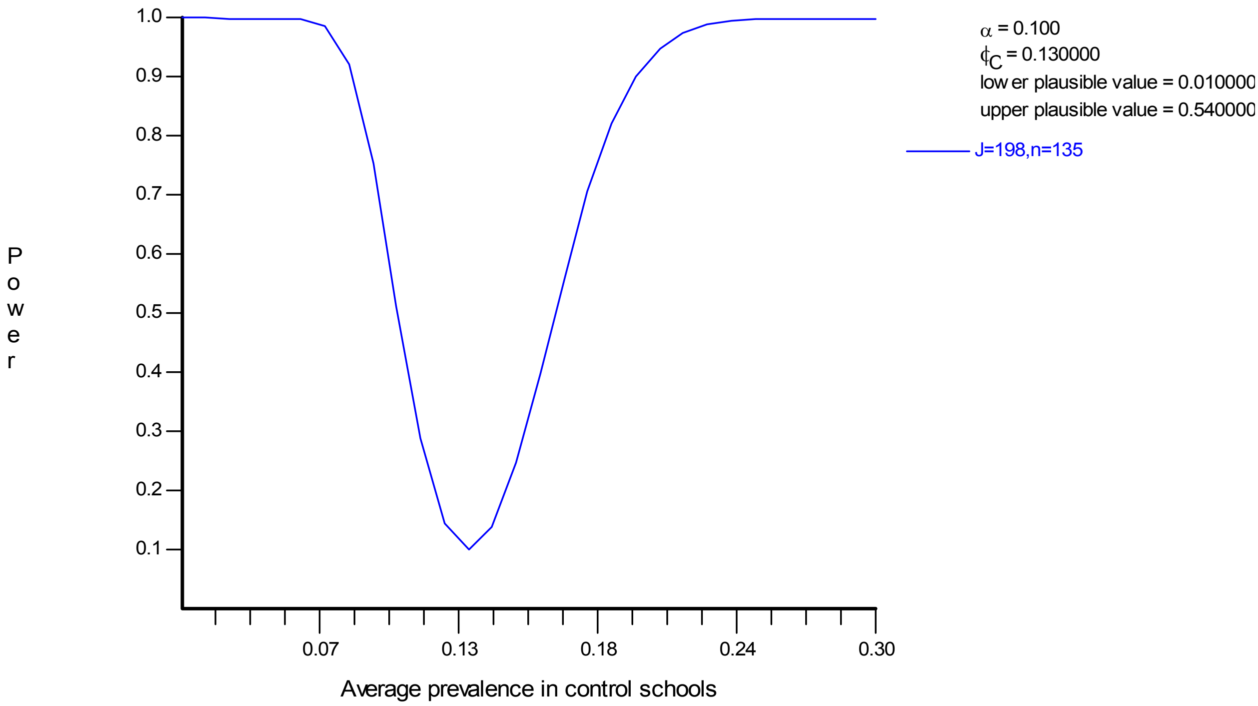

Figure 6. Statistical power distribution of final data structure, with parameters from preliminary findings

As can be seen in Figure 6, statistical power is inversely related to the effect size, i.e. the difference between prevalence in control schools (13 percent) and intervention schools. The farther away the prevalence of intervention schools is from 13 percent, the more likely it becomes to detect the effect with the given parameters of this study. The tightness of the V-shape of the curve in Figure 6 suggests that at only plus or minus five percent prevalence difference, power would be about 0.8! Optimal design calculations (see Table 3 below) and inspections of changes in the graph upon entering different parameters yielded the following important insights: a) An effect size of only two percent between control and intervention schools would have been hard to detect with the usual and entirely arbitrary alpha-level of *p*<= 0.05. Therefore, all hypothesis testing proceeded with alpha-level of *p*<= 0.10. To obtain the customary power of 0.80, the effect size would need to be about twice as large or 4% absolute prevalence difference between intervention schools and control schools. Other outcome measures may exhibit greater effect sizes and may therefore be detected with alpha-level of *p*<= 0.10.

Table 3. Possible data structures with increasing minimum N per school

Minimum # of observations within a school	**Remaining # of schools (level 2 sample size)**	**Remaining # of students (level 1 sample size)**	**Mean # of observations within a school (SD)**	**Statistical power if ES=2%, α=0.1 (0.2), prevalence(control group)=0.13**
10	198	26642	135 (118)	0.29 (0.43)
20	183	26438	144 (117)	0.28 (0.41)
30	162	25916	160 (116)	0.26 (0.39)
50	139	25017	180 (113)	0.24 (0.37)
100	92	21664	235 (100)	0.20 (0.32)

b) As can be seen in Table 3, level-2 sample size decreases with increasing minimum quota for student *N* within a given school. This means that some schools provided only very few students in 8th grade. However, even with the strictest quota of 100 students per school, the level-2 sample size does not fall below the recommended minimum number of 50.

c) Statistical power of detecting a 2% difference in prevalence drops from 0.29 to 0.20 with alpha=0.1 or by about 30%. No notable change could be observed when changing within-cluster sample size (i.e. average number of students within a school) and holding cluster sample size (i.e. number of schools) constant.

Together, these findings demonstrate that losing clusters (i.e. schools) had much more detrimental effects on statistical power than losing level-1 sample size (i.e. students). When comparing the first row of Table 3 with its last row, it appears that if higher demands were placed on minimum sample size per cluster (column 1), 53% of level-2 sample size got lost (from 198 to 92, column 2) but only about 19% of level-1 sample size (from 26,642 to 21,664, column 3). Graphical inspection of changes in Figure 6 buttressed these conclusions, as kurtosis of the u-shape became smaller (i.e. the width of the U greater) only if level-2 sample size shrunk, which indicates reductions of power. It remained virtually unchanged when average within-cluster N was reduced (moving bottom up column 4 in Table 3). In summary, data structure of the first row in Table 3 was selected because of its desirable implications for study power (level-2 $N = 198$ and level-1 $N = 26{,}642$).

Related to the issue of sample size were conclusions about representativeness of the final sample, which constituted a "convenience sample" after all, albeit a very large one. In order to be able to assess population coverage of the final sample of the population of eighth grade students in the state, Table 4 displays the coverage of the county student population in the dataset and the number of schools that contain these students. Coverage rates of the entire student population vary from 21.3 percent in Mohave County to 72 percent in Greenlee County. Overall, coverage for all counties by far exceeds 5 percent, which constitutes the threshold for applying finite population correction factors that account for the gain in precision from large samples of finite populations. In general, all statistics derived from this final dataset can be regarded as highly precise and representative for the finite population (student population per county) as a whole.

Table 4. Coverage of eighth grade student population by county

County	Enrollment (column %, per ADE)	# of students in final level-1 dataset (column %)	Coverage of entire student population by final level-1 dataset	# of schools that contain level-1 student samples	# of schools that were BTEP intervention schools (%)
Apache	1,223 (1.5%)	355 (1.3%)	29.0%	5	1 (20%)
Cochise	1,770 (2.2%)	837 (3.1%)	47.3%	10	3 (30%)
Coconino	1,604 (2.0%)	469 (1.8%)	29.2%	5	2 (40%)
Gila	709 (0.9%)	360 (1.4%)	50.8%	6	3 (50%)
Graham	462 (0.6%)	258 (1.0%)	55.8%	3	0 (0%)
Greenlee	118 (0.1%)	85 (0.3%)	72.0%	2	2 (100%)
La Paz	214 (0.3%)	145 (0.5%)	67.8%	5	5 (100%)
Maricopa	49,698 (61.6%)	17,113 (64.2%)	34.4%	98	50 (51%)
Mohave	2,148 (2.7%)	457 (1.7%)	21.3%	6	4 (66%)
Navajo	1,738 (2.2%)	484 (1.8%)	27.8%	4	0 (0%)
Pima	11,919 (14.8%)	2,542 (9.5%)	21.3%	21	18 (86%)
Pinal	2887 (3.6%)	1,069 (4.0%)	37.0%	13	6 (46%)
Santa Cruz	884 (1.1%)	341 (1.3%)	38.6%	2	0 (0%)
Yavapai	2,205 (2.7%)	609 (2.3%)	27.6%	9	3 (50%)
Yuma	3,063 (3.8%)	1,518 (5.7%)	49.6%	9	1 (11%)

The table also shows that some counties provided only ‘control schools’, i.e. schools that did not receive BTEP sponsored tobacco prevention (Graham, Navajo, and Santa Cruz County). Findings from overall, statewide analyses will therefore not be generalizable to these counties, as no intervention data could be matched to available outcome data. These counties did, however, provide prevention services (Coffman, 2007). It can also be seen that only Maricopa County provided enough schools with and without BTEP prevention services that would statistically justify separate analyses for Maricopa only. Further, available documentation on implementation guidelines for curriculum administrators in Maricopa revealed that schools who received funding through Maricopa County Tobacco Use Prevention Program (MACTUPP) had to conduct activities that exceed the actual classroom lessons. Some of these activities reached out to the entire school. Appendix B contains this document. Subsequent analyses will therefore be split into statewide, overall analyses and for Maricopa only (because of statistical feasibility and Maricopa’s ‘comprehensive’ approach that was shown to be most effective, i.e. most likely to be detectable with quantitative analyses).

3.4. Measures

This section outlines what particular measures have been utilized from the above-mentioned sources to populate the models introduced in chapter 2.6.

3.4.1. Individual-level (person) measures

In analytical terms, this group of variables consists of ‘control’ variables and outcome variables, depending on their status in the model.

3.4.2. Independent variables

Because this evaluation focused on the potential effects of curriculum-based prevention programming as a characteristic of the school, individual-level variables were only introduced to control for potentially confounding influences. Table 5 specifies measures of individual-level characteristics available from AYS that needed to be controlled for, as was pointed out in chapter 2. Some theoretically derived level-1 confounders were not available from the AYS (e.g. SES of the family & student’s pocket money/income) and could therefore not be included in further analyses.

Table 5. Potentially confounding level-1 predictors

Construct	Measures (Items)	Response categories
Parental smoking	Does anyone who lives with you now smoke cigarettes? (Only parents were regarded)	No one who lives with me now smokes cigarettes, a parent (or guardian)[responses not considered: a brother or sister, another adult who lives with us, another young person who lives with us]
Parental attitudes towards smoking	How wrong do your parents feel it would be for you to smoke cigarettes?	Very wrong, wrong, a little bit wrong, not wrong at all
Attachment to the family	Do you feel very close to your mother?	NO!, no, yes, YES!
	Do you share your thoughts and feeling with your mother?	NO!, no, yes, YES!
	Do you feel very close to your father?	NO!, no, yes, YES!
	Do you share your thoughts and feeling with your father?	NO!, no, yes, YES!
Age	How old are you?	10 or younger, 11, 12, 13, 14, 15, 16, 17, 18, 19 or older
Gender	Are you male or female?	(checkboxes)
Ethnicity	Please choose the ONE answer that BEST describes what you consider yourself to be.	White, not of Hispanic Origin Black or African American American Indian/Native American, Eskimo, or Aleut Spanish/Hispanic/Latino Asian Pacific Islander Other (Please Specify____)

Questions of scale construction and data reduction are addressed in section 4.1 (p.69).

3.4.3. Outcome variables

It is important to obtain symmetry between goals of social interventions and outcomes measured in an outcome evaluation (Wittmann et al., 2006; Wittmann et al., 2002). Otherwise, nil findings could easily result from a mismatch between factors that were truly improved but were not captured due to measurement of inappropriate outcome measures. According to Wittmann et al. (2002), it is also desirable to utilize multiple outcome criteria for the broad behavioral and cognitive domains that an intervention aims to modify, so that the evaluation study arrives at fair conclusions. This research considers a series of outcome measures that were supposedly addressed by prevention curricula most widely in use among

Arizona prevention providers. Preferably, multivariate outcome indices rather than individual items ought to be construed, as Figueredo and Sechrest (2001) have suggested. Table 6 lists level-1 outcome measures captured in AYS 2006. Items on social acceptability of smoking were examined for their suitability to form a scale in order to obtain more parsimonious models. Results are presented in 4.1.3, (p.72).

Table 6. Individual-level outcome measures

Construct	Measures (Items)	Response categories	Analytical manipulation
Self-reported smoking	"Have you ever smoked cigarettes?"	Never, once or twice, once in a while but not regularly, regularly in the past, regularly now	Binary variables for ever smoking & current smoking. Index variable that integrates all three items into a Smoking Acquisition Index
	"During the past 30 days, on how many days did you smoke cigarettes?"	0, 1 or 2, 3 to 5, 6 to 9, 10 to 19, 20 to 29, all 30 days	
	"During the past 30 days, on the days you smoked, how many cigarettes did you smoke per day?"	I did not smoke during the past 30 days, less than 1 cig./day, 1 cig./day, 2-5 cig./day, 6 to 10 cig./day, 11 to 20 cig./day, more than 20 cig./day	
Social normativity of smoking	"What are the chances you would be seen as cool if you smoked cigarettes?"	No or very little chance, little chance, some chance, pretty good chance, very good chance	Ad-hoc scale (sum score of standaridzed items)
	"How wrong do you think it is for someone your age to smoke cigarettes?"	Very wrong, wrong, a little bit wrong, not wrong at all	
	"Sometimes we don't know what we will do as adults, but we may have an idea. Please answer how true these statements may be for you. When I am an adult, I will smoke cigarettes."	NO!, no, yes, YES!	
Peer smoking	"Think of your four best friends (the friends you feel closest to). In the past year (12 months), how many of your best friends have smoked cigarettes?"	0, 1, 2, 3, 4, (# of friends)	none

Items on self-reported smoking constitute the most important outcome measures, as school-based smoking prevention programs ultimately seek to reduce smoking. In this study, peer smoking was used as another outcome measure because it is suited to triangulate findings on self-reported smoking for which it constitutes a proxy (see above).

3.4.4. School-level ('ecological') measures

Diez-Roux (1998) distinguished between two kinds of higher-order measures that can be incorporated in multi-level analyses. So-called 'derived' measures

> *"Summarize the characteristics of individuals in the group (means, proportions, or measures of dispersion; for example, percentage of persons with incomplete high school educations) [...]" (p.218)*
>
> .

These variables have an equivalent on the individual level but represent a different construct when aggregated to a higher level (Schwartz, 1994). For example, receiving prevention education as an individual may ideally instill negative attitudes towards smoking in that given person, but if provision of prevention curricula is aggregated to the school level, it may represent a proxy for anti-smoking activity in a school (Palmer, Graham, White, & Hansen, 1998).

Secondly, so-called "integral" (or global) measures have no equivalent on the individual level and therefore cannot be computed by aggregating level-1 variables. These variables constitute integral features of the object they describe. An example would be geo-physical properties of a given environment, such as availability of public transportation or play grounds in a given neighborhood. In this study, all dependent variables on the school level fall under the category of derived measures as they were all computed by aggregating student-level information. However, in the case of the most crucial level-2 independent variables (academic achievement, poverty and BTEP prevention index), they were not calculated from self-report measures which eliminates all methodological issues of self-report (social desirability, recollection issues etc.).

3.4.5. Academic achievement

School-level academic achievement can be conceptualized in different ways. One possibility is to figure out what proportion of the student body of a given school has passed or failed a certain cut-off on a standardized achievement test (Aveyard et al., 2004). The benefit of such achievement aggregates is their high accuracy. As Snijders and Bosker (1999) have pointed out, the precision of aggregated statistics (such as group means or percentages) increases with the number of observations per macro-unit. Mostly, such school level achievement indices entail the entire (or vast majority) of the student body and therefore have minuscule standard errors of means. This study employs a classification scheme provided by the Arizona Department of Education (ADE), the so-called AZLEARNS system (Arizona Department of Education, 2007). The AZLEARNS rating system assigns six-step ratings to all AZ public and charter schools based on performance scores of their student body. AZLEARNS computes an aggregate school-level rating by integrating a variety of academic performance measures. For example, scores on the Arizona Instrument to Measure Standards (AIMS) - a standardized test system to assess student knowledge in math, reading, writing and science - flow into the computation of the AZLEARNS index. Further, the AZLEARNS system takes into account the deviation scores of individual students from their predicted score on AIMS subtests, aggregated to the school level across students and study subjects in order to create a school "progress" score. Final AZLEARNS designations are "Excelling" (coded 6), "Highly Performing" (coded 5), "Performing Plus" (coded 4), "Performing" (coded 3), "Underperforming" (coded 2) and "Failing to Meet Academic Standards" (coded 1). Scores in this study came from the school year 2006/07.

3.4.6. Poverty

A commonly accepted and widely used measure of school-level poverty is the percentage of students who are eligible for free or reduced lunches (Battistich, Solomon, Kim, Watson, & Schaps, 1995). The percentage of students eligible for free or reduced lunches in the school year 2006 was therefore chosen as measure of school-level poverty. The precision, i.e.

representativeness of this indicator for a school population is very high because it stems from administrative records that encompass the entire student body.

3.4.7. BTEP Prevention Index

The main purpose of the proposed study was examining to what extent BTEP prevention programming affected tobacco-related norms & behaviors while controlling for potential level-1 and level-2 covariates. As indicated in the summary of meta-analytical findings, it can be expected that implementation of a comprehensive tobacco prevention curriculum in conjunction with other tobacco control policies (such as smoking bans and strong media campaigns) can have beneficial effects on students' attitudes and behaviors. Effects may occur in terms of reducing the very behavior of smoking, or attitudinal precursors, such as the perceived acceptability of smoking. What lends credibility to conceptualizing treatment variables on the group level is the fact that the biggest provider of school-based tobacco prevention services, Maricopa County Tobacco Use Prevention Program (MACTUPP), requires their contracting schools to undertake activities that reach out to the entire school environment (see Appendix B for a full list of requirements). For example, in addition to implementing tobacco prevention curricula, schools have to post tobacco-free signs, disseminate anti-tobacco messages to the entire school via the school newspaper etc., and conduct one student-led anti-tobacco event. Therefore, the author had strong reason to assume that BTEP presence in a Maricopa County school could be seen as proxy for anti-smoking intervention measures going beyond just implementing lesson-based curricula. About 50% of the schools in the dataset and 64% of the student sample nested within these schools came from Maricopa County. Consequently, all models were built for all schools available per outcome data and for Maricopa schools only, to fully exploit all capabilities of the data.

Generally, there are different ways to conceptualize prevention measures. Rohde et al. (2001) have computed a composite score to evaluate Oregon's tobacco prevention intervention. Implementation of tobacco use prevention curricula was one component of their composite measure. ERDU's data collection captured intervention data for a given school as outlined in the table below.

Table 7. Exemplary breakdown of school-level program data from a real school

School year	Grade	Curriculum	# of students	# of lessons	# of minutes per session
2004-05	5	Eglin Longhorn	27	9	60
2004-05	7	Project Alert	16	11	60
2005-06	7	Project Alert		11	60
2005-06	5	Eglin Longhorn	27	9	45
2006-07	5	Eglin Longhorn	20	9	45
2006-07	7	Project Alert	24	11	60

For intervention data, missing data was imputed as unconditional mean of all available column data. Sechrest et al. (1979) recommended the operationalization of strength of a treatment, as opposed to simple dichotomous indicator coding. The final BTEP Intervention Index can be thought of as 'average per-person-curriculum-minutes' across all curricula over the course of two years. It includes an indicator of 'intensity of contact' (number of students receiving interventions) with an indicator of length of interventions (minutes of a curriculum) with a measure of amount of one curriculum (number of lessons). This index was computed as follows:

$$\text{BTEP Prevention Index} = \sum (n_{ij} * l_{ij} * min_{ij}) / N_{2006}$$

n_{ij} = Number of students in curriculum i in year j

l_{ij} = Number of lessons of curriculum i in year j

min_{ij} = Number of minutes per lesson of curriculum i in year j

N_{2006} = Number of students enrolled in year 2006 in a given school

Division of prevention intensity by enrollment size takes into account that larger schools have more opportunities to educate a greater number of students than smaller schools. Without adjusting for enrollment size, larger schools would be implicitly weighted stronger. This is especially important because Maricopa County Health Department, which serves the largest student population in the state, requires implementation of curricula in entire grade levels (e.g. to all eighth graders in a school, see Appendix B).

The final BTEP prevention index will be a sum of all curriculum-specific indices within a school over two fiscal years (FY 2004-05 and FY 2005-06). This index realizes the investigator's goal of obtaining a single program performance index, aggregated over time, grades and curricula. Further, this intervention index realized Wittmann's (Wittmann et al., 2006; Wittmann, 1990) stipulation for treatment measures that have reliability higher - or rather more real - than traditional dummy codes to indicate treatment and control conditions. This research applied this concept to the school level. As Wittmann (1990) has pointed out, it is neither practical nor psychometrically justifiable to employ treatment measures with biased reliabilities. This happens by assuming that every element in the treatment group received the same amount of intervention (coded "1") and everybody in the control group received nothing (coded "0"). This leads to a rather bizarre situation of treatment reliability of one and may lower study power considerably. The BTEP prevention index provides a rather fine-grained measurement of the amount of prevention efforts put into one school with superior psychometric properties than dummy variables. The final BTEP Intervention Index will only allow estimation of an overall effect, but no inferences regarding theoretically relevant mediation processes, such as what in a given curriculum yields an effect or if any given curriculum outperforms another one.

3.4.8. School-level outcome measures

All school-level outcome measures were aggregated as school-level means from individual-level data and therefore constitute 'derived measures' (Diez-Roux, 1998). For variables that constituted a scale on the individual level (i.e. three items representing the 'normativity of smoking' scale), the scale values were aggregated as school means and utilized as level-2 outcomes. As mentioned earlier, level-1 measures that become aggregated to a higher level may undergo a conceptual change. Table 8 displays level-1 outcome variables and their new meaning as school means.

Table 8. School-level outcome measures (aggregated level-1 outcome measures)

Variable label	Content of level-1 item	Scale level before aggregation	Construct of aggregated level-1 item
ciglife	"Have you ever smoked cigarettes?"	dichotomous	Rate of students who have ever smoked
q81	"During the past 30 days, on how many days did you smoke cigarettes?"	Dichotomized (cig30dy)	Rate of students who currently smoke
		continuous	Average frequency of smoking among current smokers
q82	"During the past 30 days, on the days you smoked, how many cigarettes did you smoke per day?"	continuous	Average amount smoked per day
q25a	"What are the chances you would be seen as cool if you smoked cigarettes?"	Continuous items integrated as scale	Average of smoking as socially normative behavior
q27g	"How wrong do you think it is for someone your age to smoke cigarettes?"		
q51a	"Sometimes we don't know what we will do as adults, but we may have an idea. Please answer how true these statements may be for you. When I am an adult, I will smoke cigarettes."		
q24b	"Think of your four best friends (the friends you feel closest to). In the past year (12 months), how many of your best friends have smoked cigarettes?"	Continuous	Average number of smoking friends

3.5. Hierarchical linear modeling as data analytical tool

A brief abstract of the basic concept of HLM will follow. Details can be found in either of the many introductory textbooks such as Bryk & Raudenbush (1992), Snijders & Bosker (1999) or Goldstein (2007). Notation used here is following Bryk & Raudenbush (1992).

Hierarchical linear modeling encompasses a whole family of statistical models rather than one specific tool. Models could be ranked according to increasing complexity, starting with the one-way ANOVA with random effects, where only the intercepts of the level-2 units are allowed to vary randomly. This model is also called variance component model or random intercept model, because it allows the decomposition of residual variance into a within-group component and a between-group component. Fitting such an "empty" model allows computation of the so-called intraclass correlation (ICC) and is customarily the first step in determining if nesting of level-1 units led to non-ignorable homogeneity in outcome variables. At least two reasons necessitated the use of hierarchical linear modeling (HLM) - a more advanced data-analytical tool than standard ordinary least-squares regression (OLS) - in this study. The more important one derives from the fact that intervention data was unavailable at the individual level, but was constructed as a school characteristic. While a one-way random effects ANOVA is yielding first evidence that grouping may be influential for the outcome under study, it is limited because it cannot parse out what about the grouping accounts for the differences. Using HLM methodology enabled the author to link individual-level outcomes to school-level predictors, with amount of BTEP prevention activity as one of those school-level predictors. Using HLM in this way is also superior to just applying OLS models to level-1 outcomes aggregated to the school level. This was done, however, as a preliminary step (see section 4.2, p.87).

The second reason for using HLM was the fact that students were nested within schools and therefore observations of students within the same school could not be assumed independent from each other, a fundamental assumption of OLS methods (Palmer, Graham, White, & Hansen, 1998; Murray & Hannan, 1990; Bryk & Raudenbush, 1992; Snijders et al., 1999). Ignoring nested-data structures when they matter, i.e. when observations within groups are interdependent, leads to deflations of type-I errors (i.e. increases the chance of erroneously rejecting the null hypothesis). In other words, statistical significance (especially for level-2 predictors) will be overestimated when nesting is ignored. The deflation of the real alpha level results from a shrinkage of *effective sample size* on which tests of statistical significance are performed. The effective sample size is a function of the number of level-2 units (schools), level-1 units and the intraclass correlation (see below). The intraclass correlation (ICC) is a measure for the proportion of outcome variance that can be attributed to the students' clustering in schools. A non-trivial ICC indicates correlated residuals, a major violation of OLS prerequisites. The ICC (also denoted by the Greek letter rho ρ) is computed according to Equation (1):

$$\rho = \frac{\tau}{\tau + \sigma^2} \qquad (1)$$

with σ^2 = within-group variance and τ = between group variance.

Computation of the ICC also reveals that HLM methods are a class of tools based on the decomposition of variance. The simplest model in the HLM family, the random effects ANOVA is based on the assumption that the level-1 outcome variable Y_{ij} of a person i in group j can be expressed by three additively combined components: the grand mean of the population (γ_{00}), the unique effect of group j (u_{0j}) and the person-level effect (r_{ij}).

$$Y_{ij} = \gamma_{00} + u_{oj} + r_{ij} \qquad (2)$$

This model is called random effects ANOVA because both u_{0j} and r_{ij} are thought of stemming from random variables with mean of zero and finite variance. The variance of the group-specific effect u_{0j} is defined as τ (or τ_{00}) and the variance of the person effect is defined as σ^2. Bryk and Raudenbush (1992) refer to this model also as the *fully unconditional model* because it contains no other "predictor" variables than the grouping/clustering variable.

As for effective sample size, a hypothetical example was chosen to illustrate potential shrinkage. In this hypothetical study, 1000 students were sampled from 10 schools, and the variance attributable to the clustering of students, i.e. the ICC, equaled 0.1 (which is fairly high). The effective sample size on which statistical significance tests are performed would be derived from the following formula:

$$N_{effective} = \frac{N \times n}{1 + (n-1) \times ICC} \qquad (3)$$

with N = sample size of level-2 units (schools), n = overall level-1 sample size (students).

In the hypothetical example, the effective sample size would be 1000x10/1+(1000-1)*0.1= 99. The "real" (effective) sample size in this example would have been only 99!

In this study, about 25000 students were sampled from 198 schools and preliminary findings on the ICC of the most important outcome variables indicate an ICC of about 0.04. These leads to an effective level-1 sample size of about 4950 instead of a nominal sample size of 25,000! Although this represents an effective reduction in sample size of about 80%, it would still not impact conclusions based on results of statistical significance testing. It serves as a warning to not over-interpret significance test results in this study and maybe not even consider them at all in evaluating the results of models that involve level-1 units (i.e. all HLM analyses).

A brief outline of the formulas used in two-level hierarchical linear models follows in order to explain the basic concept underlying hierarchical decomposition of variance. The standard OLS models with only one level of analysis expresses an outcome measure Y of person i with a single predictor variable as

$$Y_i = \beta_0 + \beta_1 X_i + r_i \qquad (4)$$

with β_0 as the intercept of Y, β_1 the slope of predictor X, and r_i the unique effect of person i not explained by the linear model.

Generally, r_i is conceived of as normally distributed error with a mean of zero and finite variance. In this study, one of the level-1 predictors for current cigarette smoking is parental smoking. In the simple example of Equation 4, the regression would answer the question how much more likely it would be for a student being a smoker if either of the parents smokes. Such a relationship could be examined for every school j that contributed data to the study. For the j=198 schools in this study, there would be 198 regression coefficients and 198 slopes. For a student i in a specific school j, the level-1 regression model could be expressed as follows:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij} \qquad (5)$$

with β_{0j} as the intercept of Y in school j, β_{1j} the slope of predictor X in school j, and r_{ij} the unique effect of person i in school j.

Both intercept (β_{0j}) and the slope (β_{1j}) have a certain mean and variance over all schools j. The terminology of these variances and means was already mentioned in Equation 2 and becomes now refined. The mean of intercepts is denoted by the Greek letter gamma, with γ_0 as the population mean of all intercepts ('mean of means'). The variance of intercepts, which is identical to the variability of school-specific effects already introduced in Equation 2, is denoted by τ_{00}. The mean of slopes for the first predictor is denoted by γ_1, γ_2 for the second predictor and so on. The variance of the slopes is indicated by τ_{11}. Finally, the covariance of the slopes and intercepts is labeled τ_{01} and assumed to have a bivariate normal distribution. If τ_{01} would have a negative value in the example of parental smoking and smoking risk, this would imply that in schools with a higher smoking prevalence (β_{0j}), the amplifying impact of parental smoking on risk of student's smoking (β_{1j}) would be less pronounced. Such a finding would indicate that in school with higher smoking prevalence, it would not have been a higher proportion of students coming from smoker households that brought about the higher smoking prevalences, but 'something else' to be uncovered by further examinations. A speculative explanation for these effects could be that schools with higher smoking prevalence in grade eight implemented a lower amount of curriculum-based smoking prevention programs. The formulas to explain the variability in slopes and intercepts with a measure of amount of prevention programming (W) would constitute the level-2 models. For modeling intercepts β_{0j}, the following equation is created:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j} \qquad (6)$$

with γ_{00} as grand mean of smoking prevalence rates, γ_{01} as slope of prevention programming, and u_{0j} as unique effect of school j while controlling for prevention programming W.

The model to explain in variability in level-1 slopes β_{1j} reads as follows:

$$\beta_{1j} = \gamma_{10} + \gamma_{11} W_j + u_{1j} \qquad (7)$$

with γ_{10} the mean slope parental smoking, γ_{11} the impact (slope) of prevention programming on slopes of parental smoking on smoking risk, and u_{1j} the unique effect of parental smoking-smoking risk of school j while controlling for prevention programming W.

As pointed out for Equation 2, the u-effects in Equations 6 and 7 are assumed to be random variables with zero means, variances τ_{00} and τ_{11}, respectively, and covariance τ_{01}. In order to arrive at a full hierarchical linear model, Equation 6 and 7 are substituted into Equation 5:

$$Y_{ij} = \gamma_{00} + \gamma_{01}W_j + \gamma_{10}X_{ij} + \gamma_{11}W_jX_{ij} + u_{0j} + u_{1j}X_{ij} + r_{ij} \quad (8)$$

As Bryk and Raudenbush (1992, p.15) point out, Equation 8 differs from OLS equations in various ways and parameters cannot be estimated by least squares minimization but require maximum likelihood (ML) or restricted maximum likelihood (REML) estimation. One violation of OLS prerequisites constitutes the random error term in Equation 8 '$u_{0j} + u_{1j}X_{ij} + r_{ij}$' which is not independent among level-1 observations because all students within the same school share the unique effects u_{0j} and u_{1j}.

Parameters spelled out in Equation 8 can be calculated by maximum likelihood (ML) methods or restricted maximum likelihood (REML) methods. Without going into the details, it can be said that both methods are iterative procedures that arrive at parameter estimates most likely to have brought about the observed data. With increasing sample size, these estimators become normally distributed. REML overcomes one crucial shortcoming of ML methods by correcting the degrees of freedom used to estimate the level-1 or person-effect variance σ^2 (see Equation 2) for the number of level-1 predictors used. If there are only very few predictors, this correction will be minimal. If there are more predictors, the correction will be more substantial and REML results will differ from ML findings. Likewise, if level-2 sample size is large (as in this study), ML and REML will yield comparable results. A challenging problem in HLM is specification of the final model. As is the case for most statistical model building, final models emerge as result of an iterative process driven by theoretical considerations, empirical findings and re-adjustment of subsequent models according to decisions based on findings from previous models. Only some rules of thumb exist that have to be balanced with theoretical knowledge and re-evaluated once a model was fit and results suggest certain adjustments (Snijders et al., 1999; Bryk et al., 1992; Bickel, 2007). The author proceeded according to suggestions made by a number of HLM authors (Snijders et al., 1999; Bryk et al., 1992; Bickel, 2007), separately for each outcome variable. HLM modeling begins with proper specification of the level-1 model first, using conventional OLS methods. This step will be summarized in chapter 4.1.4 (p.75). The next step involved fitting fully unconditional models to assess the appropriateness of HLM methods and to estimate the variability of outcome variables attributable to the grouping of students in schools. Then, level-1 predictors were included as fixed effects only (in various combinations to test different model specifications), with specifying a random intercept (based on the findings of differing school means, see chapter 4.2.1, p.89). The resulting statistics on level-1 residual variance and level-2 residual variance were compared to the same parameters of the fully unconditional models. To judge the overall fit of the models, the deviance statistic (usually the -2 log likelihood estimate) of the null model is compared to the deviance statistic of subsequent models, regardless of the presence or absence of level-2 predictors or cross-level interaction terms. Comparison of such nested models is achieved by computing the difference score of ML estimates of overall fit indices. The difference score between these deviance statistics has a chi-square distribution. Its statistical significance can be assessed by comparing the difference score with the appropriate chi-square score. The appropriate chi-square score is found by computing the degrees of freedom as the difference between number of parameters used for the more complex models minus number of parameters used for the less complex model (or null model). If more complex models have a substantially lower deviance statistic than the null model (or any less complex model), it means that the more complex model is a better fit of the data.

A second method to determine model fit in HLM is conceptually equivalent to "explained variance R^2 " known from OLS regression. In HLM models, the concept is proportional reduction of prediction error (Bickel, 2007; Snijders et al., 1999). It essentially gives the proportion of residual (i.e. error) variance that was reduced by entering predictor variables. This can be done for level-1 residual variance (σ^2) and level-2 residual variance (τ_0^2). Details can be found in Snijders & Bosker (1999, p. 99-105).

3.5.1. Centering of independent variables (level-1 and level-2)

In hierarchical linear models, intercepts and slopes on level-1 can be specified as randomly varying across level-2 units, thereby opening an opportunity to examine the variation in the random components with level-2 models. This creates additional sets of variance and covariance that are not encountered in simple single-level models. Especially covariance between random intercepts and slopes can pose a problem similar to multicollinearity in simple OLS regressions. A remedy to this is centering independent variables. The second reason for centering independent variables is enhancing interpretability of predictor variables in HLM models. Bryk & Raudenbush (1992) recommend choosing sensible locations for the zero point of predictors by either centering around the grand mean or the group mean (or some other theoretically justified location, not discussed here further). These are two alternatives to using raw scores. In the first case, grand-mean centering, the mean of the predictor variable of the entire sample is subtracted from each individual value of the predictor variable ($X_{ij} - X..$). In the latter case, the mean of the specific group (in this study the school) is subtracted from each individual predictor variable value ($X_{ij} - X_{.j}$). Both methods reduce the chance of having undesired random components covariance but differ in their interpretation. In statistical models of the HLM family, intercepts get more attention than in standard single-level OLS regressions and their meaning is dependent on the location of predictor variables. The choice of which centering methods to use is guided by theoretical questions the research wants to answer. If grand-mean centering is used, then the intercept for an independent variable is the best estimate of the dependent variable when all independent variables are equalized at their mean. In group-mean centering, the intercept of independent variables gives the best estimate of the dependent variable if all independent variables are equalized at their group mean. Both methods are effectively creating deviance scores with the difference of the "anchor" that is used as point of departure.

The author of this study had no theoretically driven preferences for any particular centering method. Therefore, as Bickel (2007) recommended, all independent variables (level-1 & level-2) were grand-mean centered.

3.6. Data sanity checks & statistical tools

The author compared ADE records on grade range for each school that will go into the final dataset. Four schools that provided data from eighth graders via AYS 2006 did not serve eighth graders in the school year 2005-6 when AYS was fielded, which reduced the initial school sample size of 202 to the final 198 schools. Therefore, these four schools and contained students were excluded from further analysis. In order to exclude students with non-sensical data, the distribution of self-reported age was inspected. The grand mean of all students was 13.6 years, ranging from 10 to 18 years. Most of all students in grade eight are between 12 and 14 years old. It is virtually impossible to be younger than 12 and older than 16 in grade eight. Therefore, students who reported ages outside of this spectrum were excluded from further analysis. This resulted in the loss of 19 level-1 cases (students).

All statistical analyses were performed with SPSS 17 or HLM 6. Additionally, missing data analyses involved usage of the free-of-charge NORM[iii] software and Optimal Design[iv] was used to examine a-priori power.

4. Results

Section four is structured according to the logic of statistical model building explained earlier. It starts with examination of variables and relationships on the student level. It then reports on analyses pertaining to the school level only and is finalized by presentation of HLM models.

4.1. Individual (student) level

This chapter contains all analyses performed in order to ensure proper preparation of the ecological model and HLM models that were constructed to address the BTEP effectiveness question and the identification of proxy measures for the school tobacco epidemic. Because level-1 data are the "backbone" of this study, careful descriptive analyses were performed, including a thorough analysis on missing data. It is the author's belief that no high-powered statistical tool such as logistic regression or HLM can ever be applied in a meaningful way without prior inspection of basic properties of the variables to be used. "Preparatory" steps such as descriptive statistics, correlational analysis and construction of scales from individual items, missing data inspection and, finally yet importantly, analyses based on graphical display of data, were performed to get a deeper understanding of the data. These procedures would ultimately lead to better decisions when performing more complex data analysis.

4.1.1. Descriptive statistics and preparatory analyses of level-1 variables

This section describes descriptive analyses performed on all level-1 variables utilized in further model building and analyses. The section was divided by analytic status of level-1 variables: chapter 4.1.2 deals with independent level-1 variables; the following chapter 4.1.3 contains findings on level-1 outcome variables. Descriptive findings were based on valid data only and did not include imputed data points of any kind. Handling of missing data presented in chapters 4.1.2 and 4.1.3, further analyses and steps to "recover" missing data are shown in a separate chapter (4.1.5, p. 79).

4.1.2. Independent level-1 variables

Table 9 shows descriptive findings around level-1 independent variables from statewide AYS data and for Maricopa (Phoenix metro area) separately.

Table 9. Descriptive statistics of level-1 independent variables

Measure	Variable name	Mean		SD		Median		% missing data	
		AZ	MC	AZ	MC	AZ	MC	AZ	MC
Sex	q1	---	---	---	---	---	---	2.6	2.7
Age	q2	13.6	13.6	0.57	0.56	14	14	0	0
Ethnicity	q4	---	---	---	---	---	---	3.4	3.2
Parental smoking	q83b	---	---	---	---	---	---	7.8	8.1
Parental attitudes towards smoking	q117b	1.20	1.2	0.55	0.54	1	1	28.7	29.3
Feel close to mother	q128	3.21	3.2	0.95	0.95	3	4	28.2	29.4
Share thoughts/feelings with mother	q129	2.78	2.78	1.03	1.03	3	3	28.7	29.9
Share thoughts/feelings with father	q131	2.42	2.44	1.07	1.07	2	2	29.7	31.0
Feel close to father	q135	2.82	2.85	1.11	1.10	3	3	30.9	32.0

As can be seen in Table 9, the entire sample differs only marginally from the sample of Maricopa County alone in their statistics of level-1 independent variables. This similarity comes as no surprise because 64 percent of all student data came from Maricopa County. Overall, gender was evenly distributed in the level-1 sample (49% male). Ethnic groups were represented as follows: White, not of Hispanic origin 43.5%, Hispanic/Latino 39%, Black/African American 4.4%, American Indian/Native American 5.5%, Asian 2.2%, Pacific Islander 0.6%, Other 4.8%. In order to make the ethnicity variable more manageable analytically and better interpretable, four dummy variables were created to represent White (non-Hispanic), Hispanic (non-White) American Indian and "Other" ethnic categories. This representation effectively collapsed the remaining categories (African American, Asian, Pacific Islander and Other) into a new "Other" category. Of all students, 28% reported that some of their parents smoked. Again, findings for Maricopa alone did not notably differ from the general sample. A correlational analysis of all level-1 predictors yielded no correlations of noteworthy size with exception of the four variables on attachment to parents that correlated substantially with one another, but not with any of the other predictors. Judgments about substantiality of correlations were based on effect size guidelines suggested by Cohen (1992) rather than significance level. With sample sizes well over 20,000, even correlations of trivial magnitude (such as r =.02) result in significance levels of p<0.001 which renders significance levels useless as statistical decision criterion. Table 10 shows that the size of the correlation justified creating a scale out of these four indicators that reflects the underlying construct of "attachment to parents". Although items correlate much more strongly depending on which parent they refer to, a uniform scale was constructed rather than a separate one for attachment to father and attachment to mother because no theoretical considerations suggested differences in bonding to father or mother with respect to smoking behavior. In other words, the "attachment to parents" scale was used in a more heuristic way because prior research suggested it as a potentially useful level-1 confounder when doing multi-level modeling with emphasis on school-level influences. Its actual relationship to smoking was of no immediate interest in this research.

Table 10. Correlation matrix of variables on attachment to parents (significance level)

	Feel close to mother?	**Share thoughts and feelings with mother?**	**Feel close to father?**
Share thoughts and feelings with mother?	0.69***		
Feel close to father?	0.31***	0.28***	
Share thoughts and feelings with father?	0.26***	0.39***	0.748***

A reliability analysis yielded a Cronbach's alpha of 0.76 which can be considered acceptable for the heuristic purposes of the scale. In order to avoid problems in scaling, all items were standardized (z-transformation) before the scale score was created as mean of the four individual items.

4.1.3. Dependent level-1 variables

Table 11 reveals that most outcome variables are highly right-skewed, i.e. the majority of students were concentrated in the lowest range of the variables' response categories, as suggested by means and medians equaling the lowest endpoints of response options. Standard deviations show that the highest variability was found for age of first puff (q26b) and number of best friends who smoked (q25a). The lowest variability was found for students' anticipation of their smoking behavior as adults (q51a) and the composite measure "Smoking Acquisition Index" (SAI).

Table 11. Descriptive statistics of level-1 outcome variables

Measure	Var. name	Mean		SD		Median		Minimum		Maximum		% missing data	
		AZ	MC	AZ	MC	AZ	MC	AZ	MC	AZ	MC	AZ	MC
Ever use of cigarettes (mean gives overall prevalence)	ciglife	0.31	---	---	---	---	---	---	---	---	---	8.9	9.1
30-day use of cigarettes (mean gives overall prevalence)	cig30dy	0.10	---	---	---	---	---	---	---	---	---	9.4	9.6
Seen as cool if smoked	q25a	1.64	1.64	1.04	1.03	1	1	1	1	5	5	3.9	3.9
How wrong to smoke	q27g	1.59	1.56	0.88	0.86	1	1	1	1	4	4	9.3	8.9
When adult will smoke	q51a	1.41	1.38	0.7	0.68	1	1	1	1	4	4	6.1	6.0
# of best friends who smoke	q24b	0.76	0.72	1.25	0.72	0	0	0	0	4	4	3.4	3.5
Smoking Acquisition Index (SAI)	SAI	1.35	1.31	0.97	0.93	1	1	1	1	7	7	11.6	11.8

Further, only about 31% of all students had ever tried cigarettes in their life. The vast majority of students did not report smoking during the past 30 days (90%) and disapproved of smoking, a rather typical finding in this age group. Based on considerations around outcome measures for smoking prevention programs (chapter 2.4), possibilities for construction of a scale of the attitudinal aspects of smoking were explored. The goal was to reduce redundancy by moving away from item-level analyses to a more parsimonious approach of condensing individual items into meaningful scales. As can be seen in Table 12, three items on attitudinal aspects of smoking correlated sufficiently to warrant integration into a single scale. Item content suggested that the scale measures the social acceptability or perceived normativity of smoking.

Table 12. Correlations of items for "Social Normativity of Smoking" scale

	Cool if smokes cigarettes	How wrong to smoke cigarettes?
How wrong to smoke cigarettes?	0.338***	
Will smoke cigarettes when adult.	0.313***	0.532***

A reliability analysis yielded a Cronbach's alpha of 0.64, which indicates a moderate level of internal consistency. The higher correlation between anticipation of adult smoking and perceived wrongness of smoking (*r*=0.532, Table 12) than correlations of either one of those with the perceived coolness of smoking (*r*=0.338 and 0.313, respectively) seem to indicate that the item on coolness touches on a slightly different attitudinal aspect of smoking than the other two items. However, correlations are high enough to warrant computation of a single scale that expresses students' perception of how socially desirable smoking is. This score was computed as the mean of standardized (z-score) individual items.

A correlational analysis yielded moderate to high correlations (ranging from *r*=0.23 to *r*=0.57) for all continuous outcome measures specified in Table 6 (p.53). An exploratory factor analysis using Velicer's MAP method (Connor, 2000) for determining the number of factors to be extracted suggested the possibility of a single-factor solution. However, the author preferred the less parsimonious way of not extracting a single "principal component analysis (PCA)" factor for the following reasons:

1) Such a composite would collapse different aspects of smoking into a single measure that may "muddle" different psychological components (i.e. more conative vs. more cognitive aspects), of smoking that are better kept separate despite their inter-correlatedness. It is, however, beyond the scope of the current study to investigate structural relations between those different aspects of smoking as have been suggested by major psychological theories (Bandura, 1991; Ajzen, 1991). It suffices for the current study to investigate effects of independent variables (level-1 and level-2) on outcomes regardless of their structural connectedness.

2) Findings for a PCA factor that would collapse related but distinguishable aspects of smoking would be hard to interpret and even harder to meaningfully communicate to statistical lay people (such as program administrators, implementing teachers and health

educators etc.) as compared to analyzing and communicating content dimensions that are comprehensible to lay audiences and policy makers.

3) Even with a one-factor solution, standard analyses on binary outcomes of self-reported *ever smoking* and *current smoking* would have needed to be performed because (self-reported) behavior is the ultimate outcome for interventions geared at preventing negative behavioral outcomes such as smoking. Integrating (self-reported) behavior with other aspects of smoking, such as its social normativity, into a single factor would not do justice to the high importance of smoking behavior as ultimate outcome.

4) The first analytical step to approximate BTEP tobacco prevention program effectiveness was aggregating all outcome measures to the school level in order to perform ecological-level only analyses. The school mean of students' factor scores would be almost impossible to interpret given that even non-factorized variables may undergo a conceptual shift when they are aggregated as school means to represent school-level outcomes.

4.1.4. Predicting level-1 outcomes with level-1 variables

As pointed out earlier, the level-1 model served to identify important level-1 'confounders' to be included in subsequent multi-level analyses. To this end, all level-1 outcomes were regressed on the following predictors: sex, age, three dummies coding the most frequently mentioned ethnic groups (White, Latino, and American Indian), parental smoking, parental attitudes towards smoking and bonding to parents. No interaction terms were tested as no theoretical conjecture or empirical evidence suggested otherwise. Depending on the measurement scale of the outcome variable under consideration, either logistic regression or OLS regression models were used. Figure 7 shows a graphical depiction of the level-1 model(s). Dotted arrows from predictors to outcomes indicate logistic regression models; solid arrows indicate OLS regression models. Arrows connecting all outcome measures with one another indicate the non-independence of these variables (as mentioned above). A separate regression model was run for each outcome measure.

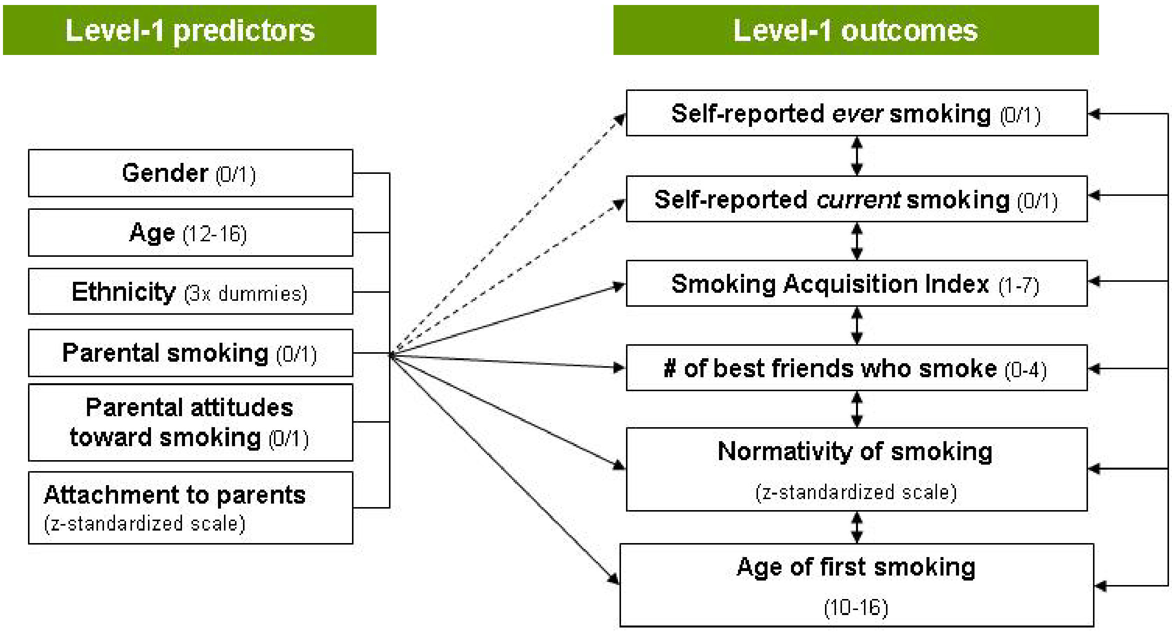

Figure 7. Level-1 regression model(s)

Table 13 displays findings for all outcome measures. A column for tolerance values was added to show (non-)redundancy of independent level-1 variables for predicting each outcome. Although tolerance values varied slightly across outcome variables, these changes were expectedly negligible and values were rounded to two decimals. *R*-square values (Nagelkerke's pseudo *R*-square for logistic models as a heuristic equivalent to OLS *R*-squares) are reported for each model. Beta weights and odds ratios are given for each predictor, respectively. Intercepts are not reported for sake of clarity. Plots of two standard regression diagnostics, i.e. histograms of residuals and Q-Q plots, were inspected to detect potential violations of OLS requirements, namely normal distribution of residuals and non-linear associations between predictors and outcome.

Table 13. Results of level-1 only regression models

Dependent Variables → / Independent variables ↓	Ever smoking	30-day smoking	Smoking Acquisition Index (SAI)	# of friends who smoke	Normativity of smoking	
	OR	**OR**	**beta**	**beta**	**beta**	**Tol.**
Sex	0.972	1.092	0.00	0.02*	0.02	0.98
Age	1.46***	1.53***	0.08***	0.08***	0.04***	0.98
Ethnicity: "White"	0.83**	0.94	0.01	-0.04***	0.02	0.37
Ethnicity: "Latino"	1.32***	1.14	-0.01	0.03*	0.05***	0.38
Ethnicity: "American Indian"	2.61***	2.39***	0.06***	0.07***	0.06***	0.72
Parental smoking	2.23***	1.95***	0.09***	0.08***	0.10***	0.96
Parental attitudes	0.44***	0.45***	-0.27***	-0.20***	-0.29***	0.96
Family bonding	0.60***	0.60***	-0.12***	-0.16***	-0.19***	0.95
R-Square (Pseudo R-Square)	0.19***	0.16***	0.13***	0.11***	0.17***	

A number of findings deserve brief discussion. For each outcome measure, the models explain at least 10% of the observed variance. Considering measurement error immanent in those predictors and outcomes, these estimates are likely an underestimation of the 'real' amount of explained variance. Without measurement error, relationships would almost certainly have been higher. Expectedly, some predictors constituted risk factors (i.e. increasing the likelihood of engaging in undesired behaviors or holding unwanted attitudes) and others protective factors (i.e. lowering the likelihood of reporting undesired behaviors or attitudes). Parental disapproval of smoking (parental attitudes) and bonding to parents were found to be strong protective factors, as indicated by their high negative parameters for continuous outcomes and odds ratios smaller than one for binary outcomes Table 13, p. 77). Each unit increase in perceived parental disapproval (e.g. moving from the response category "a little wrong" to "wrong" on the item) reduced the risk of *ever* and *current smoking* by more than half, all other things being equal. Likewise, risk for *ever* and *current smoking* was nearly halved for each unit increase in bonding to parents. For all quantitative outcomes, higher perceived disapproval of smoking by parents and higher scores of bonding to parents constituted strong protective factors against the acquisition status of smoking (SAI), having more best friends who smoke and finding smoking normative behavior. Even for analyses performed with *ever smokers* only (not shown in Table 13), a protective pattern of small magnitude was found for parental disapproval and bonding to parents. On the other hand, age, parental smoking and American Indian ethnicity constituted risk factors for all outcomes in that they were associated with more favorable attitudes towards smoking and higher odds ratios for self-reported smoking. The likelihood of reporting *ever* and *current smoking* virtually doubled if a student reported a parent was a smoker. The chances of reporting *ever smoking* almost tripled – and doubled for current smoking - if a student reported American Indian ethnicity, as opposed to reporting something else than American Indian ethnicity. Every year of age increased the chances of *ever* and *current smoking* by about 50 percent. This means that the younger a student was in grade eight, the less likely she was to report smoking. Age, parental smoking and American Indian ethnicity all exerted small but non-negligible influence on quantitative outcomes.

Sex and White and Latino ethnicity received mixed evidence concerning their status as risk or protective factor. White ethnicity appeared to be a mild protective factor, although all effect sizes were quite small. Effect sizes of sex were all negligibly small. Tolerance values were very high for all predictors (>= 0.95) except the ethnicity dummies, among which ‘White’ and ‘Latino’ were correlated r = -0.70. This results from the distribution of the categorical variable “ethnicity”, as reported in section 4.1.2 (p.70). With roughly 80 percent of students reporting either Latino or White ethnicity, these two categories became stochastically dependent. Tolerance of American Indian ethnicity dummy (0.72) was sufficiently high for inclusion into HLM modeling. Latino ethnicity appeared to proxy for unknown risk factors, however with only small effect sizes. Based on the finding presented here, the following level-1 confounders were retained for subsequent HLM model building: sex, age, the dummies coding for Latino and American Indian ethnicity, parental smoking, parental attitudes towards smoking and bonding to parents.

4.1.5. Analysis and Handling of Missing Data

In general, missing data may negatively influence a study's external and/or internal validity and therefore ought to be examined and addressed if possible. The ultimate goal of all missing data analyses is minimizing potential bias that would result from simply ignoring it (McKnight, McKnight, Sidani & Figueredo, 2007). One important reason for not simply ignoring missing data in this study was the substantial loss of level-1 sample size that would have occurred under "complete case method". Loss of level-1 cases (i.e. students) was undesirable despite the enormous sample size because some schools provided only very few students. Analyses with just complete cases could shrink the sample size of smaller schools in such a way that their effective sample size would fall below the minimum cluster size of 10 students per school, which would then result in exclusion of those schools from ecological and HLM analyses.

This section follows some principles set forth by McKnight et al. (2007). Accordingly, only level-1 variables needed for further analysis were subjected to missing data diagnostics (see Table 9, p.70 and Table 11, p.73). It was also important to keep possible retrieval or imputation methods separate for IVs and DVs because imputing missing values in DVs from IVs and then building models with the same variables would have created artificial improvement in model fit by introducing circularity.

Preparatory steps

The author followed McKnight et al.'s (2007) recommendation and computed missingness dummies for all level-1 variables of further interest so that available data on an item was coded “0” and missing data on an item was coded “1”. Likewise, missing data dummies were recoded with a procedure described in McKnight et al (2007, p.102-106) so that missingness patterns could be examined. A sum score of these recoded dummies helped identify the “cleanliness” of the missing data. In a first step, each missingness dummy was regressed on all other 15 level-1 variables to inspect potential non-ignorability of missing data due to non-MCAR patterns. The predictive power for these logistic regression models, assessed through overall pseudo R-Squares, was very low. For example, Nagelkerke’s R-Square never exceeded 0.024. Likewise, odds ratios for predictors were generally close to one, with the exception of sex and Latino/American Indian ethnicity. This may point to a MAR pattern but is not strong enough evidence of an MNAR pattern.

Amount of missing data

As can be seen, Table 9 shows descriptive findings around level-1 independent variables from statewide AYS data and for Maricopa (Phoenix metro area) separately.

Table 9 (p.70) and Table 11 (p.73), missing data of non-trivial level occurred and the amount varied over level-1 variables. Only one variable (age) had no missing data at all. The highest amounts (~30 percent) occurred for items asking about parents’ attitudes and bonding to parents. Missingness dummies of all 16 level-1 variables were summed to examine how much cumulative missing data there were across the variables for each subject. A subject with a very high sum score of missingness could be characterized as “non-compliant” subject. It seemed unwise to impute data for cases with high percentage missing (such as 90 percent). Exclusion seemed the best alternative for such cases, because data for these subjects would be largely imputed, i.e. lead to the creation of “math-based” subjects. Results in Figure 8 show that 51 percent of the sample (13674 cases) had complete data for all 16 variables of interest. The vast majority of the sample (91 percent) had a combination of six missing data points out of the 16 variables (40 percent) or less. Only five percent of students had nine (60 percent) or more data points missing.

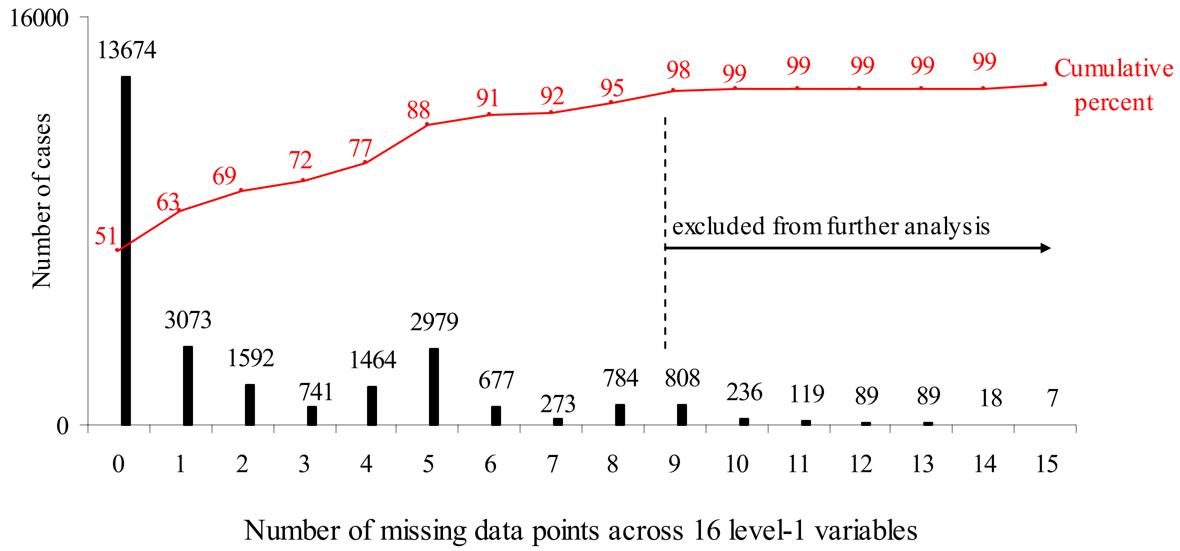

Figure 8. Distribution of cumulative missing data point across 15 independent level-1 variables

Exclusion of all cases with more than eight data points missing did not result in loss of school sample size, as the exclusion did not affect students in the smallest schools. Therefore, all these cases with more than eight missing data points were excluded from further analysis, which resulted in a loss of 1366 level-1 cases. The exclusion of these cases expectedly reduced the amount of missing data in predictor and outcome variables reported in Table 9 (p.70) and Table 11 (p.73).

Patterns of missing data

Shadish et al. (2002) pointed out that differential attrition, e.g., . missing data related to assignment to treatment or control group, is more important and troublesome than data missing at random with respect to group status. Checking associations of level-1 missing dummies with school status (control vs. intervention) is one step to approximate the existence of an MCAR pattern of missing data. In order to rule out that missing data was associated with school status, all 15 level-1 variables with missing data were cross-tabulated with school status dummy ("0" for control schools, "1" for intervention schools). In the terminology of Wittmann's (1990) five data box framework, checking the independence of missingness in predictors and school status is a variation of the $r_{PR,NTR}$ path. Relationships between missingness in outcomes and school status reflect a variation of the $r_{NTR,CR}$ path (see Figure 1, p. 30). Only minor deviations from a rectangular distribution emerged. To get a statistically defensible estimate of independence of missingness and school status, level-1 dummy variables for missingness were correlated with the school status dummy. This was done because the chi-square statistic of deviance used for comparing distributional properties between groups is highly susceptible to sample size (i.e. massively overpowered with pertinent sample size of over 25,000 students). Phi coefficients between school status and missingness dummies never exceeded trivial levels (ranging from r=0.01 to r=0.03). Together these findings demonstrate very clearly that students in control schools did not have higher degrees of missing data than students in treatment schools on any measures used in subsequent analyses. This finding is hardly surprising, as AYS survey fielding and BTEP prevention activity were not at all related, i.e. students did not complete the survey as an evaluation questionnaire in response to a BTEP prevention curriculum they had received. After ruling out associations between school status and level-1 missingness, the author continued a search for identifiable patterns of missing data with respect to items. Together with correlational findings on level-1 variables presented in chapter 4.1.2 (p.70) and chapter 4.1.3 (p.72), these insights would point to directions on how to address missing data, e.g. through replacement strategies or imputation methods.

McKnight et al. (2007) refer to the rate of missing data patterns per sample size as "cleanliness" of the missing data pattern. If every subject has a different pattern, that rate would be one and indicate very “messy” missing data. Analyzing patterns of missing data across items also helped to understand if items’ content dimensions produced missing data, e.g. though their reactive, intrusive or tiring nature. Overall, the sum score of recoded missingness dummies showed that there were 830 combinations of missing data over the remaining 25257 cases, or about 3 percent. Such a low rate indicates rather “clean” missing data pattern and may point to an MAR or MNAR pattern (McKnight et al., 2007). Correlations between missingness dummies were computed to identify the most important patterns. Three distinct clusters emerged.

First, missingness dummies related to own smoking and parental smoking were correlated about r=0.9, suggesting that most students who did not report ever smoking also did not report 30-day smoking and smoking status of their parents. This finding indicates a tendency among those individuals to refuse disclosure of information regarding smoking behavior in general. Second, a cluster emerged for dummies of bonding to parents and parental disapproval of smoking. These dummies were correlated between 0.65 and 0.89, suggesting a tendency for these subjects to not disclose information related to their parents. A third cluster could be seen among all aforementioned variables (bonding to parents and parental disapproval) and self-reported smoking and smoking status of parents. These correlations ranged around r=0.4. This third pattern indicates a general tendency for some individuals to not disclose any information on their smoking status or information related to their parents. Other missing dummies showed only modest or trivial correlations with each other, indicating that no systematic patterns occurred other than the ones just described. Such clusters of missing data across variables may point to non-ignorable missing mechanisms such as MNAR. One approximation for the existence of a more advantageous MAR pattern would be an association between missingness dummies and observed values of other variables. To that end, the author regressed all missing dummies on the remaining 15 level-1 variables. Findings suggested that the likelihood of missingness for most independent variables was higher for males and Latino students. For most outcome variables, likelihood of missing data was higher for males, students whose parents smoke, and ever smokers. These findings indicated an MAR pattern, without being able to rule out an MNAR pattern. However, overall model fit was very poor, suggesting rather week influences of these characteristics on the chance of missing data. Model fits appeared low enough to not explicitly build missingness into subsequent models, especially because these missing dummies would hardly be of practical value and could not be expected to exert a meaningful influence.

Reconstruction measures for missing data

As pointed out above, dependent and independent variables had to be handled separately to avoid circularity in subsequent model building. For independent level-1 variables that were not related to other predictors and that had high variability (sex, ethnicity, and parental smoking), retrieval or imputation methods could not be applied. These variables were, therefore, used with the irreducible amount of missing data presented in Table 14. This was, after all, not a big problem for generalizability because earlier it was shown that missing data was not related to intervention status of the school. Additionally, these three predictors did not display a pattern of missing data clearly indicative of MNAR.

Table 14. Amount of missing data in level-1 predictors before and after dealing with missing data

Item	**Missing data in initial dataset**	**Missing data after exclusion of "refusal cases"**	**Missing data after all retrieval/substitution measures**
Sex	2.6%	2.4%	2.4%
Ethnicity	3.4%	3.0%	3.0%
Parental smoking	7.8%	3.2%	3.2%
Parental attitudes towards smok.	28.7%	24.9%	0%
Bonding to parents (scale)	25%	21%	0%

The item on parental attitudes was neither related to other independent variables nor did it display high variability (Table 9, p.70). The only solution to missing data appeared to be mean/median substitution. However, mean-substituting missing data may introduce new problems, such as further reduction of variability (McKnight et al., 2007). To assess the impact of this rather "dubious" substitution method, subsequent analyses were performed with the original variable (and listwise exclusion of missing data points) and the mean-substituted variable. Results were only reported if there were major discrepancies.

Among independent variables, only the four items that constitute the scale of 'bonding to parents' were substantially correlated,; also had the highest rate of missing data. It was assumed that items would be intersubstitutable because all of them measured the same underlying construct. For all subjects who had fewer than all four of these items missing, the mean score was computed from the maximum of available z-standardized items. An irreducible core of 21 percent of all students had missing data for the scale because they were missing data for all four items. In order to avoid the loss of 21 percent of the level-1 sample, missing data for these cases was replaced by the scale mean, which was 0 due to standardization. Again, subsequent analyses were run with the original scale (with listwise exclusion) and the mean-substituted scale to check sensitivity of this rather 'crude' replacement method. Results were only reported if they differed for the original scale with 21 percent missing.

For three reasons, addressing missing data in level-1 outcomes was much less problematic than in predictors. First, the amount of missing outcome data was rather small to begin with (see Table 11, p.73). Second, outcome measures were rather highly correlated, allowing retrieval strategies such as multiple imputation. Third, items on self-reported smoking (ever smoking, current smoking and smoking acquisition index) represented some kind of repeated measures as their response categories partially overlapped, allowing the use of available data on one item to substitute missing data in a different one for overlapping response categories (see Table 6, p.53). Table 15 shows the rate of missing data in level-1 outcome measures after all retrieval strategies were performed.

Table 15. Amount of missing data in level-1 outcomes before and after handling of missing data

Item	**Missing data in initial dataset**	**Missing data after exclusion of "refusal cases"**	**Missing data after all retrieval/substitution measures**
Ever use of cigarettes	8.9%	4.2%	0.3%
30-day use of cigarettes	9.4%	4.7%	2.0%
Smoking Acquisition Index	11.6%	6.9%	2.5%
# of friends who smoke	3.4%	2.7%	2.7%
Normativity scale	0.8%	0.5%	0.5%

Overall, the missing data problem was substantially reduced with methods that minimized the threat of biased statistical conclusions. For rather crude mean replacements, sensitivity analyses were performed to assess the impact of this problematic method. To that end, level-1 regression models presented in Table 13 (p.77) and Figure 6 (p.76) were re-run with reconstructed variables. Only minor deviations from initial findings occurred, suggesting no detrimental influence of this replacement method. In subsequent model building, reconstructed and mean-imputed variables were used.

4.2. Ecological level (school level)

This chapter will present findings on the school-level only. This "sociological" perspective adds another level of evidence to assessing the effects of BTEP prevention programming. Again, this section was conceptually based on Wittmann's (1990) five-data-box framework in that all paths specified in the "southwestern" route were statistically tested. That is, the relationship between school-level predictor variables and school status was tested to demonstrate the quasi-experimental nature of the data ($r_{PR,NTR}$ path, see Figure 1, p.30). If schools differed on relevant predictors depending on their status as control school or intervention school, that would have implications for modeling or statistical conclusions. Further, by switching to the school level, policy-relevant outcomes readily understood by all stakeholders, such as school prevalence rates of *ever smoking* and *current smoking*, could be examined. That prospect was appealing because findings expressed as difference in prevalence rates between control and intervention schools may be intuitive to many stakeholders in the public health domain. Additionally, more statistically powerful regression analyses with the continuous prevention index measure were performed to assess the impact of BTEP's prevention efforts. Investigating relationships between intervention measures and outcomes represented the $r_{NTR, CR}$ path in Wittmann's (1990) framework.

A total of 198 schools were available for analysis. Independent level-2 variables came from administrative sources (e.g. BTEP prevention index, academic achievement, poverty) and aggregated level-1 variables. In order to account for the potentially confounding influence of demographic differences of students contributing data to AYS across schools, a number of level-1 aggregates were included in the analysis. Those were the percentage of male students for a given school contributing data to the AYS sample, percentage of American Indian students, and average age of students in AYS sample. All these predictors were checked for their interrelation to avoid multicollinearity and unstable regression results. Outcome measures were all aggregated as school means from level-1 outcome data. It was expected that school means of aggregated level-1 outcomes would be even higher correlated than level-1 outcomes themselves. To avid redundancy, possibilities of extracting a principal component factor were explored.

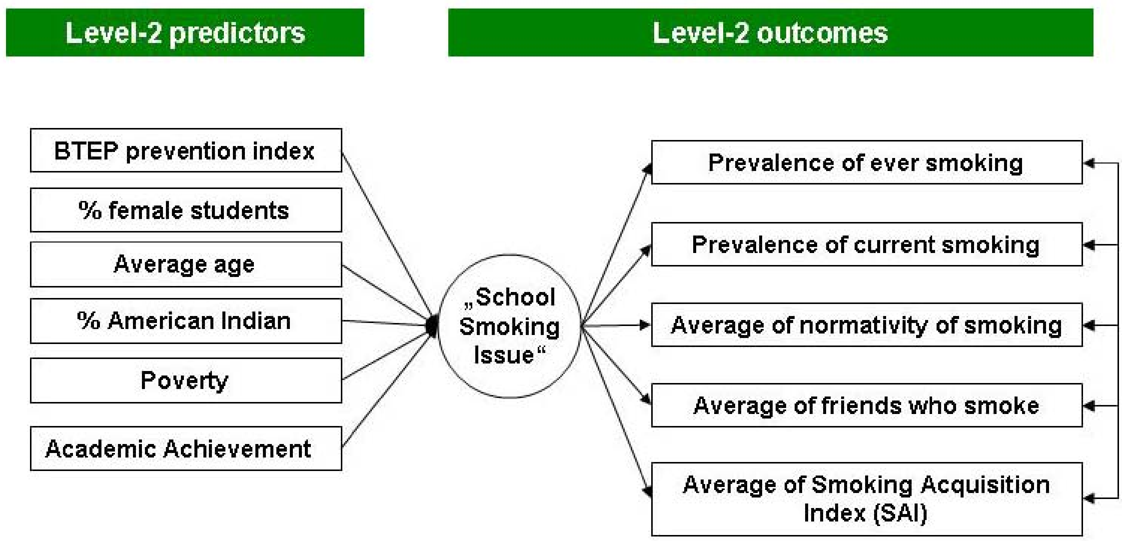

Figure 9. Structural level-2 only model

The figure above shows the structural model that was used to test the influence of BTEP prevention intensity on outcomes on the school level.

A number of preparatory steps had to be undertaken before the final model presented in Figure 9 could be investigated. These are briefly outlined in the following chapter. The final level-2 regression model and statistical results terminate this section, after which HLM models are presented.

4.2.1. Descriptive statistics and preparatory analyses of level-2 variables

Table 16 below shows descriptive statistics of level-2 variables. Columns labeled "AZ" show findings for data of the entire state sample (N=198 schools) and "MC" columns display findings for Maricopa County only (N=98 schools).

Table 16. Descriptive statistics of level-2 variables

Measure	**Mean**		**SD**		**Median**		**Minimum**		**Maximum**		**% missing data**	
Independent variables												
	AZ	MC	AZ	MC	AZ	MC	AZ	MC	AZ	MC	AZ	MC
BTEP Prevention Index (intervention schools only)	164	130	117	90	129	106	6	6	504	437	0	0
School Poverty (%)	58	52	27	30	63	50	1	1	99	99	9	7
Academic Achievement	4	4.3	1.2	1.4	4	4	1	2	6	6	0	0
Percent Male (%)	52	52	9	9	52	52	14	14	89	89	0	0
Mean Age	13.6	13.6	0.2	0.2	13.6	13.6	12.8	12.8	14.5	14.1	0	0
Percent Am. Indian (%)	8	3	21	4	2	2	0	0	100	31	0	0
Dependent variables												
Prevalence of ever smoking (%)	34	31	15	14	33	30	0	6	100	76	0	0
Prevalence of current smoking (%)	13	11	9	9	10	9	0	0	54	54	0	0
Mean # of best friends who smoke	1.8	1.8	0.4	0.5	1.8	1.73	1	1.2	4.4	4.4	0	0
Average normativity of smoking	0.0	0.0	0.2	0.2	0.0	0	-0.5	-0.4	1	1	0	0
Average Smoking Acquisition Index	1.4	1.4	0.4	0.4	1.3	1.3	1	1	3.7	3.7	0	0

Findings on the standard deviation in the table above indicate that there was great variability in BTEP prevention activity across schools in the dataset. Control schools with presumably zero BTEP prevention programming were excluded to obtain a more accurate picture of prevention statistics at the school level. Intervention schools administered about 160 minutes (or about 2.5h) of some curriculum per student over the course of two years. This translates into only about 1h per year per person. Given that prevention classes were typically not delivered to the entire school, this amount may be much higher for those students who actually received the programming (depending on curriculum up to 12h). Very few schools implemented a very high amount of tobacco prevention curricula as can be inferred from the high maximum values of the BTEP index. On the lower end of the range, some schools delivered only marginal amounts of "treatment" when extrapolated to the entire school. Little differences where found between Arizona overall and Maricopa County, again because Maricopa contributed the bulk of students to the final dataset and is the most populated county in the state by far (containing more than 60% of the entire state population). Figure 10 below shows the distribution of the BTEP Prevention Index by Arizona's 15 counties from those schools that provided outcome data per AYS.

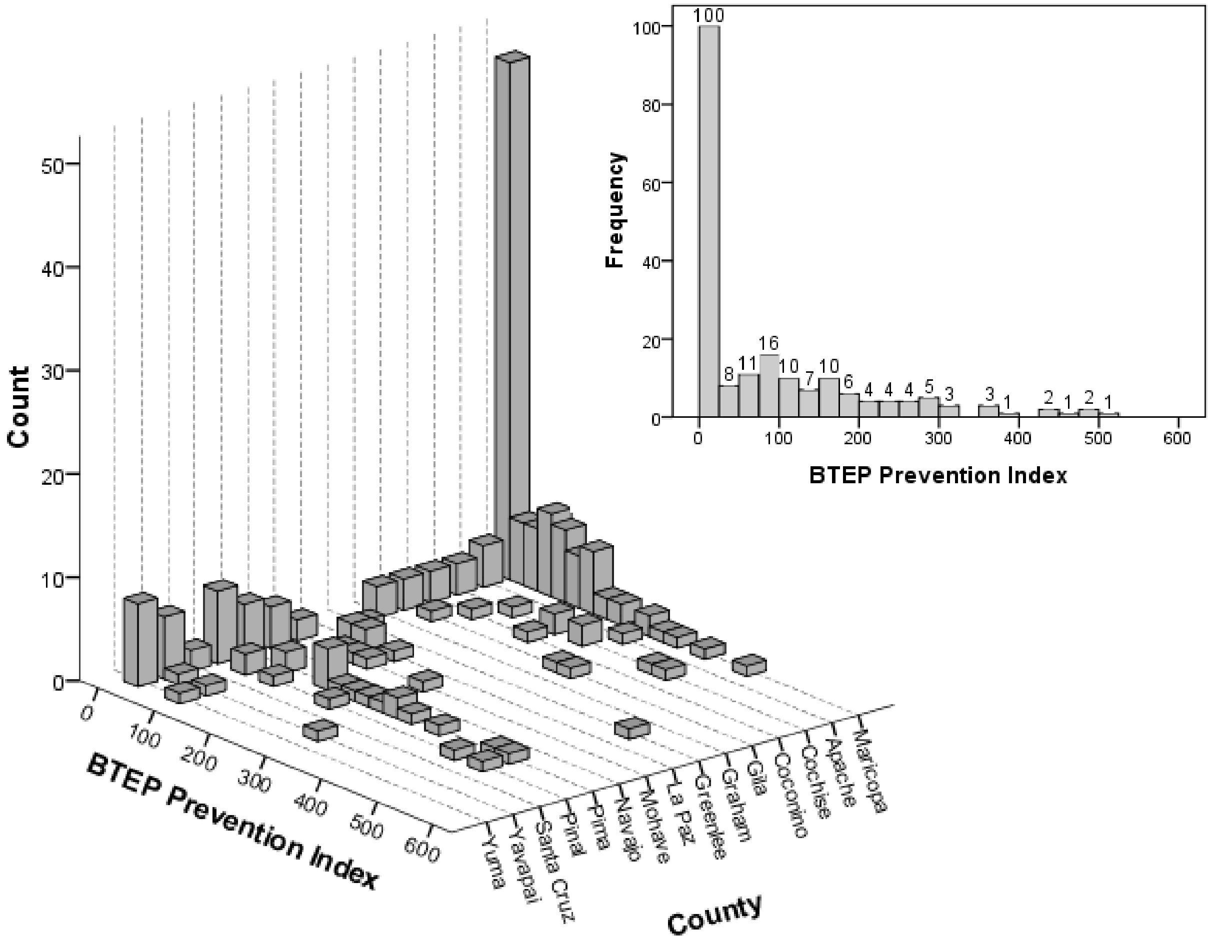

Figure 10. Histogram of BTEP Prevention Index by County

A number of features of the intervention data become apparent in Figure 10 and deserve mentioning. 1) For all but two counties (La Paz & Greenlee), 'control' and 'intervention' schools were available from outcome data. Control schools are those that have not implemented BTEP prevention curricula at all over the course of the two years considered here (school year 2004-05 and 2005-06). Overall, about 50 percent of schools (N=100) were 'control' schools. 2) A sufficient amount of 'control' and 'intervention' schools was only available for the state as a whole and Maricopa County so that questions around BTEP effectiveness could be addressed for Maricopa and assuming a statewide perspective. However, it must be kept in mind that the opportunistic use of available outcome data may have resulted in an inadequate sample or even non-sampling of intervention schools in counties other than Maricopa. This means that more sparsely populated counties not strongly represented in the outcome data could not be examined with respect to BTEP effectiveness and statewide findings cannot be applied to these counties.

On average, one of two students is eligible for free or cost-reduced lunches across schools, with high variability of the poverty indicator across the entire spectrum (ranging from 1% to 99%). Plotting revealed a left-skewed distribution with the majority of schools concentrated in the higher ranges of poverty levels. This indicator was also the only level-2 variable with any missing data due to unavailability of the indicator from administrative records.

In terms of academic achievement of the school as a whole, most schools displayed overall academic performance rated "performing plus" on the AZLearns system. Academic achievement displayed a smaller variability than the BTEP prevention index or poverty and plotting revealed an approximately normal distribution.

Most school samples provided a balanced gender ratio and only few school samples contributed a disproportionate gender distribution to the data. Average ages for eighth graders showed very little variability across schools, which was expected due to them all being eighth grade students. However, even small variability in age could make difference for tobacco-related outcomes. Older students may be in grade eight for a number of reasons relevant to tobacco use, such as academic failure etc. American Indian students constituted small shares of the school populations under consideration. There was large variability for all schools in the sample but very small variability of the share of American Indian students in Maricopa County schools. The average rate of American Indian students in Maricopa was also lower than in the full sample. Ten schools provided 85% or more American Indian students to AYS. Correlations among level-2 predictors were examined to minimize the likelihood of redundancy in subsequent regression models. A very high correlation was found between academic achievement of the school and the poverty indicator (r=-0.74). This means that schools with high rates of students on free or reduced lunch performed worse academically than did schools with lower rates of students from impoverished backgrounds. Inclusion of two such highly correlated predictors would destabilize any regression model. To examine which of the two indicators (or an index of both) had more predictive power, a number of preliminary regression models with block-wise inclusion of level-2 predictors were performed with 30-day school prevalence as level-2 outcome. Results showed that once academic achievement was controlled for, poverty rates could not explain additional variance in school prevalence. This was not the case when poverty was statistically controlled. Academic achievement of the school could still explain substantial portions of remaining outcome variance. Finally, a sum score of the standardized poverty indicator and standardized academic achievement performed worst in explaining variance of 30-day prevalence. Together, these findings demonstrated that academic achievement of the school is a more powerful predictor of the school smoking prevalence than is poverty rate or a combined measure. The poverty indicator was therefore excluded from further analysis.

Academic achievement of the school was mildly negatively correlated with average age of the student sample (r=-.35) and percentage of American Indian students (r=-.2). This means that schools with a higher average age of students and higher rates of American Indian students received lower ratings on the AZLearns score.

A very low but noteworthy negative correlation was found between rate of American Indian students and BTEP prevention index (r=-.16). An analysis that excluded the ten schools with rates of American Indian students higher than 80 percent revealed that this correlation then dropped to r=-.05. This finding may reflect BTEP's lower involvement in schools with very high rates of American Indian students. All schools with higher rates than 80 percent of students were located on reservation lands. During the period considered here (school year 2004-05 and 2005-06), BTEP has not funded prevention programming in reservation schools, which are not under the jurisdictive authority of the Arizona Department of Education. No other correlations were found between level-2 predictor variables. Outcome variables were aggregated as school sample means from level-1 outcome measures. On average, school prevalence of ever smoking was 34% and 13% for current (past 30 day) smoking. Histogram plots revealed that distribution for both prevalence rates were approximately normally distributed with some "outliers" of extremely high school prevalence of ever smoking and current smoking. The scatter plot below shows that both prevalence rates were closely correlated and that there were four cases with exceptionally high rates (in the upper right corner, indicate by the "x" markers).

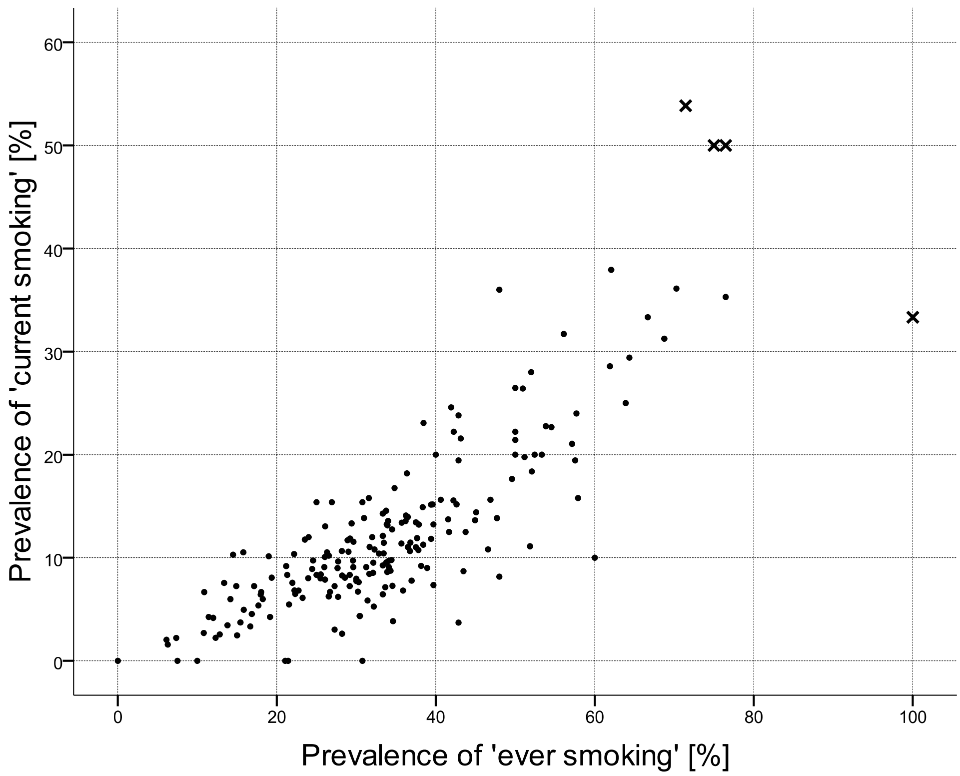

Figure 11. Scatter plot of prevalence of ever smoking and prevalence of current smoking

This may be an artifact of AYS sampling and had to be taken into account for further analyses (i.e. by running regression models with and without "outliers" to check for their potentially biasing influence). All level-2 outcome measures were very highly correlated (from 0.73 to 0.93). In order to avoid highly redundant analyses from analyzing each level-2 outcome separately, a PCA factor analysis with MAP test was performed to determine the possibility of extracting a single outcome factor. Results from the PCA eigenvalues and the mathematically superior MAP test lent strong support to a one-factor solution. Loadings of the five school outcomes ranged from 0.86 to 0.96. This general factor represented the severity of the smoking epidemic in a given school through combining aggregates of self-reported smoking behavior (ever, current, and average acquisition stage), average attitudes towards smoking and perceived smoking in the social environment. It was therefore labeled

"School Smoking Issue". Although this general factor encompassed *ever* and *current smoking*, separate analyses for these measures were performed because findings could be easily translated into policy-relevant implications that are understandable to policy makers and stakeholders who may not fully comprehend findings from a psychometric construction such as a general PCA factor.

4.2.2. Effects of BTEP prevention programming at the school level

The continuous measure of BTEP Prevention Index was dichotomized to run initial and preliminary ANOVAs to examine basic differences between intervention and control schools. This approach is not capable of controlling for potential confounders and may be less statistically powerful because of dichotomization of the predictor of focal interest (BTEP Prevention Index). However, it allowed for an initial and crude assessment of possible effects of BTEP's prevention efforts on crucial outcomes. Table 17 below shows results for different samples: the full sample (198 schools), after exclusion of schools with extreme prevalence rates (leaving 194 schools), Maricopa schools only (98 schools, and Maricopa schools without the four outlier schools (leaving 95 schools).

Table 17. ANOVA results of dichotomized BTEP Prevention Index, using different samples

	Full data (all schools)		**Without four extreme outliers**		**Maricopa only**		**Maricopa only without four outliers**	
	Control	Interv.	Control	Interv.	Control	Interv.	Control	Interv.
Outcome measure								
Prevalence of ever smoking (%)	37%	32%	35%	32%	33%	29%	30%	29%
Difference between control and intervention, sign. level (Effect Size: Cohen's *d*)	5%* (0.31)		3% (0.22)		4% (0.30)		2% (0.12)	
Prevalence of current smoking (%)	14%	11%	12%	11%	13%	9%	11%	9%
Difference between control and intervention, sign. level (Effect Size: Cohen's *d*)	3%* (0.29)		1% (0.15)		4%* (0.45)		2% (0.28)	
School Smoking Issue (PCA factor)	0.17	-0.15	0.01	-0.16	0.1	-0.38	-0.18	-0.38
Difference between control and intervention, sign. level (Effect Size: Cohen's *d*)	0.32* (0.32)		0.14 (0.2)		0.48* (0.45)		0.2 (0.27)	

Before exclusion of 'outlier' schools, effect sizes ranged from small to medium, according to Cohen's (1992) classification for effect sizes. Effect sizes were especially large for Maricopa schools and maintained an acceptable level even after control schools with extreme outcome values were excluded. This finding may reflect that the presumably more comprehensive implementation of prevention programming in Maricopa schools resulted in more readily detectable effects of programming on outcomes (see section 3.1.1, p. 38). Results in Table 17 indicate that initial effects of the binary BTEP indicator fade after schools with excessive smoking rates were excluded. Prevalence differences between intervention and control schools decrease for the entire state sample and Maricopa County only, for both indicators *ever smoking* and *current smoking*. The difference between the PCA factor 'School Smoking Issue' also diminished after excluding the most extreme schools. However, effects did not fully disappear or become reversed in direction, nor did effect sizes fade to trivial levels. Most reductions in effect sizes (and associated loss of statistical significance) resulted from a convergence of the statistics of the control to the statistics of the intervention schools. In other words, exclusion of the most extreme schools affected only control school statistics and made their results more similar to the intervention schools' statistics.

If the difference of two percent in current smoking would hold for Maricopa County in general, an extrapolation of this effect to the entire eighth grade population of Maricopa County (~49,000) would result in roughly 1,000 smokers prevented from current smoking through all direct and indirect effects of BTEP's prevention activity, in that grade level alone. Assuming an effect of four percent difference between control an intervention schools, this would translate into roughly 2,000 prevented current smokers in grade eight.

After these exploratory ANOVA analyses, regression models with the continuous BTEP Prevention Index were performed which also allowed to control for more potential confounders than just the 'outlier' schools. Table 18 shows results for all predictors on the "School Smoking Issue". Findings were broken down for a number of steps aimed at controlling for potentially confounding influences. Assessment of the impact of confounding factors was achieved by two means: a) exclusion of 'extreme' cases and b) a modified regression model in which all predictors except the BTEP Prevention Index were entered first, the so-called "block-wise" model building. After its initial exclusion as a predictor, the prevention measure was included to assess its incremental predictive power on outcomes. Separate analyses for Maricopa County were also performed because of the exceptional implementation that the county measure proxied for, as well as due to its unique position to provide equally high amounts of intervention and prevention schools. The final step consisted of testing the predictive power of all predictors among schools who had received some amount of prevention activity. Analyzing effects of intervention schools only served to examine possible dose-response relationships more closely by excluding all schools who have not received services at all.

Table 18. Beta weights of predictors in 'layered' regression models with outcome "School Smoking Issue"

Selection criteria → **Independent variables ↓**	**Full dataset (N=198)**	**Exclusion of four outliers (N=194)**	**Maricopa only (N=98)**	**Maricopa without outliers (N=94)**	**Interven-tion schools only (N=101)**	**Interven-tion schools within Maricopa (N=49)**
Percentage female	-0.19**	-0.06	-0.31***	-0.15*	-0.09	-0.07
Mean age	0.29***	0.30***	0.35***	0.33**	0.24**	0.17
Percentage American Indian	0.19*	0.21**	0.16*	0.06	0.13	0.09
Academic achievement	-0.28*	-0.28	-0.31***	-0.38***	-0.28***	-0.34**
BTEP Prevention Index	-0.05	-0.02	-0.14*	-0.13	-0.06	-0.05
R-Square change after inclusion of BTEP Prevention Index	0.0	0.0	0.02*	0.02	0.00	0.00
R-Square (with BTEP Index)	0.33***	0.31***	0.48***	0.33***	0.22***	0.18

Not reported in Table 18 are tolerance values of predictor variables. These were expectedly very high for all models, i.e. demonstrating the independence of predictors from one another. On the content level, this tolerance finding also implies that BTEP prevention activities were not disproportionately concentrated in schools with higher overall academic achievement.

The ‘layered’ approach of sequentially running a series of regression models (Table 18) allowed for more comprehensive results than an isolated, individual model. All steps presented in Table 18 were performed to assess BTEP effectiveness on the ‘School Smoking Issue” factor in the context of potentially confounding school-level predictors. The underlying rationale behind running a sequence of models was to estimate the sensitivity of findings under varying circumstances. In general, excluding extreme cases reveals the stability of findings in ordinary least-squares methods.

The most stable predictor with a large effect size of about d=0.7 throughout all models was average academic achievement as expressed in the AZLearns score. The higher the AZLearns score of the school was, the lower was the school smoking issue, which includes prevalence rates of eighth graders and the social climate towards smoking. The effect of this predictor remained roughly unchanged even in the model for Maricopa County Intervention schools only, in which all other predictors and the overall explained variance dropped substantially below the level in all other models (in addition to losing significance level of 0.1).

Average age of the students contributing data was an equally important predictor of the school smoking issue with a similar effect size than academic achievement. The influence of this predictor dropped only for the final model on Maricopa intervention schools only, but retained a medium effect size (β=0.17 translates into a Cohen's d of $d \sim 0.3$). In other words, the higher the average age of a student body in a given grade relative to other schools, the higher the school smoking issue. This finding may reflect the fact that students who were older in grade eight than most of their peers in this grade smoke at higher rates and have more favorable views of smoking as part of a more general syndrome that also contributed to their still being in grade eight at a higher age.

The influence of the share of female students who contributed data changed substantially when the four outlier schools were excluded from the entire sample. A similar drop occurred when the "outlier" schools were excluded from the Maricopa-only sample. This was largely due to the exceptionally low rate of female students in these four outlier schools who contributed data to the final AYS dataset. All other predictors showed no or little change after the exclusion of these schools from the entire sample.

The rate of American Indian students who provided data was substantially positively related to the size of the school smoking issue. The higher the rate of American Indian students contributing data to AYS, the higher was the school smoking issue. When restricting the model to Maricopa only, the effect of the percentage of American Indian students dropped strongly after exclusion of the four outlier schools. This was due to the fact that Maricopa schools had much less American Indian students contributing data to AYS 2006 (see Table

16, p.89), and that one of the outlier schools in Maricopa County had 100% American Indian students.

A small effect favoring intervention schools emerged that dropped very slightly after exclusion of outlier schools, but remained its incremental predictive power (in terms of effect size, but not in statistical significance), as expressed in the gain of explained variance. According to Cohen's rule-of-thumb, the effect size of the BTEP Prevention Index would fall into the lower range of medium effect sizes. The beta weight of 0.14 would translate into a Cohen's *d* of about 0.3. The higher the amount of hours spent per student on any of the prevention curricula, the lower the school smoking issue. The effect of the BTEP Prevention Index dropped to small effect sizes when only looking at prevention schools, statewide and for Maricopa only (see last two columns in Table 18, p.98). The beta weight of 0.05 translates into a Cohen's *d* of about 0.1, which is very small. All findings related to the BTEP Prevention Index may suffer from limited ability of the index to take into account that different curricula may have different effect sizes. The index did not measure fidelity of the treatment implementations.

An additional series of models was run that included all predictors from the previous models but included the percentage of students who reported having a smoking parent. These models were run in order to assess if differences in the rate of reported smoking parents between schools would mediate the effects of BTEP prevention activities. The underlying notion was that in schools with higher rates of students coming from smoker households, BTEP prevention contractors would engage in an "uphill battle" because these students are exposed to smoking role models at home, which may put them at increased risks of finding smoking acceptable and being more prone to smoking initiation. For brevity reasons, only findings for Maricopa County without outlier schools are discussed for this last set of model including the rate of parental smoking per school. The effect of the BTEP Prevention Index remained unchanged. The most important change occurred for the relative importance of the school's academic achievement in predicting the school smoking issue. The beta weight dropped from -0.38 (see Table 18, p.98) to -0.27, without losing its statistical significance. At the same time, the beta weight of the rate of reported smoking parents attained a magnitude of 0.41, boosting the overall explained variance of the school smoking issue factor to R^2=0.46. Tolerance values for academic achievement of the school (0.78) and rates of smoking parents (0.78) indicated a medium-sized colinearity issue. In fact, these two measures correlated r=0.38. Together with findings presented in Table 18 (p. 98), these results demonstrate that low ratings on the AZLearns score was an ecologically highly valid proxy measure for the rate of parental smoking reported by students. Low overall academic achievement not only signaled an aggravated smoking issue among the schools' eighth graders, but also indicated that students in these schools came from families with increased parental smoking. Administering curriculum-based prevention programming in such schools may consequently facing increased difficulties, as the receiving student body tends to come from smoker households and may therefore represent a 'hard' target.

Finally, because the factor 'school smoking issue' was derived from a factor analytical approach, its metric did not lend itself readily to a straightforward interpretation of unit changes in this dependent variable per unit change in the independent variable. Therefore, a final series of models were run with 30-day school smoking prevalence as outcome. Only effects of the BTEP Prevention Index are discussed here as it could be expected that the relative influence of predictors was very similar to findings for the PCA factor that included 30-day prevalence rates.

The unstandardized regression weight of the BTEP prevention Index for Maricopa County schools (without the four 'outlier' schools) and 30-day prevalence as outcome was b=0.000098 (p<= 0.12). The metric of the outcome variable was percentage points expressed as decimals, i.e. one percent was 0.01 and 100 percent were one. An increase of one unit in the BTEP Prevention Index, which may be interpreted as one curriculum minute per student over the course of two years resulted in a prevalence reduction of 0.01 percent. Put differently, administering each student 100 minutes of prevention curriculum over the course of two years (or 50 minutes per year, per student), in conjunction with all other implementation effects resulting from MACTUPPs presence in a school, would reduce 30-day point prevalence among eighth graders by one percent. It has to be kept in mind that construction of the BTEP Prevention Index averaged across all ages that had been served in a school. Additionally, these calculations assume that teaching all students 50 minutes of prevention curricula in a middle school of 100 students is equally effective as teaching 50 students 100 minutes. Certainly, such assumptions are unwarranted. Rather, the above calculation were performed in the attempt to quantify effects of BTEP intervention in a meaningful metric rather than sticking to the abstract and hard-to-interpret scaling of the PCA factor 'school smoking issue'.

For all models discussed in this chapter, graphical regression diagnostics did not reveal major violations in model assumptions: residuals appeared roughly normally distributed and Q-Q plots did not show major deviations from the assumption of linearity.

4.3. Hierarchical Linear Modeling of BTEP prevention effectiveness

Arriving at final models and their results is usually a process driven by a-priori considerations and exploratory data analytical steps. Building hierarchical linear models represents an extreme case of this rule because models have to be fit at various levels and findings may indicate the necessity to change specifications at either level. Therefore, the following sections describe the processes and steps that were taken to arrive at the final models. In this study, the complexity was increased because models had to be fit for five outcomes, at two levels and for two different sets of data; all study schools, and only those in Maricopa County. As Snijders and Bosker (1999) have pointed out, the complexity of model specification already known from OLS regression become substantially compounded in HLM because many more parameters have to be specified that are based on different assumptions, have different interpretations and implications. No ‘cook book’-style procedures exist for HLM. Instead, model specification is based on a mix of a priori, theoretical considerations, empirical findings with resulting adjustments and ‘common sense’.

4.3.1. Random effects ANOVA (intraclass correlations of level-1 predictors)

The first step in conducting multi-level modelling is fitting fully unconditional models (random effects ANVOA) to all the level-1 outcome measures to find out if the nesting of subjects in level-2 units reaches a sufficient magnitude that justifies using HLM (Bryk et al., 1992; Snijders et al., 1999). At the same time, such analyses yield estimates of within-group or level-1 variability and help in deciding how to fit complete models. A higher proportion of variability at the student level calls for more predictor variables at the student level as compared to the school level. Table 19 displays results from fully unconditional models for all level-1 outcome variables, i.e. after fitting random effects ANOVAs. Formulas of models involving random components are shown with in Appendix C to save space in this text. As mentioned earlier, because of differences in implementation and potential consequences for evaluation findings, models for all schools and only those in Maricopa County were run separately and thus are displayed and discussed separately. Intraclass correlations reported in Table 19 are so-called unconditional ICCs because they reflect variability due to clustering before any level-1 variables have been controlled for.

Table 19. Average school means with average reliability, variance components and ICC of level-1 outcome variables (statewide "AZ" and Maricopa County only "MC" in red)

Outcome measure	**Variable name**		**Average school mean γ_{00} (average reliability)**	**Standard error of γ_{00}**	**Level-1 variance component σ^2**	**Intercept variance component τ_{00}**	**(ICC) ($\tau_{00}/\sigma^2+\tau_{00}$)**
Ever use of cigarettes	ciglifeR	AZ	0.332 (0.807)	0.009	0.204	0.014	0.064***
		MC	0.301	0.012	0.195	0.012	0.058***
30-day use of cigarettes	cig30dyR	AZ	0.112 (0.634)	0.005	0.089	0.003	0.033***
		MC	0.095	0.005	0.080	0.002	0.024***
Social Norm Scale	smknorm	AZ	0.027 (0.738)	0.013	0.600	0.026	0.041***
		MC	0.027	0.018	0.575	0.025	0.042***
# of friends who smoke	q24brec	AZ	0.807 (0.792)	0.025	1.502	0.096	0.060***
		MC	0.753	0.034	1.461	0.095	0.061***
Smoking Acquisition Index	SAIR	AZ	1.382 (0.770)	0.018	0.882	0.048	0.052***
		MC	1.340	0.027	0.806	0.060	0.070***

In general, Maricopa County and the entire sample did not differ greatly which could be expected. It is noteworthy that for the Smoking Acquisition Index (SAI), Maricopa County's ICC was notably higher than that for all schools.

Overall, the intraclass coefficients (ICC) revealed that a small, but non-ignorable amount of variability occurred at the school level with all ICCs attaining statistical significance. This suggests that an HLM approach was warranted as all ICCs exceeded a conventional cut-off ICC < 0.01 (which is still arbitrary but useful). Looking at all schools ("AZ" rows), the clustering effect was strongest for *ever use* of cigarettes and for the number of friends who smoke. As in most HLM studies, most variability was at the student level. As for content, the findings reflect largely what has been reported for level-2 only results. The mean prevalence of *ever smoking* was 33.2 percent for the 198 schools, ranging from 31.4 percent to 34.9 percent (95% CI). All average reliabilities (except the one for the school mean of 30-day smoking) of school means γ_{00} reached acceptable levels of nearly 0.8, which means that most samples means are good estimators of the true school means. More results on school-level averages for other outcomes were reported earlier (see chapter 4.2.1, p.89).

After fitting fully unconditional models to all outcome variables, the author entered level-1 predictors identified in chapter 4.1.4 (p.75) as fixed effects only. This means that all relationships between outcomes and level-1 predictors were considered invariant across schools, i.e. not dependent on any school-level characteristics. In addition to that, models were developed to account for variability in school intercepts, with two explanatory school-level characteristics identified earlier: school-level academic achievement (AZLearns) and BTEP Prevention Index.

4.3.2. Means-as-Outcomes models

The following student-level predictor variables were included into subsequent models: sex (dummy), age, American Indian ethnicity (dummy), Latino ethnicity (dummy), parental smoking (dummy), parental attitudes (continuous), and bonding to parents (continuous). No interaction terms were included as there were no theory-based considerations suggesting this. Including level-1 predictors was done to adjust for differences in student composition between schools, not to assess the relative importance of these predictors. Only two level-2 predictor variables were considered or modeling the intercept variability, in accordance with the research questions of this study. School-level academic achievement, as measured by the AZLearns rating, was included in models d) and e) in order to find out if this measure was an ecologically valid indicator for a student's risk of engaging in smoking or finding smoking normative. The crucial level-2 predictor of this study, the BTEP Prevention Index, was introduced in models c) and e) to examine if it represented a protective factor for students to not engage in smoking or for finding smoking less normative.

Five outcome variables specified earlier were analyzed separately to maximize the likelihood of detecting effects of BTEP prevention and maintain interpretability that would have been lost by examining an artificial factor score. These were *ever smoking* (dummy), *current smoking* (dummy), normativity of smoking (standardized scale), number of best friends who smoke (continuous), and Smoking Acquisition Index SAI (continuous). For all outcomes, variables were used after missing data was addressed, as explained in chapter 4.1.5, p.79. Fixed effects for level-1 predictors are only reported for the first model in each outcome because neither were they the focal interest of this study nor did they change substantially after introduction of level-2 predictors. These models constituted a 'hybrid' version of analytical paths suggested by Wittmann's (1990) five-data-box framework: predicting outcomes with level-1 predictors was an instantiation of the $r_{PR,CR}$ path, while the examination of school intercepts with level-2 predictors was a version of the $r_{NTR,CR}$ path (see Figure 1, p.30).

In all models, the intercepts of the level-1 regression models were allowed to vary randomly across schools. An important statistic to consider before modeling random slope variability is the reliability of intercept estimates. This depends on the sample size for each school. Reliability estimates were r_{tt} = 0.59 for the intercept of normativity of smoking, r_{tt} = 0.66 for the intercept of *ever smoking*, r_{tt} = 0.43 for the intercept of *current smoking*, r_{tt} = 0.70 for the intercept of number of best friends who smoked, and r_{tt} = 0.59 for the intercepts of the Smoking Acquisition Index. Reliability estimates for these intercepts were moderate and in the case of *ever smoking* rather low. Nevertheless, random variability for all intercepts was analyzed with level-2 predictors under the assumption that results would represent the lower boundary of effects because unreliability works against statistical power. Resulting level-1 and level-2 variance components were important indicators to assess the overall model fit and level-2 variability that was explained by level-2 predictors. Likewise, the deviance statistic, estimated per maximum likelihood (ML) method, is an indicator that can be used to compare nested models to one another to inspect improvements (or worsening) of overall model fit. Variance components were computed by using restricted maximum likelihood (REML), as pointed out earlier. The research questions that motivated construction of the models presented below were:

(1) Does the introduction of level-1 predictors increase the overall model fit?

(2) Does school-level academic achievement, or BTEP prevention activity or both variables simultaneously explain the variability in intercepts? By how much does the intraclass correlation change after level-2 predictors are introduced (conditional intraclass coefficient in comparison to the unconditional intraclass coefficient)?

Question 2 addressed two theoretically different notions. Testing the influence of one level-2 predictor at a time answered if either one could explain variability in school intercepts, without 'controlling' for the influence of another level-2 predictor. Entering both level-2 predictors simultaneously answered the question if either one of the level-2 predictors remained a significant explanatory variable once the influence of the other predictor was factored in. Regarding content, this meant examining the question if BTEP prevention activity continued being an important predictor once academic achievement of the school was taken into account. Table 20 shows results for all outcome variables for the following five models: a) no predictors at all, the so-called random effects ANOVA, b) level-1 predictors as fixed effects only, with randomly varying intercepts, c) BTEP Prevention Index introduced as sole level-2 predictor, d) academic achievement (AZLEARNS) introduced as sole level-2 predictor, and e) both BTEP Prevention Index and AZLEARNS as level-2 predictors simultaneously. Models c) to e) try to explain the variability of the level-1 regression intercepts that were allowed to vary randomly. A lower conditional intraclass coefficient means that introducing the level-2 predictors reduces the level-2 variability, i.e. pairs of students in schools with the same value of the level-2 predictor are more similar than a pair of students from school with different values of the level-2 predictor.

Table 20. Means-as-outcomes results

(see next page)

Outcome variable	Model #	Level-1 fixed effects (odds ratios for outcomes 1 and 2, unstandardized regression weights for outcomes 3, 4 and 5)							Level-2 fixed effects			Random effects		Deviance (ML)
		Sex	Age	Latino	Am. Ind.	Parental smoking	Parental attitudes	Bonding to parents	Intercept	Acad. Achievem.	BTEP Index	σ^2	τ_{00}	-2 log likelih.
1. Ever smoking[1]	1.a	---	---	---	---	---	---	---	0.335***	---	---	0.204	0.014	31768
	1.b	0.98	1.43***	1.39***	2.16***	2.044***	0.452***	0.619***	0.317***	---	---	0.182	0.005	26492
	1.c	-|-	-|-	-|-	-|-	-|-	-|-	-|-	-|-	---	1.000	0.182	0.005	26492
	1.d	-|-	-|-	-|-	-|-	-|-	-|-	-|-	-|-	0.834***	---	0.182	0.004	26450
	1.e	-|-	-|-	-|-	-|-	-|-	-|-	-|-	-|-	0.833***	1.000	0.182	0.004	26450
2. Current smoking[1]	2.a	---	---	---	---	---	---	---	0.112***	---	---	0.089	0.003	10602
	2.b	1.10*	1.56***	1.24***	2.04***	1.858***	0.448***	0.621***	0.107***	---	---	0.084	0.001	8341
	2.c	-|-	-|-	-|-	-|-	-|-	-|-	-|-	-|-	---	-0.000	0.084	0.001	8341
	2.d	-|-	-|-	-|-	-|-	-|-	-|-	-|-	-|-	0.909**	---	0.084	0.001	8333
	2.e	-|-	-|-	-|-	-|-	-|-	-|-	-|-	-|-	0.909**	-0.000	0.084	0.001	8333
3. Normativity of smoking	3.a	---	---	---	---	---	---	---	0.027*	---	---	0.596	0.026	58870
	3.b	0.002	0.06***	0.07***	0.14***	0.177***	-0.413***	-0.188***	0.008	---	---	0.518	0.011	50643
	3.c	-|-	-|-	-|-	-|-	-|-	-|-	-|-	-|-	---	-0.000	0.517	0.011	-|-
	3.d	-|-	-|-	-|-	-|-	-|-	-|-	-|-	-|-	-0.025*	---	0.517	0.010	50632
	3.e	-|-	-|-	-|-	-|-	-|-	-|-	-|-	-|-	-0.026**	-0.000	0.517	0.010	50631
4. Number of best friends who smoke	4.a	---	---	---	---	---	---	---	0.807***	---	---	1.502	0.096	80097
	4.b	0.025	0.17***	0.19***	0.32***	0.207***	-0.427***	-0.250***	0.771***	---	---	1.373	0.054	71308
	4.c	-|-	-|-	-|-	-|-	-|-	-|-	-|-	-|-	---	-0.000	1.373	0.053	71306
	4.d	-|-	-|-	-|-	-|-	-|-	-|-	-|-	-|-	-0.039*	---	1.372	0.053	71302
	4.e	-|-	-|-	-|-	-|-	-|-	-|-	-|-	-|-	-0.041**	-0.000	1.373	0.052	71299
5. Smoking Acquisition Index	5.a	---	---	---	---	---	---	---	1.382***	---	---	0.882	0.048	67129
	5.b	0.015	0.13***	0.01	0.20***	0.188***	-0.460***	-0.137***	1.363***	---	---	0.812	0.017	60328
	5.c	-|-	-|-	-|-	-|-	-|-	-|-	-|-	-|-	---	-0.000	0.812	0.018	-|-
	5.d	-|-	-|-	-|-	-|-	-|-	-|-	-|-	-|-	-0.035***	---	0.812	0.016	60317
	5.e	-|-	-|-	-|-	-|-	-|-	-|-	-|-	-|-	-0.036***	-0.000	0.812	0.017	60316

'---' means these parameters were not part of the model.

'-|-' means that changes from the model above were miniscule and their change was of no interest to the pertaining research questions. [1] For binary outcomes, variance components, deviance and intercepts were obtained by running models as if these variables were continuous. Results for these parameters may be misleading because binary variables cannot be normally distributed. The purpose of running these statistics under flawed assumptions, however, was showing the change in these parameters, not interpreting directly.

Findings presented in the table above are from all schools in the final dataset. The following section summarizes the most important findings from these sequentially run models. The way to read results in Table 20 is comparing findings within each outcome by reading downwards in columns. The fixed effects of level-1 predictors were only reported for models b), because their values and associated interpretation were not of primary interest for the current study and changed only minimally upon entering of level-2 predictors. It suffices to broadly summarize results of level-1 fixed effects. Sex had no meaningful influence on any of the outcome variables but was kept in all models because additional analyses indicated that it substantially increased the overall model fit as expressed in the deviance statistic (not reported here). Latino ethnicity increased the risk for *ever* and *current smoking* slightly and was positively associated with having more best friends who smoke and finding smoking normative behavior. The same was found for American Indian ethnicity, but for this predictor effect sizes were sometimes twice as large as they were for Latino ethnicity. Being American Indian doubled a student's chance of being a current smoker as compared to a white student, all other considered predictors being equal. Parental smoking constituted a strong risk factor for a student's risk of *ever* and *current smoking* and of holding favorable views of smoking, as well as for the severity of smoking (as expressed by the Smoking Acquisition Index). Higher scores on the scale that measured bonding to parents and higher perceived parental disapproval of smoking were protective factors against the risk of smoking, against having more friends who smoke, against holding favorable views of smoking and against being a more established smoker.

For all outcomes, it can be seen that the change in student-level error variance (shown in the σ^2 column) was substantial for all outcomes when comparing model a) to b). This validated the relevance of these level-1 predictors, and can be interpreted as a reduction of within-school variability on those outcomes once all these predictors have been introduced. Even more importantly, level-2 variability on all outcomes was strongly diminished after introduction of level-1 predictors. This confirms the fact that much of the level-2-engendered variability in outcomes was in fact caused by differences in student demographics between schools. For all three continuous outcomes, where the reduction in level-2 variability could reliably be estimated, it turned out that for normativity of smoking, 58 percent of between-school variability was due to differences in level-1 predictors. For the number of best friends who smoked, 44 percent of school-level variability was engendered by level-1 predictors. Finally, in smoking acquisition status, 65 percent of school-level variance was attributable to differences in student populations on level-1 predictors. Similarly, deviance statistics (which express goodness of fit in a 'smaller-is-better' fashion) shrank drastically after entering level-1 predictors, suggesting profoundly better overall model fit. All changes in deviance from models a) to b) were highly statistically significant (as indicated by chi-square tests, not shown here). Introducing level-2 predictors had no influence on within-school variability estimates. This could be expected from the logic of hierarchical multi-level modeling, as these predictors had no within-group variability.

Of greatest interest to the current study were the fixed effects of level-2 predictors and the change in variability of school-specific intercepts (τ_{00}) from models b) to models c), and from models b) to models d). As can be seen in Table 20, fixed effects were statistically significant for academic achievement of the school but not for BTEP Prevention Index, for all outcome variables. Schools with higher academic achievement than the entire sample of schools put their students at about a ten percent lower risk of *ever* and *current smoking* than did schools with average or below-average academic achievement (with all other predictors being equal). Likewise, higher-than-average academic achievement of the school was associated with less favorable views of smoking, fewer reported best friends who smoke, and less established smoking status. Small but relevant amounts of level-1 intercept variability were explained by attending a school with above-average academic achievement. On the content level, these finding suggest that academic achievement as a characteristic of the school environment had a fostering effect on desired outcomes, such as lowering the perceived normativity of smoking and lowering the risk of self-reported smoking, once potentially confounding influences on the student level had been controlled for.

Fixed effects of BTEP Prevention Index did not attain substantial effect sizes (nor statistical significance) for any of the five outcomes. Introducing BTEP Prevention Index did not substantially reduce intercept random variability for any of the outcomes, either. However, for reported number of best friends who smoked, the index approached statistical significance of $p<=0.10$ and a small amount of random intercept variance could be explained by this predictor. Likewise, this outcome (number of reported friends who smoke) was the only one were a small improvement in model fit was observed after BTEP Prevention Index was introduced as level-2 predictor (albeit not reaching statistical significance of $p<=0.10$). Findings summarized in Table 20 seem to suggest that BTEP Prevention Index had no discernible influence on student-level outcomes when all schools that were available for this study were considered. It has been pointed out earlier that the organizational structure of prevention service administration justified restricting all analyses to Maricopa County. To that end, all analyses presented in Table 20 were performed for Maricopa County schools only. The author restricted the presentation of findings to fixed effects of level-2 predictors, random effects and overall model fit (deviance) and explained level-2 variability, because fixed effects of level-1 predictors were neither of focal interest in this study nor did they change in any meaningful way when only Maricopa schools were considered. To ensure clarity, results for *ever* and *current smoking* were not put into Table 21 as analyses revealed nil effects of BTEP Prevention Index for Maricopa schools only, a parallel finding to what was reported for all schools in Table 20.

Table 21. Means-as-outcomes results, Maricopa only (without level-1 effects and ever /current smoking)

Outcome variable	**Model #**	**Level-2 fixed effects**		**Random effects**		**Deviance (ML)**	**Explained level-2 variability[2]**
		Acad. Achievem.	BTEP Index	σ^2	τ_{00}	-2 log likelihood	
3. Normativity of smoking	3.a	---	---	0.574	0.025	37186	---
	3.b	---	---	0.496	0.010	31819	60%
	3.c	---	-0.0001*	0.496	0.010	31818	0%
	3.d	-0.017*	---	0.496	0.010	31816	0%
	3.e	-0.018*	-0.0002*	0.496	0.010	31814	---
4. Number of best friends who smoke	4.a	---	---	1.461	0.095	50901	---
	4.b	---	---	1.323	0.053	45123	44%
	4.c	---	-0.0006*	1.323	0.052	45119	1.9%
	4.d	-0.021	---	1.323	0.054	45121	---
	4.e	-0.025	-0.0006*	1.322	0.053	45117	---
5. Smoking Acquisition Index	5.a	---	---	0.806	0.060	41712	---
	5.b	---	---	0.741	0.018	37371	70%
	5.c	---	-0.0003*	0.741	0.020	37369	---
	5.d	-0.030*	---	0.741	0.019	37366	---
	5.e	-0.033*	-0.0003*	0.741	0.021	37363	---

'---' means these parameters were not applicable to the model.

In accordance to what has been found for analyses on all schools, introducing level-1 predictors dramatically lowered residual level-1 variance (σ^2) and school mean variability (τ_{00}) after level-1 predictors were introduced (models b) when compared to the unconditional models a). The vast majority of school means variability in normativity of smoking (models 3), reported number of best friends smoking (models 4) and Smoking Acquisition Index (models 5) was attributable to differences in student demographics. Likewise, the overall model fit improved substantially after introduction of level-1 predictors. More important for the purpose of this study, however, were the results on the BTEP Prevention Index, which attained small but statistically significant values for all three outcomes presented in Table 21. The reported number of best friends ho smoked was the only outcome for which explained level-2 variance (school mean variance) attained a meaningful result. About two percent of the school-mean variability in that outcome was attributable to the intensity with which MACTUPP had targeted that school. While interpreting these results, it is important to keep in mind that the metric for the BTEP Prevention Index was ‘average person-minute of any curriculum over two years’. In other words, in order to reduce the Maricopa school average of perceived normativity of smoking by one unit, MACTUPP/BTEP would need to put in about 42 hours (1/0.0002*2*60) of prevention education per student per year. If a typical elementary school year contains about 1,100 hours, 42 hours would be about four percent of that time. It has to be kept in mind, however, that these data contained an unknown amount of students who may have received no prevention curriculum at all. Therefore, these estimates represent the lower boundary of what would have been found had all students been subjected to prevention services, because the weak effect size of BTEP prevention activity may have resulted from the fact that it had to be conceptualized as school-level intervention measure. Therefore, the hypothesis that was tested here linked individual-level outcomes to effects that must have been instantiated by changes that BTEP prevention activity had on the overall school environment.

Because the amount of prevention services received by individual students could not be discerned in this study, it was argued that student’s reports on the number of their best friends who smoked would serve as “radar” into the school environment and proxy for the general prevalence of smoking in a given school. It is therefore a positive finding for program effectiveness that more intensively targeted schools had a lower average of reported number of smoking best friends than those that were targeted less intensively. The effect size was very small, but with negative sign and statistically significant. It may be an indicator that MACTUPP's implementation of curriculum-based smoking prevention affected those schools' overall environment. This could have led to fewer students smoking or finding smoking normative behavior. Finally, higher-than-average BTEP Prevention Index was also associated with students being less committed to smoking, as expressed in the Smoking Acquisition Index.

Academic achievement was negatively associated with all outcomes presented in Table 21. This finding confirms that academic achievement as a characteristic of the school was an important protective factor against undesired public health outcomes, such as viewing favorably at smoking or having friends who smoke.

4.3.3. Random coefficient modeling

After variability in school means was analyzed for all outcomes in the previous section, the final step in hierarchical linear modeling was the examination of variability in slopes of level-1 predictors and subsequent modeling of such slope variability with the two level-2 predictors ('BTEP Prevention Index' and 'Academic Achievement, AZLEARNS)'. If intercepts and slopes are allowed to vary from school to school, the covariance between slopes of different level-1 predictors and covariance between slopes and the intercept constitute new sources of information. On the other hand, estimates on these covariance parameters penalize the modeler with lost degrees of freedom. In general, theoretically derived arguments should guide the decision as to which parameters in the models should be set as fixed or random. For the purpose of interpretability and parsimony, as few predictors as possible should be allowed to vary randomly. Allowing too many predictors to vary randomly may jeopardize the stability of resulting estimates, as the data may not be strong enough to support the “flurry” of parameters estimates.

Questions to be addressed by random coefficient models were the following:

(1) How did slopes of level-1 predictors vary between schools? Was that variability significantly different from zero?

(2) What was the reliability of slope estimates?

(3) What was the correlation between random intercepts and random slopes? That is, had schools with higher average outcomes also stronger or weaker associations between level-1 predictors and outcomes?

One example for question three would be a negative correlation between school smoking prevalence and the regression weight of parental disapproval of smoking. This would indicate that in schools with an above-average smoking rate, the protective influence of parents’ disapproval of smoking was weaker. Of the seven level-1 predictors, only three were considered for randomly varying slopes. These were parental smoking, parental disapproval of smoking (attitudes) and bonding to parents. Excluded from allowing their slopes to vary randomly were age, sex, and the two ethnicity dummies (Latino, American Indian). Preliminary analyses showed that the random variability differed significantly from zero for all level-1 predictors, including the ones not considered for random slopes. The decision to exclude the aforementioned variables, however, was based on the consideration that the ultimate goal of this research was examining the effects of BTEP prevention efforts. Modeling the random slope variability of the ethnicity dummies (or age, or sex) would mean that the author hypothesized that the relationship of these variables with the outcomes was moderated by school-level predictors (BTEP Index and academic achievement). This seemed inappropriate because the curricula were not designed to ameliorate the potentially detrimental influence of age, sex or ethnicity on the risk of smoking. Additionally, the theoretical status of the variables excluded from random slope modeling (age, sex, and ethnicity) was that of 'proxy variables'. In other words, the true causal agents behind these manifest variables for smoking risk were unknown and could only be speculated about. Interpretation of explained variability of the slopes for these variables (age, sex, and ethnicity) without knowing the actual causal agents would have been close to impossible. The author preferred avoiding possible interpretative statements in the case of positive findings for the BTEP Index, such as: "Higher average prevention activity in a school protected against the risk factors of being older (or male, or American Indian)".

However, because the curricula in use were designed to reduce youth's susceptibility to negative influences of the social environment, these variables were allowed to vary randomly across schools so that their random variability could later be examined with level-2 predictors. The next step was examining if the selected level-1 predictors varied randomly for each outcome, so that subsequent modeling of these random slopes with level-2 predictors was statistically justified. This answered question one above. For reasons of brevity and the secondary importance for the questions of this study, fixed effects for level-1 predictors were not reported anymore.

Table 22 below shows that random slope variability differed for all outcomes. Asterisks indicate that most slope variances differed significantly from zero. These significance tests, however, are based on maximum-likelihood engendered chi-square tests. These are known to be overpowered with large sample size. For a lack of a readily available effect size measure for random slope variance components, the author chose the improvement in model fit of the overall models as decision criterion on what random slopes should be kept. The difference in deviance statistics is a chi-square distributed estimator. It tells how much a more complicated model surpasses a less complicated model in fit, given that the less complicated model is fully nested in the more complicated model (i.e. that all parameters that were estimated in the more complex model were also included in the less complex model).

Table 22. Random slope variance of level-1 predictors

Outcomes	**Random slope variance of level-1 predictors**			**Deviance with random slopes**	**Deviance without random slopes**	**Difference in deviances (Chi-square, df)**
	Parental smoking	Parental attitudes	Bonding to parents			
1. Ever smoking	0.047	0.041*	0.018*	26468	26492	$\chi^2_{(9)}=23$**
2. Current smoking	0.066*	0.009	0.025	8341	8146	$\chi^2_{(9)}=163$***
3. Normativity of smoking	0.005***	0.016***	0.003***	50569	50642	$\chi^2_{(9)}=72$***
4. Number of best friends who smoke	0.011*	0.046***	0.010***	71163	71307	$\chi^2_{(9)}=116$***
5. Smoking Acquisition Index	0.025***	0.087***	0.008**	59925	60328	$\chi^2_{(9)}=402$***

As can be seen in Table 22, more complex models were a much better fit than less complex models. All differences in deviance statistics (last column in Table 22) were highly significant chi-square values with nine degrees of freedom. This means that for all outcomes, models with randomly varying slopes for parental smoking, parental disapproval of smoking (attitudes), and bonding to parents fit much better than those without randomly varying slopes. This was true when all models with and without random slopes were run for Maricopa schools only (findings not shown here).

With various level-1 predictors being allowed to have random slopes, it became a necessity for further modeling to inspect their reliabilities as these reliabilities set the lower boundaries of model accuracy. That is, modeling random coefficients with low reliability de facto means that the signal-noise ratio is low and failure of detecting meaningful effects may be more a function of low reliability than truly small, irrelevant effect sizes. As Bryk and Raudenbush (1992) have pointed out, the reliability of slope estimates depends on the sample size for each group (school) and the underlying variability in predictor variables within schools. That is, low reliability of slopes may be a function of not only small sample size, but also a result of low variability (high homogeneity) in predictor variables within schools. The OLS regression between level-1 predictor and outcome would then be highly instable (or unreliable) because of the low variability of the predictor variables. As can be seen in Table 23 below, reliabilities for random slopes for the three level-1 predictors were rather low. Random slopes for parental smoking had the best reliability across all outcomes. Slope reliability for the outcome *current smoking* (smoking during the past 30 days) was low for all. Overall, however, even with low reliabilities, exploratory analyses were conducted by regressing level-2 predictors on random slope estimates (see next section).

Table 23. Correlations between random intercepts and slopes, and reliabilities of random slopes

		1. Ever smoking	2. Current smoking	3. Normativity of smoking	4. Number of best friends who smoke	5. Smoking Acquisition Index
Correlations between random components	Intercept-Parent smoking	-0.419	-0.262	---	0.539	0.793
	Intercept-Parental disapproval	-0.421	0.593	-0.503	-0.603	-0.724
	Intercept-Bonding to parents	0.542	---	-0.543	-0.66	-0.888
	Parental smoking-Parental disapproval	-0.585	-0.38	-0.218	-0.456	-0.454
	Parental smoking-Bonding to parents	-0.851	---	---	-0.786	-0.600
	Parental disapproval - Family bonding	0.239	0.719	---	0.202	0.402
Reliability of random slopes	Parental smoking	0.638	0.394	0.564	0.661	0.546
	Parental disapproval	0.158	0.116	0.159	0.14	0.339
	Bonding to parents	0.142	0.036	0.368	0.374	0.592

'---' Correlations were smaller than 0.2.

Another important set of results were the covariances (correlations) between the randomly varying coefficients. A brief summary sufficed for the secondary purpose of these findings to this study. As for the correlation between school intercepts and the within-schools correlations of parental smoking and respective outcomes, the following observations were made. High rates of ever smoking in a school were associated with a lower influence of parental smoking on ever smoking. The same was true for current smoking but to a lesser degree. This means that in schools with high average rates of smoking, the detrimental effect of parental smoking on self-reported smoking was weaker. The opposite was true for the average of number of friends who smoked and the influence of parental smoking on reported number of friends who smoke. The higher the reported average number of friends who smoked in a school was, the more influential was parental smoking on having more friends who smoked. Higher average smoking acquisition status in a school was highly positively correlated with the detrimental influence of parental smoking on the acquisition status of smoking. Together, these finding suggest that individual-level risk factors have an aggravated influence in environments where these risk factors have a higher overall group average. A careful interpretation would be to assume that school environments with a higher accumulation of risk factors exacerbate the influence of risk factors from the personal background. The opposite pattern was found for the relationship between outcome intercepts and the influence of the protective factor of bonding to parents on outcomes. It appeared that in higher risk environments (higher outcome intercepts), i.e. those with higher school averages in perceived normativity of smoking or higher reported number of friends who smoked, the protective effect of bonding to parents was more pronounced (indicated by the negative correlations between intercepts and slopes of parental bonding for these outcomes). A careful interpretation of these findings suggested that in higher-risk environments, the protective effects of higher bonding to parents exerted a higher influence than in lower-risk environments.

As for correlations between slopes, consistent and meaningful patterns emerged. Parental smoking was a dummy, coded zero for non-smoking parents and one if at least one parent smoked. It was found to be an important risk factor with positive slopes, meaning a smoking parent increased the risk of smoking for a student. Parental disapproval of smoking and bonding to parents, however constituted protective factors with a positive regression coefficient (see same tables). Therefore the very high negative correlations between parental smoking and parental disapproval, and between parental smoking and bonding to parents demonstrate that in schools were the influence of parental smoking is more pronounced, the effects of protective factors is also more pronounced. This seemed to suggest that in environments where parental smoking was more influential in increasing a students risk to smoke, protective factors of parental disapproval and bonding to parents were more influential in protecting students from risk of smoking or findings smoking more acceptable. Finally, the slope estimates for higher parental disapproval and higher bonding to parents were positively correlated, suggesting a synergistic effect of those two protective factors.

4.3.4. Slopes-as-outcomes modeling

Modeling slopes as outcomes represented the very final set of models to be built in this study. Their purpose was examining the potential role of BTEP prevention activity and academic achievement as moderators of smoking risk and attitudes around smoking. The term 'moderator' was chosen here purposefully as modeling slopes in hierarchical models is following the notion of interactions known from OLS regressions or ANOVA. More specifically, in HLM this is considered cross-level interactions because relationships between variables at the student-level, expressed as regression weights (i.e. slopes), now become themselves subject to variability that is examined by regressing slope estimates on school-level predictors. Modeling slopes with level-2 predictors is basically an extension of the models presented earlier where random intercepts were modeled (see chapter 4.3.2, p.105). The research question to be answered was the following: Did the strength of association between level-1 predictors and level-1 outcomes vary systematically within schools as a function of level-2 predictors (school-level academic achievement and, more importantly, the BTEP Prevention Index)? In other words, what amount of variability in level-1 regression weights, if any, could be explained by those focal level-2 predictors? The necessary condition for modeling this variability – sufficient variability in regression slopes between schools - was found to be fulfilled in the previous chapter (4.3.3).

Parallel to modeling intercepts with level-2 predictors, consecutive models for explaining slope variability started with entering solely the BTEP Prevention Index (models f), then solely the academic achievement measure (models g), and finally both level-2 predictors simultaneously (models h). Again, all models are shown with their formulas in Appendix C. Fixed effects of level-2 predictors represent regression weights for the level-2 model. Fixed effects of level-1 predictors are not shown here anymore, for the same reasons they were no longer reported in Table 21 (p.113) and Table 22, (p.118), which was their secondary relevance for the main research questions. Overall model fit is reported as difference score of deviance statistics (again, computed by using full maximum likelihood estimation rather than restricted maximum likelihood for comparability of nested models). The basis for difference scores of models f) and g) shown in Table 24 (p.124) and Table 25 (p.131) were models with identical specifications, but without level-2 predictors (not shown here). As such, these base models were fully nested within those shown in the tables below (a prerequisite for computing such difference scores). The basis for the difference score of models h) was the deviance statistic (and degrees of freedom) of model g). This answered the question if model fit improved after introducing the BTEP Prevention Index once academic achievement was already included (as it exerted a much stronger influence on overall model fit than the BTEP index). All slopes-as-outcome models were also run for Maricopa County only, because of its statistical suitability and unique implementation parameters.

Table 24. Fixed effects of level-2 predictors on slopes of three selected level-1 predictors

Predictors		**Level-2**	BTEP Index			Academic achievement			**Difference in deviance**[2]
		Level-1	Parental smoking	Parental attitudes	Bonding to parents	Parental smoking	Parental attitudes	Bonding to parents	
Outcomes	1. Ever smoking[1]	1.f	-0.0002	0.0002	0.0003	---	---	---	1.4 (3)
		1.g	---	---	---	0.057*	-0.081**	-0.056**	11.6 (3)**
		1.h	-0.0002	0.0002	0.0002	0.065*	-0.082**	-0.061**	0.3 (3)
	2. Current smoking[1]	2.f	0.0002	0	0.0004	---	---	---	0.6 (3)
		2.g	---	---	---	0.100**	-0.029	-0.073**	6.1 (3)
		2.h	0.0003	0	0.0003	0.102**	-0.029	0.070**	0.9 (3)
	3. Normativity of smoking	3.f	-0.0001	0.0002	0.000	---	---	---	4 (3)
		3.g	---	---	---	0.014*	-0.001	-0.021**	13.7 (3)**
		3.h	-0.0001	0.0002	-0.0001	0.013*	0.000	-0.022***	4.2 (3)
	4. Number of best friends who smoke	4.f	-0.0001	0.0003	-0.0002	---	---	---	4.4 (3)
		4.g	---	---	---	0.038*	-0.031*	-0.022*	12.8 (3)**
		4.h	0	0.0002	-0.0002*	0.038*	-0.029	-0.023*	4.7 (3)
	5. Smoking Acquisition Index	5.f	0	0	0.0001	---	---	---	2.1 (3)
		5.g	---	---	---	0.030*	-0.027	-0.015*	11.4 (3)**
		5.h	0.0001	0	0	0.031*	0.028	-0.014*	2.0 (3)

[1] For these outcomes, deviance differences had to be calculated by assuming these were continuous variables. Accordingly, results are more heuristic in nature and need to be interpreted with caution.

[2] ML computed deviance scores of these models were compared against deviance scores of the same models without the level-2 predictors. The resulting difference in a chi-squared distributed difference score.

Modeling BTEP prevention effectiveness as moderator of level-1 regression coefficients yielded very weak effects that failed to attain statistical significance at the $p <= 0.10$ level. Only one coefficient in model 4h went below this (arbitrary) alpha level. A cautious interpretation of this very small effect would be that more BTEP prevention activity increases the protective effect of higher bonding to parents guarding against reporting more friends who smoke. The very small size of effects of the BTEP Prevention Index has three potential reasons that may be at work individually or together in causing its smallness. 1) The metric of this indicator was average person-minutes of any smoking prevention curriculum over the course of two years. In other words, small effect sizes may be a reflection of the calibration of that measure, rather than an expression of zero effects. 2) The coefficients' magnitude depends on the reliability of utilized variables/measures that went into the analysis. These reliabilities must be assumed (and were shown) to be fairly low. Especially reliabilities of slopes estimates were low. However, this explanation is somewhat weakened by the fact that non-trivial effect sizes emerged for academic achievement which was affected by the same limitation. 3) The underlying hypothesis tested here was that BTEP prevention activity affected even students who may not have been exposed to prevention curricula directly but who were indirectly affected by possible changes in the school environment due to BTEP prevention activities. These indirect effects may be too weak to be measurable or may not have occurred at all.

Adding academic achievement as level-2 predictor did not substantially alter any effect size estimates for the BTEP Prevention Index. This means that controlling for the overall academic achievement of the school had no effect on the ways BTEP prevention activity may or may not have affected students' behaviors and attitudes. It suggests that effects of BTEP activity – if they occurred - were unaffected by the academic achievement of the school and the possible causal agents this measure proxied for. Finally, model fit did not improve in terms of effect size nor reach statistical significance after the BTEP Prevention Index was introduced. Overall model fit for models f) did not improve as compared to the baseline model, as indicated by small and non-significant chi-square difference scores (last column Table 24, p.124). All findings on BTEP prevention activity summarized above proved true for all five outcome variables. BTEP Prevention activity was largely unrelated to the slopes of parental smoking, parental disapproval and bonding to parents on all smoking-related outcomes (self-reported smoking, normativity and number of smoking friends).

A different picture emerged for academic achievement as moderator of level-1 slopes. Model fit improved substantially and attained statistical significance for all outcomes when academic achievement was introduced as level-2 predictor. Chi-square difference scores of models g) improved substantially as compared to the baseline model. The positive relationship between the level-1 risk factor of parental smoking and all five outcomes became more positive with increasing academic performance of the school as a whole, as indicated by the positive coefficients in the column of "Academic achievement-Parental smoking". This means that the level-1 risk factor of parental smoking was more influential in schools with higher academic achievement. It appears that higher academic achievement of the school did not constitute a protective factor against the personal risk of having smoking parents. Rather, in schools with lower academic achievement this level-1 risk factor exerted a smaller influence on a student's risk to smoke (or hold favorable views of smoking). The full picture, however, becomes apparent when findings on covariation of random components is taken into consideration (see Table 23, p.119). The correlation between parental smoking and intercept estimates was slightly negative, meaning that in schools with higher prevalence of 30-day smoking, the influence of parental smoking was less influential. Additionally, in Table 20 (p.108) it was reported that school prevalence of current smoking was mildly inversely related to academic achievement, suggesting that, on average, schools with higher academic achievement have lower smoking rates. Together, these findings suggest that a school environment characterized by higher academic achievement than the average school (because of the grand mean centering) may not buffer against the personal risk factor of parents who smoke because higher smoking prevalence estimates are concentrated in schools with lower overall academic achievement. For students in schools with relatively higher academic achievement, however, having a smoking parent increased the risk of current (and ever) smoking more than for students in lower-achieving schools. The figure below depicts these relationships graphically.

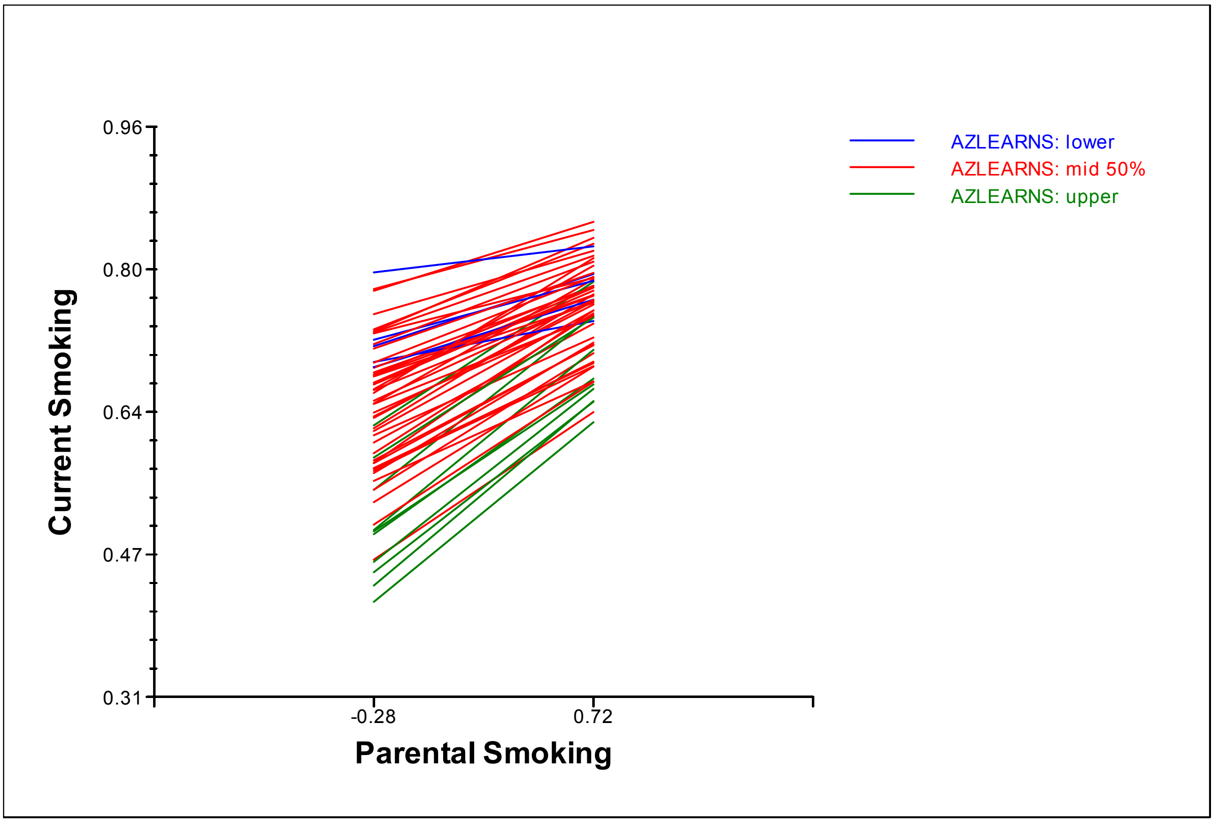

Figure 12. Influence of school-level academic achievement on the relationship between parental smoking and risk of current smoking

It can be seen that overall, slopes become less steep with increasing intercepts (expressing the negative correlation between slopes of the parental smoking-current smoking and intercept estimate for current smoking). Further, slopes are much steeper for schools with higher academic achievement (upper 75 percentile on academic achievement 'AZLEARNS', green color code) and intercepts are much lower than for schools from the lower 25 percentile of AZLEARNS. Slopes flatten out considerably from the lowest AZLEARNS percentile to the highest AZLEARNS percentile. Conversely, intercept estimates increase with decreasing AZLEARNS percentile, confirming higher smoking prevalences in schools with below-average academic achievement.

Consistent findings, but with opposite directionality, emerged for the level-1 protective factors of parental disapproval of smoking (parental attitudes) and bonding to parents. These protective factors exerted a stronger reductive influence on risk of *ever* and *current smoking* as well as a more protective influence of perceived normativity of smoking (and reporting fewer friends who smoked) in schools with above-average academic achievement. Technically speaking, the negative correlation between protective level-1 predictors became even more negative with higher relative academic achievement of the school. The negative relationship between these protective factors and all outcomes was, consequently, less pronounced in schools with below-average academic achievement. That is, if the school environment was characterized by low academic achievement, the protective effect of parental disapproval or higher bonding to parents turned out smaller than in schools with higher academic achievement. It appears that factors associated with an environment of below-average achievement weakened the protective effects of parental disapproval on smoking or higher bonding to parents. Identifying such factors, however, was beyond the scope of this study. Figure 13 below shows these relationships graphically.

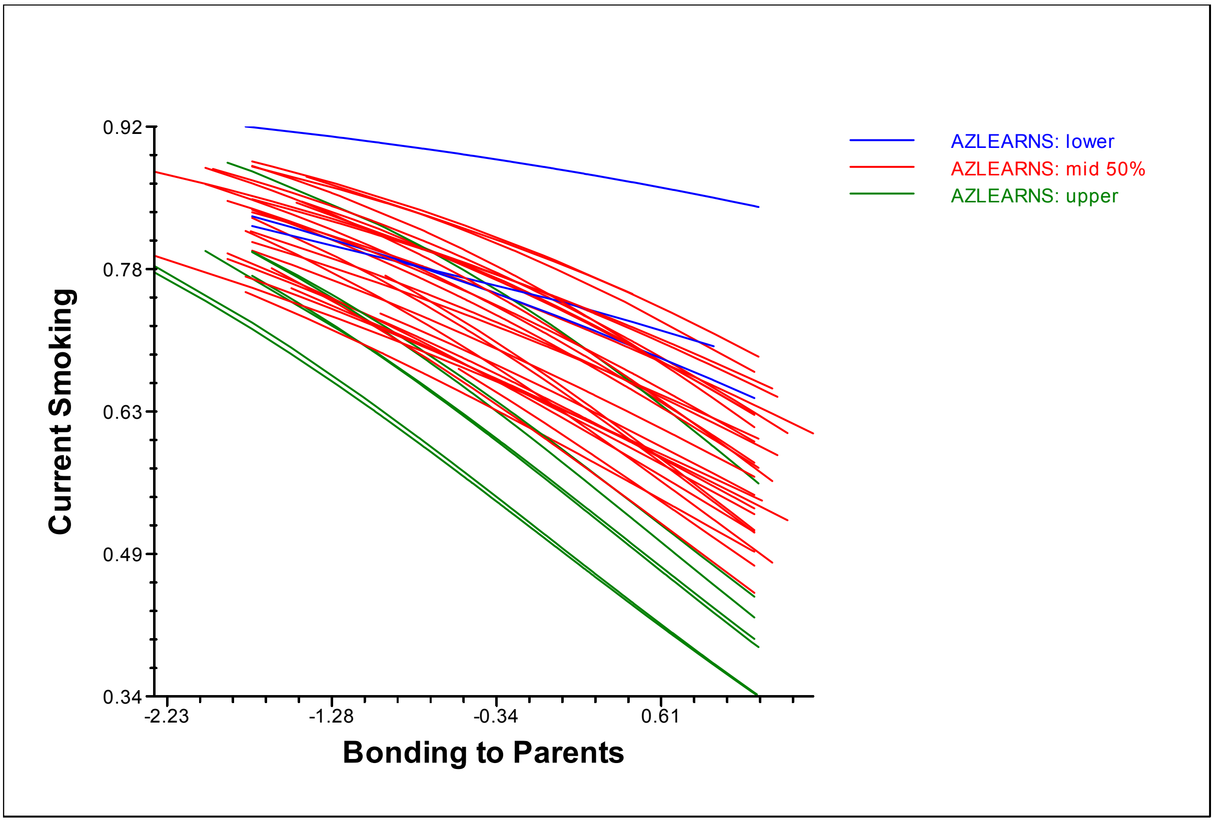

Figure 13. Influence of school-level academic achievement on the relationship between bonding to parents and risk of current smoking

Again, it can be seen that the slopes between bonding to parents (protective level-1 predictor) and current smoking were much steeper for schools with above-average academic achievement, only that signs of these slopes were now negative. This explains the negative signs of coefficients in the columns 'Academic achievement-Parental attitudes' and 'Academic achievement-Bonding to parents' in Table 24 (p.124) and

Table 25 below. In schools with below-average academic achievement, protective factors played less of a role than in schools with above-average academic achievement. Together with findings on level-1 risk factors above, this seems to suggest that school environments characterized by below-average achievement constitute risk factors in themselves, over and above individual-level risk factors and with reduced effectiveness of level-1 protective factors. It appears fair to say that if eighth-grade students in schools with above-average academic achievement smoked, they came most likely from a personal background with accumulated individual-level risk factors, i.e. smoking parents, parental non-disapproval (ignorance) with respect to their children's smoking, below-average attachment to parents etc. All models discussed above were also built for Maricopa County schools only.

Table 25 below summarizes findings of modeling slopes for Maricopa schools only.

Table 25. Fixed effects of level-2 predictors on slopes of three selected level-1 predictors, Maricopa County schools only

Predictors		Level-2	BTEP Index			Academic achievement			Deviance (ML)
		Level-1	Parental smoking	Parental attitudes	Bonding to parents	Parental smoking	Parental attitudes	Bonding to parents	-2 log likelihood
Outcomes	1. Ever smoking[1]	1.f	-0.0005	0.0002	0.0003	---	---	---	1.4 (3)
		1.g	---	---	---	0.066*	-0.053*	-0.072**	4.8 (3)
		1.h	0.0004	0	0.0003	0.063*	-0.052*	-0.070**	0.6 (3)
	2. Current smoking[1]	2.f	0.0004	0.0000	0.0005	---	---	---	1.4 (3)
		2.g	---	---	---	0.129**	-0.026	-0.072*	10.5 (3)*
		2.h	0.0006	0	0.0004	0.132**	-0.027	-0.070*	1.3 (3)
	3. Normativity of smoking	3.f	-0.0001	0.0002	-0.0001	---	---	---	3.9 (3)
		3.g	---	---	---	0.017	-0.017	-0.027***	15.9 (3)**
		3.h	0	0.0003	-0.0001	0.016	-0.019	-0.028***	5.6 (3)
	4. Number of best friends who smoke	4.f	0	0.0003	-0.0003*	---	---	---	3.5 (3)
		4.g	---	---	---	0.020	-0.014	-0.015	3.2 (3)
		4.h	0	0.0002	-0.0003*	0.020	-0.013	-0.017*	2.8 (3)
	5. Smoking Acquisition Index	5.f	0	0	0.0001	---	---	---	1.2 (3)
		5.g	---	---	---	0.032*	-0.012	-0.011	6.0 (3)*
		5.h	0.0001	0.0001	0.0001	0.032*	-0.018	-0.010	1.0 (3)

Findings suggest very similar effects of BTEP prevention activity and academic achievement on level-1 random slopes. The BTEP Prevention Index attained only small effect sizes, and only one effect fell below the arbitrary significance level of $p <= 0.1$. It appeared that schools with above-average BTEP prevention increased the protective effect of higher bonding to parents to reporting below-average number of friends who smoked. As such, a careful assessment of this effect suggests that the level-1 protective factor of higher bonding to parents was slightly augmented by implementation of prevention curricula. However, as was the case for all study schools, introducing BTEP prevention activity did not substantially improve overall model fit. This was true for models f) when compared to baseline models, as well as for models h) that compared the index' capability of improving model fit once the influence of academic achievement was taken into account. All remaining findings were similar to results for all study schools. That is, in schools with above-average academic achievement, the detrimental influence of parental smoking on students' risk to currently smoke was more pronounced. Conversely, such schools augmented the protective effects of bonding to parents on lowering the risk of current smoking and finding smoking normative behavior.

4.3.5. Examination of model assumptions (residual analysis)

Every thorough quantitative study subjects final models to a careful analysis of residuals, as this can yield insights that have important implications for the validity of conclusions that are derived from model building findings. Steps geared at checking the validity of model assumptions are interrelated and some serve multiple purposes. For example, inspection of level-1 residuals can help identify outlying cases and the assumption of normality of residuals. Deviations from normality can have many causes and each calls for a different remedy, such as considering the exclusion of outliers, transformation of predictors or outcomes or adjustments to model specifications (including/excluding variables). However, as Snijders and Bosker (1999) have pointed out, assumptions can never be proved to be met 100 percent, if only for the simple reason that in social science there is always the possibility of omission bias, i.e. the exclusion of one or more potentially relevant variables. Some model assumptions were already checked in the previous chapters during model building, such as inspection of model fit after introduction of new terms into multi-level equations. Most importantly, normality of residuals is desired as this indicates appropriateness of level-1 specifications and ascertains unbiased computation of standard errors of fixed effects at both levels and related significance tests. To that end, histograms of unstandardized residuals and Q-Q plots of all five outcome variables were inspected for fully unconditional models (i.e. no predictors at any level) vs. a final model with random slopes for the three level-1 predictors, random intercepts and level-2 predictors BTEP Index and AZLEARNS as level-2 predictors for all random components.

The figures below show the reduction of curviness of the residuals vs. the expected value.

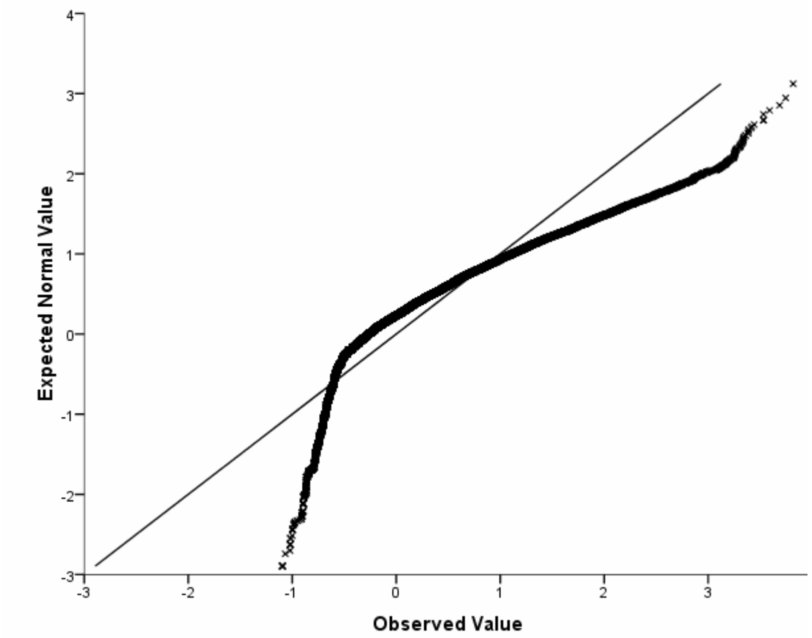

Figure 14. Q-Q plot of level-1 residuals of Normativity of Smoking scale of fully unconditional model

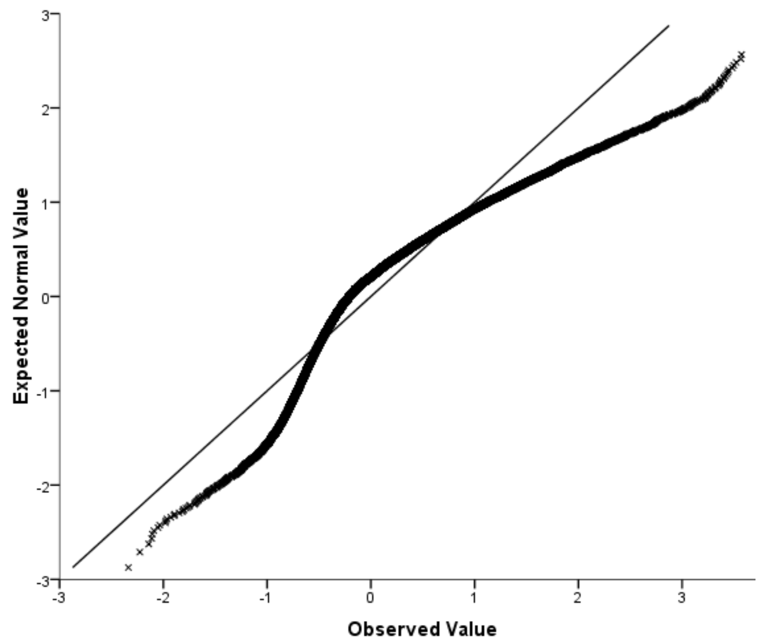

Figure 15. Q-Q plot of level-1 residuals of Normativity of Smoking scale of final models

Curvilinearity did not disappear completely but was reduced to an acceptable level. As such, model specifications and the normality assumption appeared to be met to a satisfying degree.

5. Discussion

This study tried to shed light on the effects of curriculum-based prevention programming on smoking related outcomes in eighth grade students. As a ‘by-product’, it yielded findings on the ecological validity of a school-level indicator (academic achievement) for indicating an aggravated smoking issue. A number of methodological-conceptual, statistical, and content-related issues deserve discussion. The very basis for this research was the desire and need for empirical evidence on a part of Arizona’s tobacco control program that costs over $4,000,000 annually (Arizona Department of Health Services, personal communication), as the new BTEP leadership is committed to modeling Arizona’s program according to ‘evidence-based’ practices. The study design followed limitations imposed by availability of data. Without the Arizona Youth Survey and extensive implementation data that was only available after many years of careful planning and finally collecting, no results could have been produced. As such, it is fortunate to have any empirical results at all. The archival approach used here, i.e. analyzing available data in ways that it was not initially collected for, has a number of important limitations, but also a few benefits. These are outlined in the remaining sections.

5.1. Review of methodology and statistical approach

In general, quasi-experimental designs cannot strongly suggest causality but only make a more or less plausible case for an ‘association’ between variables of interest, especially with cross-sectional data only. Such designs gain strength in causal inference if rivaling explanations for effects can be ruled out, e.g. by ‘controlling’ potentially confounding factors through statistical means. Ideally, longitudinal studies would be necessary to establish causal effects of school-level factors (West et al., 2004). Effects of prevention activity on outcomes were first modeled by examining schools as unit of analysis. This could only be done after aggregating student-level outcomes as school means. This 'ecological approach' where relationships are examined on a level higher than the individual has a number of shortcomings. The most important one is its inability to sufficiently account for differences in student-level confounder variables. In other words, differences in e.g. smoking prevalence among eighth grade students could be a function of differences in the demographic composition of the student samples in AYS data, rather than 'true' effects of factors of the school environment. One important measure to reduce such confounding in this study was adjusting for differences in student samples with respect to important potential 'confounders', i.e. variables associated with the outcome but a-priori unaffected by the intervention in the HLM framework. Such variables typically include demographic items like sex, age, and ethnicity that were included in level-1 structural models. In this respect, this study addressed this important methodological shortcoming brought up by West et al. (2004) and Aveyard et al. (2004).

Wittmann's (1990) five data boxes guided the author in linking intervention measures to outcomes determined by the EVA box (see Figure 1, p.30). Utilized outcomes of this study are widely considered crucial by most experts and stakeholders in the field of tobacco prevention. However, archival data restricts the analyst to building models with what is available, not with what should have been collected ideally. 'Omission bias' at level-1 (students) may have led to "under controlling" of level-1 residual heterogeneity, which then affects fixed parts of random slopes and intercepts. This problem, however, also plagues research studies that collect their data prospectively. Indeed, potentially influential but unmeasured factors can be thought of for most social science phenomena. The most apparent level-1 predictor that was not included in available data sets was a measure of student pocket money, which has been shown to be an influential factor in youth's decision (and feasibility) to smoke that is not moderated by the school environment. One review, however, found socioeconomic status not strongly associated with smoking (Reid et al., 1995).

The available sample of schools and included students constituted a ‘convenience’ sample. That is, AYS administration allowed schools that were not selected through the random drawing of schools to ‘opt in’. Many schools do this in order to be provided with data. This means that some of the schools in the final dataset of this study may have differed systematically from schools not included in the final sample. The author investigated the representativeness of his final sample with respect to academic achievement scores (AZLearns scores). The mean of AZLearns scores of all Arizona schools was practically identical to the overall mean of the 198 schools in the final sample (4.03 vs. 3.93, respectively, $p = .326$). This indicates representativeness of the sample in terms of academic achievement. Moreover, because academic achievement and school poverty were so highly correlated, the final sample was also representative in terms of school poverty.

As for the representativeness of the resulting student sample, it was shown that the final student sample was a quite accurate reflection of the eighth grade student population of the state with respect to distribution over counties. In addition, the large level-1 sample size resulted in a high coverage rate of the counties’ eighth grade student population, for one county even exceeding 70 percent. For such high coverage, finite population corrections could have been applied to reduce standard errors of aggregate statistics. This was, however, beyond the scope of this research.

All level-1 measures came from student self-reports. Self-reports have a number of crucial shortcomings, among which is their propensity for social desirability. Especially in the context of program evaluations, where participants complete self-report instruments following the interventions, there may be special bias in terms of respondents trying to answer in a way they think pleases the intervention provider. In the case of smoking, this would lead to underreporting, because most interventions teach students that smoking is unwanted and non-normative behavior. AYS, however, was unrelated to the interventions and was therefore not affected by this shortcoming. Another challenging problem of self-report data that affected this research was missing data. Careful analyses ruled out missing data dependent on school status (control vs. intervention). Some systematic missingness patterns were identified and addressed by a combination of exclusion and retrieval strategies. Subsequent re-runs of analyses with imputed data indicated no negative effects of imputation on results but improved statistical accuracy by reducing standard errors due to higher sample size.

A methodological strength of this study is usage of school-level predictors (poverty rates, academic achievement and BTEP Prevention Index) that were not based on self-report. All level-2 predictor measures came from administrative records and can be assumed highly reliable because they were not based on student samples but the entire school population. By examining the incremental predictive utility of poverty and academic achievement and excluding poverty from final models, stability of HLM models was established by reducing multicollinearity on the school-level. In fact, if one knows about the academic achievement score of a school, knowledge of poverty rates will not improve prediction of the smoking issue in that school. A host of other school-level factors could have been included in HLM models to reduce level-2 error variance more (such as ethnic diversity of the school, measures on school ethos or educational style, etc.). The focus and scope of this study, however, was to shed a light on BTEP prevention effectiveness and effects of academic achievement. It seems fair to say that other conceivable level-2 predictors would have affected results of BTEP Prevention index little because even the powerful predictor of academic achievement did not moderate its results. School-level outcomes were reduced through extracting a PCA common factor, a more parsimonious procedure than that for analyzing level-1 outcomes (item-by-item), where such a data reduction method seemed less appropriate.

A number of features of the BTEP Prevention Index deserve brief discussion. The measure was a solution to the problem of lack of student-level intervention data. Not having student-level intervention data renders this study rather weak in terms of its power to detect possible BTEP intervention effects. Yet, conceptualizing the strength of intervention as a school-level characteristic was the only possible solution after all in order to obtain quantitative outcome results. One shortcoming of the index was its inability to account for implementation fidelity or distinguish between effects of different curricula or to represent the intervention as a multidimensional construct. The construction of the BTEP Prevention Index followed some recommendations of Sechrest et al. (1979) on how to assess 'strength of treatment'. It resembles somewhat the economic concept of the 'person-hour'. As such, it can be used whenever a comparison of outcomes is to be made between organizations (such as schools) depending on some input. This continuous index can also be assumed superior both conceptually and statistically to simple ANOVA-style distinctions between control group (coded 0) and intervention groups (coded 1). A cautionary note, however, must be made regarding the assignment of values of 'zero' on the BTEP Prevention Index to schools that did not receive anti-smoking interventions through BTEP contractors. It is very possible and likely that at least some schools with a BTEP Intervention Index of zero had some other form of prevention education, e.g. through funding mechanisms of the federal Office of Safe and Drug-free Schools within the US Department of Education. In that sense, some control schools were actually intervention schools. If at all, however, this would imply that effects of BTEP Intervention Index would have been much larger had only 'true' intervention schools been included.

This study was restricted to examining students in grade eight only. It was argued that most of these eighth grade students were most proximal in time to having received interventions when compared to tenth grade or twelfth grade students. For the proportion of students who had directly received programming in the past (i.e. prior to grade eight), this study was effectively a follow-up study on their smoking behavior and attitudes because the vast majority of prevention programming happened in grades below eight. In that sense, this study tested intervention effects at a ‘follow up’ time point and thereby realized a stipulation made by Dielman (1994) who suggested that behavioral effects of prevention education may occur only delayed and therefore should be measured at a follow-up time point.

It appears warranted to say that although no clear results emerged favoring the interventions (which, again, may be a function of the methodology and available data), the interventions did also not harm the subjects, e.g. by increasing the perceived favorability of smoking. Such adverse effects could happen in higher grades where students in “pubertal rebellion” would see smoking as cool due to the interventions because “cool” is whatever is diametrically opposed to what teachers say. Such counter-effects, however, may be limited to students who are more prone to high-risk behavior or deviance in the first place.

5.2. Review of results

Power calculations presented earlier (see chapter 3.3, p. 45) showed that small effect sizes may not be easily detected with available data. Overall, a number of findings deserve brief discussion. On the individual level, important risk factors and protective factors were identified that corroborate previous evidence. Older age, parental smoking, and American Indian and Latino ethnicity were found to increase both favorable views of smoking and risk of smoking. Parental disapproval of smoking and stronger bonding to parents were found to be protective factors. Looking only at the school level, the BTEP Intervention Index attained small effect sizes, even after controlling for 'outlier schools'. This was true for 'crude' ANOVAs as well as ecological regression analyses in which a number of confounders were entered. Percentage of American Indian students in a school was indicative of an aggravated smoking problem. This suggests that BTEP may seek collaboration with tribal education authorities to provide prevention services to schools with high percentages of American Indian students. Academic achievement as measured by the AZLearns rating system proved to be a strong predictor in all models. In that respect, it is a highly ecologically valid proxy measure for the tobacco epidemic in a given school. Its predictive validity surpassed poverty rates of the student body. BTEP may revise their policy recommendations so that providers are encouraged to first serve schools that are below average on the AZLearns system before targeting other schools. It needs to be kept in mind, however, that schools with below-average academic achievement may need a different quality (and quantity) of intervention programs.

Academic achievement remained a powerful variable even in HLM models, in which effects of the BTEP Prevention Index largely vanished. Academic achievement had substantial influence on school intercepts of level-1 outcomes after level-1 predictors had been introduced and improved model fit for almost all models and all datasets (all schools and Maricopa County schools only). Surprisingly, AZLearns scores also indicated higher parental smoking. This finding suggests an unexplored route of interventions to BTEP: students from low-achieving schools could be employed as anti-smoking messengers at home and encouragement of parents to utilize quit services provided by BTEP (e.g. the Arizona Smokers Helpline). With reach rates of 60,000 students, a potentially high number of parents could be reached with rather inexpensive means (such as info flyers).

BTEP Prevention Index exerted a weak influence on school intercepts in HLM models for Maricopa schools only (Table 21, p.113). Adjusted school means of number of best friends who smoked and Smoking Acquisition Index were smaller for Maricopa schools that had implemented more prevention activities than average. This finding seems to suggest that higher BTEP activity in these schools was associated with a climate that was less tolerant of smoking and may have led some students to either smoke not at all or be less committed to smoking (as measured by the Smoking Acquisition Index).

Further, the BTEP Intervention Index was largely unrelated to slopes between level-1 predictors and outcomes. Even narrowing analyses down to only Maricopa County schools did not yield substantial effects. This absence of effects may be due to the low power of the study to discover very small effects.

6. Conclusions

There are essentially two ways to do social science research. One paradigm suggests gathering data to answer research questions. The other asks what questions could be answered with already available data. The first philosophy is routinely followed in most research studies in psychology and medicine. The second one is oftentimes used by economists, epidemiologists and evaluators. Studies with archival data are especially attractive under notorious constraints of time and resources. There is, however, another set of circumstances that warrants so-called archival methods, i.e., the use of available data to answer questions for which they data may not necessarily have been collected. Empirical evidence is often needed in an applied setting when social programs need be evaluated but no evaluation data collection was implemented as the program was initiated. It could be said that the current study is an example of attempting to generate empirical, quantitative evidence under conditions of sub-optimal program evaluation planning and 'messy' program implementation parameters. In fact, the way BTEP lets its contractors implement prevention curricula would render a textbook research design (i.e. experimental design with longitudinal measurement of individuals and institutions) cost-prohibitive.

BETP have spent $4,400,000 on prevention activities in FY 06-07. Such expenditures have opportunity costs. Very small effects may not be the best expenditure of tax money if more cost-beneficial interventions could be funded instead. If all direct and indirect effects of BTEP school-based intervention efforts have prevented 1,000 students from picking up smoking until they passed the age of initiation (i.e. about 20 years of age), each prevented smoker would have cost $4,400. A little excursion into a possible cost-benefit-ratio of prevention programming seems appropriate, however, only very crude and without getting into the full technical and ethical complexity of calculating costs of smoking and by taking a reduced and ethically questionable societal perspective only concerned with saving tax dollars, i.e. costs of smoking not born out by the smoker or her family. Sloan et al. (2004) estimated that a 24-year old smoker accrues about \$6,200 in so-called external costs (in year 2000 dollars), net of money spent on cigarette taxes, which offset some of those external costs (it should be pointed out that these costs are tiny compared to the costs the smoker imposes on himself, which have been estimated at about \$144,000). The benefit-cost ratio in such a scenario (again, only looking at external costs) would have been 1.4 (\$6,200/\$4,400), so clearly in favor of the doing the prevention programming, even under such a dubiously narrow societal perspective. Under the scenario of just 100 prevented smokers due to all direct and indirect effects of BTEP activities, the ratio would have been 0.14 (\$6,200/\$44,000), a very bad investment.

Now, if one includes the perspective of the student who was prevented from becoming a smoker in the future into the societal perspective, which is an ethically more tenable perspective, the benefit-cost ratio with 1,000 prevented smokers would have been 34 (\$144,000+\$6,200/\$4,400), a truly astronomical return on investment. Even under the assumption of only 100 prevented smokers (that previously yielded a dismal return on investment), the benefit-cost ratio under the new societal perspective that includes the costs to the prevented future smoker would be 3.4 (\$144,000+\$6,200/\$44,000), or an excellent investment. Together, this suggests that even under very conservative assumptions, providing prevention activities may still be a cost-beneficial investment of tax money if the financial well-being of the smoker is taken into the societal perspective. Again, these estimates are very crude and should be interpreted heuristically rather than literal.

A number of policy recommendations appear to follow from findings in this study that may inform BTEP on ways implementation and strategies around youth prevention could be improved. A short bulleted list outlines recommendations that follow from results of this study.

School with high rates of American Indian students should be targeted. Culturally sensitive interventions must be selected to properly address and respect the different cultural meaning of tobacco to many Arizona tribes and their tobacco-related practices.

School with low scores on the AZLearns system should be targeted.

Schools with either high proportions of American Indian students or low AZLearns scores may benefit from offering cessation services (as they have higher prevalence rates).

Schools with serious smoking problems may benefit the most from a multi-pronged approach that also includes intervening with parents and the community at large (lower AZLearns ratings proxied for a higher rate of smoking parents at home). It seems reasonable to assume that the more comprehensive an anti-tobacco approach is, the more likely it is to “deliver”. That is, if implementing curriculum-based interventions, it should ideally be part of a grander strategy that involves parents, teachers & school administrators and the ‘community at large’, in order for anti-smoking lessons to be effective. Teaching students in the classroom about the dangers of smoking may not be very effective if a large share of the teacher body continues to smoke. The findings reported by Pizacani et al. (2008) suggest a synergistic effect of curriculum-based interventions and policy measures aimed at addressing the school environment and families of receiving students. It must be kept in mind that providing services in low-achievement schools constitutes an ‘uphill battle’ because students come from impoverished family backgrounds where smoking may be much more normative than for students from higher socio-economic strata.

Implementation of curricula should target the same students repeatedly in ascending grades. That is, curricula with booster sessions are recommended (e.g., Project Alert).

Increasing the dosage for those students most at risk instead of increasing the reach of prevention programming seems likely to increase impact (=reach x efficacy).

Most of BTEP's programming happens in lower grades (such as fourth and fifth). It may be beneficial or increase the overall impact to advise prevention providers to target relatively more higher grades (six through eight), because only in those grades does smoking become a behavior with personal relevance to students.

Further evaluation studies could either examine changes in prevalence rates of cohorts in intervention schools and control schools or gather longitudinal data on a select number of students in a random sample of schools. Such studies would hugely benefit from careful planning of data collection, both in terms of logistics and constructs. It would be desirable to get high-quality measures of intervention constructs, conceptualized as a multidimensional. These constructs would consist of several measures of implementation fidelity (e.g. what contents of a curriculum have been implemented), modality of exercises conducted during the lessons (e.g. extent to which more behavior-based exercises have been conducted, such as role play), and important other implementation features such as instructor training and expertise, school policies regarding smoking (such as sanction for tobacco violators), and instantiations of comprehensive school health programs that address the underlying common causes of health-compromising behavior.

Reference List

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human* Decision Processes, 50, 179-211.

Arizona Criminal Justice Commission (2006). *Arizona Youth Survey - State Report 2006*.

Arizona Department for Health Services (2007). *2006 Biennial Evaluation Report* Arizona Department for Health Services; Bureau for Tobacco Education and Prevention.

Arizona Department of Education (2002). 2002 Arizona School Health Education Profiles.

Arizona Department of Education (2007). Arizona's School Accountability System 2006 - Technical Manual.

Arizona Department of Health Services, (2003). Standards & recommendations for school-based prevention - Intensive curriculum. Internet Pamphlet

Aveyard, P., Markham, W. A., & Cheng, K. K. (2004). A methodological and substantive review of the evidence that schools cause pupils to smoke. *Social Science & Medicine, 58,* 2253-2265.

Aveyard, P., Markham, W. A., Lancashire, E., Bullock, A., Macarthur, C., Cheng, K. K. et al. (2004). The influence of school culture on smoking among pupils. *Social Science & Medicine, 58,* 1767-1780.

Bandura, A. (1991a). Social cognitive theory of self-regulation. *Organizational Behavior* and Human Decision Processes, 50, 248-287.

Battistich, V., Solomon, D., Kim, D. I., Watson, M., & Schaps, E. (1995). Schools As Communities, Poverty Levels of Student Populations, and Students Attitudes, Motives, and Performance - A Multilevel Analysis. *American Educational Research Journal, 32,* 627-658.

Bialous, S. & Glantz, S. (1997). Tobacco Control In Arizona 1973-1997.

Bickel, R. (2007). Multilevel analysis for applied research : it's just regression! New York : Guilford Press, c2007.

Botvin, G. J., Dusenbury, L., Baker, E., James-Ortiz, S., Botvin, E. M., & Kerner, J. (1992). Smoking prevention among urban minority youth: assessing effects on outcome and mediating variables. *Health Psychology, 11,* 290-299.

Bruvold, W. H. (1993). A Metaanalysis of Adolescent Smoking Prevention Programs. American Journal of Public Health, 83, 872-880.

Bryant, A. L., Schulenberg, J. E., O'Malley, P. M., Bachman, J. G., & Johnston, L. D. (2003). How academic achievement, attitudes, and behaviors relate to the course of substance use during adolescence: A 6-year, multiwave national longitudinal study. *Journal of Research on Adolescence, 13,* 361-397.

Bryk, A. S. & Raudenbush, S. W. (1992). Hierarchical Linear Models: Applications and Data Analysis Methods. Sage Pubns.

Cameron, R., Brown, K. S., Best, J. A., Pelkman, C. L., Madill, C. L., Manske, S. R. et al. (1999). Effectiveness of a social influences smoking prevention program as a function of provider type, training method, and school risk. *American Journal of Public Health, 89,* 1827-1831.

CDC (2006). Cigarette Use Among High School Students - United States, 1991-2005. MMWR, 55.

CDC (2007). Cigarette Smoking Among Adults - United States, 2006. *CDC MMWR, 56.*

CDC (2007). Best Practices for Comprehensive Tobacco Control Programs Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

Coffman, A. S. (2007). Results of the Comprehensive statewide Evaluation of TEPP's Intensive School-based Tobacco Prevention Education Program in Grades 4 - 8, 2005-06.

Cohen, J. (1992). A Power Primer. *Psychological Bulletin, 112,* 155-159.

Connor, O. (2000). SPSS and SAS Programs for Determining the Number of Components Using Parallel Analysis and Velicer's MAP Test. *Behavior Reseach Methods Instruments and Computers, 32,* 396-402.

Conrad, K. M., Flay, B. R., & Hill, D. (1992). Why Children Start Smoking Cigarettes - Predictors of Onset. British Journal of Addiction, 87, 1711-1724.

Datta, G. D., Subramanian, S. V., Colditz, G. A., Kawachi, I., Palmer, J. R., & Rosenberg, L. (2006). Individual, neighborhood, and state-level predictors of smoking among US Black women: A multilevel analysis. *Social Science & Medicine, 63,* 1034-1044.

Dielman, T. E. (1994). School-Based Research on the Prevention of Adolescent Alcohol Use and Misuse: Methodological Issues and Advances. *Journal of Research on Adolescence, 4,* 271-293.

Diez-Roux, A. V. (1998). Bringing context back into epidemiology: Variables and fallacies in multilevel analysis. American Journal of Public Health, 88, 216-222.

Ellickson, P. L., Bell, R. M., & Harrison, E. R. (1993). Changing Adolescent Propensities to Use Drugs - Results from Project Alert. *Health Education Quarterly, 20,* 227-242.

Evaluation Research and Development Unit (2008). Comprehensive Statewide Evaluation of BTEP's Intensive School-based Tobacco Prevention Education Program in Grades 4 – 8, FY 2006-07.

Figueredo, A. J. & Sechrest, L. (2001). Approaches used in conducting health outcomes and effectiveness research. Evaluation and Program Planning, 24, 41-59.

Flay, B. R. (1986). Efficacy and Effectiveness Trials (and Other Phases of Research) in the

Development of Health Promotion Programs. *Preventive Medicine, 15,* 451-474.

Flay, B. R. (1999). Understanding environmental, situational and intrapersonal risk and protective factors for youth tobacco use: The theory of triadic influence. *Nicotine and Tobacco Research, 1,* 111-114.

Friis, R. H. & Sellers, T. A. (2004). *Epidemiology for Public Health Practice*. Jones & Bartlett Publishers.

Harris, D. N. (2007). High-flying schools, student disadvantage, and the logic of NCLB. American Journal of Education, 113, 367-394.

Hawkins, J. D., Catalano, R. F., & Miller, J. Y. (1992). Risk and Protective Factors for Alcohol and Other Drug Problems in Adolescence and Early Adulthood - Implications for Substance-Abuse Prevention. *Psychological Bulletin, 112,* 64-105.

Kairouz, S. & Adlaf, E. M. (2003). Schools, students and heavy drinking: A multilevel analysis. Addiction Research & Theory, 11, 427-439.

Kaplan, R. M. (1990). Behavior as the central outcome in health care. *American* Psychologist, 45, 1211-1220.

Kumar, R., O'Malley, P. M., Johnston, L. D., Schulenberg, J. E., & Bachman, J. G. (2002). Effects of school-level norms on student substance use. *Preventive Scence., 3,* 105-124.

Leatherdale, S. T., Cameron, R., Brown, K. S., Jolin, M. A., & Kroeker, C. (2006). The influence of friends, family, and older peers on smoking among elementary school students: Low-risk students in high-risk schools. *Preventive Medicine, 42,* 218-222.

Leatherdale, S. T., Cameron, R., Brown, K. S., & McDonald, P. W. (2005). Senior student smoking at school, student characteristics, and smoking onset among junior students: a multilevel analysis. *Preventive Medicine, 40,* 853-859.

Liu, X.-F., Spybrook, J., Congdon, R., Martinez, A., & Raudenbush, S. (2008). Optimal Design (Version 1.77) [Computer software].

Maas, C. & Hox, J. (2005). Sufficient Sample Sizes for Multilevel Modeling. Methodology, 1, 86-92.

MacKinnon, D. P. (2008). Introduction to Statistical Mediation Analysis. Lawrence Erlbaum Associates.

McKnight, P. E., McKnight, K. M., Sidani, S. & Figueredo, A. J. (2007). *Missing Data: A Gentle Introduction*. The Guilford Press.

Murray, D. M. & Hannan, P. J. (1990). Planning for the appropriate analysis in school-based drug-use prevention studies. Journal of Consulting and Clinical Psychology, 58, 458-468.

Palmer, R. F., Graham, J. W., White, E. L., & Hansen, W. B. (1998). Applying multilevel analytic strategies in adolescent substance use prevention research. Preventive Medicine: An International Journal Devoted to Practice and Theory, 27, 328-336.

Pentz, M. A. (1989). The power of policy: the relationship of smoking policy to adolescent smoking. American Journal of Public Health 79 [7], 857-862.

Peterson, A. V., Kealey, K. A., Mann, S. L., Marek, P. M., & Sarason, I. G. (2000). Hutchinson smoking prevention project: Long-term randomized trial in school-based tobacco use prevention - Results on smoking. *Journal of the National Cancer Institute, 92,* 1979-1991.

Pizacani, B. A., Dent, C. W., Maher, J. E., Rohde, K., Stark, M. J., Biglan, A. et al. (2008). Smoking Patterns in Oregon Youth: Effects of Funding and Defunding of a Comprehensive State Tobacco Control Program. *Journal of Adolescent Health*.

Reid, D. J., McNeill, A. D., & Glynn, T. J. (1995). Reducing the prevalence of smoking in youth in Western countries: an international review. *British Medical Journal, 4,* 266-277.

Resnick, M. D., Bearman, P. S., Blum, R. W., Bauman, K. E., Harris, K. M., Jones, J. et al.

(1997). Protecting adolescents from harm - Findings from the National Longitudinal Study on Adolescent Health. *Journal of the American Medical Association, 278,* 823-832.

Rohde, K., Pizacani B., Stark, M., Pietrukowicz, M., , M. C., Romoli, C. et al. (2001). Effectiveness of school-based programs as a component of a statewide tobacco control initiative--Oregon, 1999-2000. *MMWR Morbidity and Mortality Weekly Report, 50,* 663-666.

Schwartz, S. (1994). The Fallacy of the Ecological Fallacy - the Potential Misuse of A Concept and the Consequences. *American Journal of Public Health, 84,* 819-824.

Sechrest, L., West, S. G., Phillips, M. A., Redner, R., & Yeaton, W. (1979). Some neglected problems in evaluation research: Strength and integrity of treatments. *Evaluation Studies Annual Review, 4,* 15-35.

Sellstrom, E. & Bremberg, S. (2006). Is there a "school effect" on pupil outcomes? A review of multilevel studies. Journal of Epidemiology and Community Health, 60, 149-155.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-* Experimental Designs for Generalized Causal Inference. Houghton Mifflin Co. Boston, MA.

Shohaimi, S., Luben, R., Wareham, N., Day, N., Bingham, S., Welch, A. et al. (2003). Residential area deprivation predicts smoking habit independently of individual educational level and occupational social class. A cross sectional study in the Norfolk cohort of the European Investigation into Cancer (EPIC-Norfolk). *Journal of Epidemiology & Community Health 57*[4], 270-276.

Sloan FA, Ostermann J, Picone G, Conover C, Taylor D, Jr. (2004) The price of smoking. Cambridge (Massachusetts): MIT Press.

Snijders, T. A. B. & Bosker, R. J. (1999). Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling. Sage Publications Inc.

Sussman, S., Hansen, W. B., Flay, B. R., & Botvin, G. J. (2001). Re: Hutchinson Smoking Prevention Project: Long-Term Randomized Trial in School-Based Tobacco Use Prevention-Results on Smoking. *Journal of the National Cancer Institute 93* [16], 1267.

Thomas, R. (2002). School-based programmes for preventing smoking. *The Cochrane* Database of Systematic Reviews, 1465-1858.

Tyas, S. L. & Pederson, L. L. (1998). Psychosocial factors related to adolescent smoking: a critical review of the literature. *Tobacco Control, 7,* 409-420.

West, P. (2006). School effects research provide new and stronger evidence in support of the health-promoting school idea. *Health Education, 106,* 421-424.

West, P., Sweeting, H., & Leyland, A. (2004). School effects on pupils' health behaviours: evidence in support of the health promoting school. *Research Papers in Education, 19,* 261-291.

Wiehe, S. E., Garrison, M. M., Christakis, D. A., Ebel, B. E., & Rivara, F. P. (2005). A systematic review of school-based smoking prevention trials with long-term follow-up. *Journal of Adolescent Health, 36,* 162-169.

Wills, T. A., Pierce, J. P., & Evans, R. I. (1996). Large-scale environmental risk factors for substance use. American Behavioral Scientist, 39, 808-822.

Wittmann, W. & Klumb, P. (2006). How to fool yourself with experiments in testing theories in psychological research. In Richard R.Bootzin & Patrick E.McKnight (Eds.), *Strengthening research methodology : psychological measurement and evaluation* (Washington, DC: American Psychological Association.

Wittmann, W., Nuebling, R., & Schmidt, J. (2002). Evaluationsforschung und Programevaluation im Gesundheitswesen. *Zeitschrift fuer Evaluation, 1,* 39-60.

Wittmann, W. W. (1990). Brunswik-Symmetrie und die Konzeption der Fuenf-Datenboxen--Ein Rahmenkonzept fuer umfassende Evaluationsforschung. *Zeitschrift fuer Paedagogische Psychologie, 4,* 241-251.

World Health Organization (2008). WHO Report on the Global Tobacco Epidemic, 2008: The MPOWER package Geneva.

Appendix A – Class Form School Year 2005-06

Class Form 2005-06 Tobacco Class Evaluation

Please complete one form for each class that participates in the evaluation. Thank you!

1. Today's Date: ________________________________

2. School: ____

3. Grade level: ___ 4th ___5th ___6th ___7th ___8th

5. Name of Classroom Teacher:____

6. Class name (example: 7th Grade Science): ____

7. Name of Tobacco Class Instructor: ____

8. Persons involved in teaching this TEPP Tobacco Prevention Education Class (check all that apply):

Staff
___ TEPP Local Project Staff
___ Classroom teacher
___ School Counselor
___ Social Worker
___ Coach/PE teacher
___ School Nurse
___ School Prevention Educator/Coordinator
___ Other Staff
(specify ___________________________)

Volunteers
___ Peer educators
___ Fireman
___ Police officer
___ Former tobacco user
___ Other volunteer
(specify ___________________________)

9. TEPP Tobacco Education was taught during:

___ Homeroom ___ Science ___ Health ___ PE ___ Health/PE ___ Math
___ Language Arts ___ Social Studies ___ Other (specify: ___________________________________)

10. Date tobacco class started: ___ / ___ /______ (mm dd yyyy) Date tobacco class ended: ___ / ___ /______ (mm dd yyyy)

11. Number of students enrolled in this tobacco class: ______________

12. Number of students who participated in the tobacco class evaluation: ______________

13. Name of curriculum taught ___ Project Alert ____ Get Real About Tobacco ____ TNT

Other (specify) ____

14. Total number of lessons covered from the curriculum: __________

15. How many class sessions were taught? _______

16. On average, how many days per week was the class taught? __1 __2 __3 __4 __5

17. On average, how many minutes did each class session last? ______

18. Approximately what proportion of this class was taught in Spanish? __ 0% __ 25% _50% _75% _100%

19. Comment on key successes or barriers to the desired outcome of this class (continue on back if desired).

____

Appendix B – Contractual Requirements for Implementing School-Based Prevention, Maricopa County

Maricopa County
Tobacco Use Prevention Program
Program Contractual Requirements

The following program components are required as part of the partnership between MACTUPP and your school. These activities will assist your school in creating an effective program.

- "Tobacco-Free Campus" signs must be posted and visible on school campus. Your campus may already have a 'Drug-Free' sign, however most adult smokers don't consider tobacco a drug. Therefore, you must post a Tobacco-Free campus sign.

- Attend and participate in an individual or group (preferred) site coordinator orientation provided by your assigned MACTUPP Prevention Specialist or MACTUPP trainer.

- Tobacco Prevention Curriculum must come from the Maricopa County Tobacco Use Prevention Program's approved list of curricula. The curriculum must be implemented in **ONE ENTIRE GRADE LEVEL** within the 4^{th}-8^{th} grade target population. The curriculum must be implemented with fidelity, meaning that the required lessons are taught fully and in the designated order.

- The site coordinator will be responsible for administering the tobacco class evaluation survey to all classes receiving the curriculum. All materials for the survey will be provided. **The survey should be administered the day of the last lesson or the next day**. Surveys are mailed out to you based on the number of completed lessons indicated on your monthly report. Contact your assigned Prevention Specialist if you need assistance.

- Ten tobacco prevention messages are required during the contract period. These messages can include, but are not limited to morning announcements, articles in the school paper, announcements at sporting events, print ads in yearbooks, and videotaped commercials from the project.

- One youth-driven tobacco use prevention event must take place on campus during the school year. **This activity should be student-led**.

- Monthly reports must be completed accurately and submitted consistently by the **3rd of each month**.

- Contact your assigned Prevention Specialist if you need assistance completing any of the contractual requirements outlined above.

www.mactupp.org

Appendix C – HLM6 Formulas

1) Fully unconditional models (models (a) presented in Table 20)

1a) Binary outcome (model 1a in Table 20)

LEVEL 1 MODEL (bold: group-mean centering; bold italic: grand-mean centering)

Prob(CIGLIFER=1|β) = φ

Log[φ/(1 - φ)] = η

$\eta = \beta_0$

LEVEL 2 MODEL (bold italic: grand-mean centering)

$\beta_0 = \gamma_{00} + u_0$

$$\eta = \gamma_{00} + u_0$$

1b) Continuous outcome (model 3a in Table 20)

LEVEL 1 MODEL (bold: group-mean centering; bold italic: grand-mean centering)

SMKNORM = $\beta_0 + r$

LEVEL 2 MODEL (bold italic: grand-mean centering)

$\beta_0 = \gamma_{00} + u_0$

$$\text{SMKNORM} = \gamma_{00} + u_0 + r$$

2) Level-1 predictors with randomly varying intercepts (models (b) presented in Table 20)

2a) Binary outcome (model 1b in Table 20)

LEVEL 1 MODEL (bold: group-mean centering; bold italic: grand-mean centering)

Prob(CIGLIFER=1|β) = φ

Log[φ/(1 - φ)] = η

$\eta = \beta_0 + \beta_1(\boldsymbol{SEX01}) + \beta_2(\mathbf{Q2}) + \beta_3(\boldsymbol{LATINO01}) + \beta_4(\boldsymbol{AMIND01}) + \beta_5(\boldsymbol{Q83B01}) + \beta_6(\boldsymbol{Q117BREC}) + \beta_7(\boldsymbol{FAMILYBO})$

LEVEL 2 MODEL (bold italic: grand-mean centering)

$\beta_0 = \gamma_{00} + u_0$

$\beta_1 = \gamma_{10}$

$\beta_2 = \gamma_{20}$

$\beta_3 = \gamma_{30}$

$\beta_4 = \gamma_{40}$

$\beta_5 = \gamma_{50}$

$\beta_6 = \gamma_{60}$

$\beta_7 = \gamma_{70}$

2b) Continuous outcome (model 3b in Table 20)

LEVEL 1 MODEL (bold: group-mean centering; bold italic: grand-mean centering)

SMKNORM = $\beta_0 + \beta_1(\boldsymbol{SEX01}) + \beta_2(\mathbf{Q2}) + \beta_3(\boldsymbol{LATINO01}) + \beta_4(\boldsymbol{AMIND01}) + \beta_5(\boldsymbol{Q83B01}) + \beta_6(\boldsymbol{Q117BREC}) + \beta_7(\boldsymbol{FAMILYBO}) + r$

LEVEL 2 MODEL (bold italic: grand-mean centering)

$\beta_0 = \gamma_{00} + u_0$

$\beta_1 = \gamma_{10}$

$\beta_2 = \gamma_{20}$

$\beta_3 = \gamma_{30}$

$\beta_4 = \gamma_{40}$

$\beta_5 = \gamma_{50}$

$\beta_6 = \gamma_{60}$

$\beta_7 = \gamma_{70}$

SMKNORM = $\gamma_{00} + \gamma_{10}*\boldsymbol{SEX01} + \gamma_{20}*\mathbf{Q2} + \gamma_{30}*\boldsymbol{LATINO01} + \gamma_{40}*\boldsymbol{AMIND01} + \gamma_{50}*\boldsymbol{Q83B01} +$ *
$\gamma_{60}*\boldsymbol{Q117BREC} + \gamma_{70}*\boldsymbol{FAMILYBO} + u_0 + r$

3) Random intercepts with level-2 predictors and level-1 predictors (models (e) presented in Table 20)

3a) Binary outcome (model 1e in Table 20)

LEVEL 1 MODEL (bold: group-mean centering; bold italic: grand-mean centering)

Prob(CIGLIFER=1|β) = φ

Log[φ/(1 - φ)] = η

$\eta = \beta_0 + \beta_1(\boldsymbol{SEX01}) + \beta_2(\mathbf{Q2}) + \beta_3(\boldsymbol{LATINO01}) + \beta_4(\boldsymbol{AMIND01}) + \beta_5(\mathbf{Q83B01}) + \beta_6(\boldsymbol{Q117BREC}) + \beta_7(\boldsymbol{FAMILYBO})$

LEVEL 2 MODEL (bold italic: grand-mean centering)

$\beta_0 = \gamma_{00} + \gamma_{01}(\boldsymbol{ALLCURR}) + \gamma_{02}(\boldsymbol{AZLEARNS}) + u_0$

$\beta_1 = \gamma_{10}$

$\beta_2 = \gamma_{20}$

$\beta_3 = \gamma_{30}$

$\beta_4 = \gamma_{40}$

$\beta_5 = \gamma_{50}$

$\beta_6 = \gamma_{60}$

$\beta_7 = \gamma_{70}$

$$\eta = \gamma_{00} + \gamma_{01}{*}\boldsymbol{ALLCURR} + \gamma_{02}{*}\boldsymbol{AZLEARNS} + \gamma_{10}{*}\boldsymbol{SEX01} + \gamma_{20}{*}\boldsymbol{Q2} + \gamma_{30}{*}\boldsymbol{LATINO01} + \gamma_{40}{*}\boldsymbol{AMIND01} + \gamma_{50}{*}\boldsymbol{Q83B01} + \gamma_{60}{*}\boldsymbol{Q117BREC} + \gamma_{70}{*}\boldsymbol{FAMILYBO} + u_0$$

3b) Continuous outcome (model 3e in Table 20)

LEVEL 1 MODEL (bold: group-mean centering; bold italic: grand-mean centering)

$SMKNORM = \beta_0 + \beta_1(\boldsymbol{SEX01}) + \beta_2(\mathbf{Q2}) + \beta_3(\boldsymbol{LATINO01}) + \beta_4(\boldsymbol{AMIND01}) + \beta_5(\mathbf{Q83B01}) + \beta_6(\boldsymbol{Q117BREC}) + \beta_7(\boldsymbol{FAMILYBO}) + r$

LEVEL 2 MODEL (bold italic: grand-mean centering)

$\beta_0 = \gamma_{00} + \gamma_{01}(\boldsymbol{ALLCURR}) + \gamma_{02}(\boldsymbol{AZLEARNS}) + u_0$

$\beta_1 = \gamma_{10}$

$\beta_2 = \gamma_{20}$

$\beta_3 = \gamma_{30}$

$\beta_4 = \gamma_{40}$

$\beta_5 = \gamma_{50}$

$\beta_6 = \gamma_{60}$

$\beta_7 = \gamma_{70}$

$$SMKNORM = \gamma_{00} + \gamma_{01}{*}\boldsymbol{ALLCURR} + \gamma_{02}{*}\boldsymbol{AZLEARNS} + \gamma_{10}{*}\boldsymbol{SEX01} + \gamma_{20}{*}\boldsymbol{Q2} + \gamma_{30}{*}\boldsymbol{LATINO01} + \gamma_{40}{*}\boldsymbol{AMIND01} + \gamma_{50}{*}\boldsymbol{Q83B01} + \gamma_{60}{*}\boldsymbol{Q117BREC} + \gamma_{70}{*}\boldsymbol{FAMILYBO} + u_0 + r$$

4) Random intercepts & random slopes, no level-2 predictors

4a) Binary outcome (results in Table 22)

LEVEL 1 MODEL (bold: group-mean centering; bold italic: grand-mean centering)

Prob(CIGLIFER=1|β) = φ

Log[φ/(1 - φ)] = η

$$\eta = \beta_0 + \beta_1(\textbf{\textit{SEX01}}) + \beta_2(\textbf{Q2}) + \beta_3(\textbf{\textit{LATINO01}}) + \beta_4(\textbf{\textit{AMIND01}}) + \beta_5(\textbf{\textit{Q83B01}}) + \beta_6(\textbf{\textit{Q117BREC}}) + \beta_7(\textbf{\textit{FAMILYBO}})$$

LEVEL 2 MODEL (bold italic: grand-mean centering)

$$\beta_0 = \gamma_{00} + u_0$$
$$\beta_1 = \gamma_{10}$$
$$\beta_2 = \gamma_{20}$$
$$\beta_3 = \gamma_{30}$$
$$\beta_4 = \gamma_{40}$$
$$\beta_5 = \gamma_{50} + u_5$$
$$\beta_6 = \gamma_{60} + u_6$$
$$\beta_7 = \gamma_{70} + u_7$$

$$\eta = \gamma_{00} + \gamma_{10}*\textbf{\textit{SEX01}} + \gamma_{20}*\textbf{Q2} + \gamma_{30}*\textbf{\textit{LATINO01}} + \gamma_{40}*\textbf{\textit{AMIND01}} + \gamma_{50}*\textbf{\textit{Q83B01}} + \gamma_{60}*\textbf{\textit{Q117BREC}} + \gamma_{70}*\textbf{\textit{FAMILYBO}} + u_0 + u_5*\textbf{\textit{Q83B01}} + u_6*\textbf{\textit{Q117BREC}} + u_7*\textbf{\textit{FAMILYBO}}$$

4b) Continuous outcome (model 3e in Table 22)

LEVEL 1 MODEL (bold: group-mean centering; bold italic: grand-mean centering)

$$\text{SMKNORM} = \beta_0 + \beta_1(\textbf{\textit{SEX01}}) + \beta_2(\textbf{Q2}) + \beta_3(\textbf{\textit{LATINO01}}) + \beta_4(\textbf{\textit{AMIND01}}) + \beta_5(\textbf{\textit{Q83B01}}) + \beta_6(\textbf{\textit{Q117BREC}}) + \beta_7(\textbf{\textit{FAMILYBO}}) + r$$

LEVEL 2 MODEL (bold italic: grand-mean centering)

$$\beta_0 = \gamma_{00} + u_0$$
$$\beta_1 = \gamma_{10}$$
$$\beta_2 = \gamma_{20}$$
$$\beta_3 = \gamma_{30}$$
$$\beta_4 = \gamma_{40}$$
$$\beta_5 = \gamma_{50} + u_5$$
$$\beta_6 = \gamma_{60} + u_6$$
$$\beta_7 = \gamma_{70} + u_7$$

$$\text{SMKNORM} = \gamma_{00} + \gamma_{10}*\textbf{\textit{SEX01}} + \gamma_{20}*\textbf{Q2} + \gamma_{30}*\textbf{\textit{LATINO01}} + \gamma_{40}*\textbf{\textit{AMIND01}} + \gamma_{50}*\textbf{\textit{Q83B01}} + * \; \gamma_{60}*\textbf{\textit{Q117BREC}} + \gamma_{70}*\textbf{\textit{FAMILYBO}} + u_0 + u_5*\textbf{\textit{Q83B01}} + u_6*\textbf{\textit{Q117BREC}} + u_7*\textbf{\textit{FAMILYBO}} + r$$

5) Intercepts/Slopes-as-outcomes models (level-2 predictors for randomly varying slopes)

5a) Binary outcome (model 1h in Table 24)

LEVEL 1 MODEL (bold: group-mean centering; bold italic: grand-mean centering)

Prob(CIGLIFER=1|β) = φ

Log[φ/(1 - φ)] = η

$$\eta = \beta_0 + \beta_1(\textbf{\textit{SEX01}}) + \beta_2(\textbf{Q2}) + \beta_3(\textbf{\textit{LATINO01}}) + \beta_4(\textbf{\textit{AMIND01}}) + \beta_5(\textbf{Q83B01}) + \beta_6(\textbf{\textit{Q117BREC}}) + \beta_7(\textbf{\textit{FAMILYBO}})$$

LEVEL 2 MODEL (bold italic: grand-mean centering)

$$\beta_0 = \gamma_{00} + \gamma_{01}(\textbf{\textit{ALLCURR}}) + \gamma_{02}(\textbf{\textit{AZLEARNS}}) + u_0$$
$$\beta_1 = \gamma_{10}$$
$$\beta_2 = \gamma_{20}$$
$$\beta_3 = \gamma_{30}$$
$$\beta_4 = \gamma_{40}$$
$$\beta_5 = \gamma_{50} + \gamma_{51}(\textbf{\textit{ALLCURR}}) + \gamma_{52}(\textbf{\textit{AZLEARNS}}) + u_5$$
$$\beta_6 = \gamma_{60} + \gamma_{61}(\textbf{\textit{ALLCURR}}) + \gamma_{62}(\textbf{\textit{AZLEARNS}}) + u_6$$
$$\beta_7 = \gamma_{70} + \gamma_{71}(\textbf{\textit{ALLCURR}}) + \gamma_{72}(\textbf{\textit{AZLEARNS}}) + u_7$$

$$\eta = \gamma_{00} + \gamma_{01}*\textbf{\textit{ALLCURR}} + \gamma_{02}*\textbf{\textit{AZLEARNS}} + \gamma_{10}*\textbf{\textit{SEX01}} + \gamma_{20}*\textbf{Q2} + \gamma_{30}*\textbf{\textit{LATINO01}} + \gamma_{40}*\textbf{\textit{AMIND01}} + \gamma_{50}*\textbf{Q83B01} + \gamma_{51}*\textbf{\textit{ALLCURR}}*\textbf{Q83B01} +$$
$$\gamma_{52}*\textbf{\textit{AZLEARNS}}*\textbf{Q83B01} + \gamma_{60}*\textbf{\textit{Q117BREC}} + \gamma_{61}*\textbf{\textit{ALLCURR}}*\textbf{\textit{Q117BREC}} + \gamma_{62}*\textbf{\textit{AZLEARNS}}*\textbf{\textit{Q117BREC}} + \gamma_{70}*\textbf{\textit{FAMILYBO}} + \gamma_{71}*\textbf{\textit{ALLCURR}}*\textbf{\textit{FAMILYBO}}$$
$$\gamma_{72}*\textbf{\textit{AZLEARNS}}*\textbf{\textit{FAMILYBO}} + u_0 + u_5*\textbf{Q83B01} + u_6*\textbf{\textit{Q117BREC}} + u_7*\textbf{\textit{FAMILYBO}}$$

5b) Continuous outcome (model 3h in Table 24)

LEVEL 1 MODEL (bold: group-mean centering; bold italic: grand-mean centering)

$$\text{SMKNORM} = \beta_0 + \beta_1(\textbf{\textit{SEX01}}) + \beta_2(\textbf{Q2}) + \beta_3(\textbf{\textit{LATINO01}}) + \beta_4(\textbf{\textit{AMIND01}}) + \beta_5(\textbf{Q83B01}) + \beta_6(\textbf{\textit{Q117BREC}}) + \beta_7(\textbf{\textit{FAMILYBO}}) + r$$

LEVEL 2 MODEL (bold italic: grand-mean centering)

$$\beta_0 = \gamma_{00} + \gamma_{01}(\textbf{\textit{ALLCURR}}) + \gamma_{02}(\textbf{\textit{AZLEARNS}}) + u_0$$
$$\beta_1 = \gamma_{10}$$
$$\beta_2 = \gamma_{20}$$
$$\beta_3 = \gamma_{30}$$
$$\beta_4 = \gamma_{40}$$
$$\beta_5 = \gamma_{50} + \gamma_{51}(\textbf{\textit{ALLCURR}}) + \gamma_{52}(\textbf{\textit{AZLEARNS}}) + u_5$$
$$\beta_6 = \gamma_{60} + \gamma_{61}(\textbf{\textit{ALLCURR}}) + \gamma_{62}(\textbf{\textit{AZLEARNS}}) + u_6$$
$$\beta_7 = \gamma_{70} + \gamma_{71}(\textbf{\textit{ALLCURR}}) + \gamma_{72}(\textbf{\textit{AZLEARNS}}) + u_7$$

$$\text{SMKNORM} = \gamma_{00} + \gamma_{01}*\textbf{\textit{ALLCURR}} + \gamma_{02}*\textbf{\textit{AZLEARNS}} + \gamma_{10}*\textbf{\textit{SEX01}} + \gamma_{20}*\textbf{Q2} + \gamma_{30}*\textbf{\textit{LATINO01}} + \gamma_{40}*\textbf{\textit{AMIND01}} + \gamma_{50}*\textbf{Q83B01} + \gamma_{51}*\textbf{\textit{ALLCURR}}*\textbf{Q83B01} +$$
$$\gamma_{52}*\textbf{\textit{AZLEARNS}}*\textbf{Q83B01} + \gamma_{60}*\textbf{\textit{Q117BREC}} + \gamma_{61}*\textbf{\textit{ALLCURR}}*\textbf{\textit{Q117BREC}} + \gamma_{62}*\textbf{\textit{AZLEARNS}}*\textbf{\textit{Q117BREC}} + \gamma_{70}*\textbf{\textit{FAMILYBO}} + \gamma_{71}*\textbf{\textit{ALLCURR}}*\textbf{\textit{FAMILYE}}$$
$$+ \gamma_{72}*\textbf{\textit{AZLEARNS}}*\textbf{\textit{FAMILYBO}} + u_0 + u_5*\textbf{Q83B01} + u_6*\textbf{\textit{Q117BREC}} + u_7*\textbf{\textit{FAMILYBO}} + r$$

Eidesstattliche Erklärung/Affidavit

Ich versichere, dass ich die beiliegende Dissertation mit dem Titel "Outcome Evaluation of Arizona’s School-Based Smoking Prevention Program – A Multilevel Study" ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

I hereby declare that I have developed and written this doctoral dissertation entitled "Outcome Evaluation of Arizona’s School-Based Smoking Prevention Program –
A Multilevel Study" entirely by myself and have not used sources or means other than those cited in the text. Any quotations from these sources were clearly marked as such. This doctoral dissertation was not submitted in the same or a similar version, not even partially, to any other authority to achieve an academic title or degree and was not published elsewhere.

Tucson, Arizona, USA, Sept.2009

____

Frederic Malter

Endnotes

---

[i] http://www.dhss.mo.gov/InterventionMICA/Tobacco/GroupEducation/index_3.html, accessed 08-10-2008.

[ii] http://www.rmc.org/K12/Docs/FactSheets.pdf, accessed 08-10-2008.

[iii] http://www.stat.psu.edu/~jls/misoftwa.html, accessed 10-01-2008

[iv] http://sitemaker.umich.edu/group-based/optimal_design_software, accessed 11-12-20