

Contract No.: LC91015001
MPR Reference No.: 8014

**DO WE KNOW WHOM TO SERVE?
ISSUES IN USING RISK FACTORS
TO IDENTIFY DROPOUTS**

June 1998

Author:
Philip Gleason
Mark Dynarski

A Research Report from the School Dropout
Demonstration Assistance Program Evaluation

Submitted to:

U.S. Department of Education
Planning and Evaluation Service
600 Independence Avenue, S.W.
Room 4168, FOB-10
Washington, DC 20202-4246

Project Officer:
Audrey Pendleton

Submitted by:

Mathematica Policy Research, Inc.
P.O. Box 2393
Princeton, NJ 08543-2393
(609) 799-3535

Project Director:
Mark Dynarski

CONTENTS

| | Page |
|---|-------------|
| EXECUTIVE SUMMARY | iv |
| A. INTRODUCTION | 1 |
| B. PAST RESEARCH HAS IDENTIFIED RISK FACTORS | 2 |
| C. LONGITUDINAL DATA ALLOW RISK FACTORS TO BE ASSESSED | 5 |
| 1. Longitudinal Data Are from Middle Schools and High Schools in Four Cities | 6 |
| 2. Single Risk Factors | 7 |
| 3. Risk Factors Based on Multiple Characteristics | 8 |
| D. MOST RISK FACTORS ARE INEFFICIENT | 10 |
| E. CAN OTHER FACTORS PREDICT DROPPING OUT BETTER? | 17 |
| REFERENCES | 21 |
| APPENDIX A: CREATING THE PROPENSITY SCORE RISK FACTOR | A.1 |

TABLES

| Table | | Page |
|--------------|--|-------------|
| 1 | MIDDLE SCHOOL RISK FACTOR EFFICIENCY | 11 |
| 2 | HIGH SCHOOL RISK FACTOR EFFICIENCY | 15 |

EXECUTIVE SUMMARY

A common approach taken by dropout prevention programs is to provide intensive services to a targeted group of at-risk students. Programs using this approach typically identify at-risk students for participation by selecting those with particular characteristics or measures of past school performance that the research literature has found to correlate with dropping out. The idea is that these are the students most in need of dropout prevention services--that is, most likely to drop out if they do not receive such services. But do these characteristics, or *risk factors*, efficiently predict which students ultimately will drop out (in the absence of program intervention)?

In this report, we evaluate the predictive efficiency, or validity, of risk factors commonly used by dropout prevention programs. We also construct and evaluate a more comprehensive dropout prevention scale--the propensity score risk factor--to determine the benefits of incorporating more information into the process of predicting dropping out. Much research has examined which factors are correlated with dropping out. However, less attention has been paid to assessing how well these factors predict dropping out from the perspective of a dropout prevention program using risk factors to try to identify future dropouts.

We use data from the evaluation of the School Dropout Demonstration Assistance Program (SDDAP), sponsored by the U.S. Department of Education. In particular, we measure risk factors among students beginning the 7th and 10th grades at schools in four SDDAP sites undertaking schoolwide reform activities (as well as at comparison group schools in these sites). The dropout status of these students is measured two to three years later, when the students in the middle school sample should be completing the 8th or 9th grade and students in the high school sample should be completing the 11th or 12th grade.

The risk factors we examine include about 20 student characteristics or measures of past school performance that previous studies have found to be correlated with dropping out. In addition, we construct the propensity score risk factor by taking into account these factors and about 20 additional student characteristics, using a logistic regression model to translate the characteristics into a single measure of students' propensity to drop out. We consider the students to be at risk if their "propensity score" exceeds a specific threshold.

Commonly used risk factors are inefficient predictors of whether students drop out. In both the middle and high school samples, at-risk students identified by these risk factors are more likely to have remained in school two to three years later than to have dropped out. For example, for the middle school sample, high absenteeism and being overage for grade by two or more years are the most efficient risk factors among individual student characteristics. However, only 15 percent of students with either of these risk factors have dropped out within two to three years. Although this dropout rate is much higher than the overall dropout rate among the sample (six percent), a dropout prevention program using either of these risk factors would serve many more students who do not need dropout prevention services than who do. Furthermore, a dropout prevention program of a reasonable size (with one slot for each future dropout in the school) using either of these risk factors would fail to serve a large fraction of the future dropouts in a school.

Risk factors predict dropping out no better at the high school level. When individual student characteristics are used as risk factors, the best of these (high absenteeism, overage by two years, low grades, has a child) identify a group of students among whom less than one-third drop out. Again, this dropout rate exceeds the 15 percent dropout rate among the full high school sample but would identify future dropouts relatively inefficiently for a dropout prevention program.

The propensity score risk factor predicts dropping out only slightly more efficiently, despite using much more information than individual risk factors to attempt to distinguish dropouts from those who remain in school. The dropout rate among those defined as at risk by the propensity score risk factor is 23 percent among middle school students and 42 percent among high school students. Thus, a dropout prevention program using this risk factor to identify students destined to drop out within two to three years would be incorrect more often than correct in identifying these future dropouts. The program would serve many students who would not need dropout prevention services and fail to serve many others who do.

One implication of these findings is that dropout prevention programs that identify at-risk students by using risk factors of the type examined in this report are unlikely to reduce the dropout rate substantially, no matter how effective their dropout prevention services. Unless programs can efficiently identify and serve the students destined to drop out without intervention, they cannot prevent these students from dropping out. At the high school level, for example, a hypothetical program that prevented half the future dropouts it served from dropping out and that used the propensity score risk factor to identify participants would lower the overall dropout rate in a school from 15 percent to 12 percent.

If programs are to continue to combat dropping out by targeting services toward at-risk students, they must work hard to find efficient means of identifying future dropouts. Our results suggest that this task is far more difficult than simply looking at known risk factors that are correlated with dropping out. Programs somehow must obtain student information that more reliably predicts whether students will drop out. We suggest four types of information not used in this study but perhaps available to program operators: (1) ecological characteristics, such as neighborhood conditions or peer group effects; (2) unobserved psychological factors; (3) measures of the persistence of specific characteristics over time; and (4) transitory events that occur closer to the time that students decide to drop out.

We do not know whether this additional information will allow dropout prevention programs to identify future dropouts more efficiently. However, we do know that the risk factors examined in this report do not identify future dropouts efficiently enough to allow the programs using them to substantially reduce the dropout rate. Programs must better understand why students drop out if they are to more efficiently identify program participants who will benefit from dropout prevention services.

A. INTRODUCTION

Dropout prevention programs commonly provide intensive services to a group of students within a school or district deemed most likely to drop out. To identify these “at-risk” students, program operators typically use risk factors--student characteristics or measures of past school performance thought to be associated with future dropping out. If programs using this approach are to be effective in reducing the dropout rate, then risk factors must effectively identify the students who would, in fact, drop out if they did not receive program services. Even the highest-quality dropout prevention programs will have little influence on the dropout problem if the wrong students are served.

Although research has examined thoroughly student characteristics or past performance measures correlated with dropping out, many of these studies have not directly assessed the predictive validity of these risk factors. In particular, if a dropout prevention program uses a given risk factor to identify participants, what proportion of these participants would drop out in the absence of receiving services? Alternatively, what proportion of the “future dropouts” of a school or community would be served by a program if that program used a particular risk factor to identify program participants? And, if a particular risk factor used by a dropout prevention program does not predict the rate of dropping out well, how does that weakness affect the impact of the program on a school’s overall dropout rate? This paper uses data from a recent evaluation of dropout prevention programs to address these questions.

The main result of the analysis is that risk factors commonly used by dropout prevention programs do not identify dropouts effectively. At both the middle and high school levels, students identified by common risk factors are more likely to remain in school than to drop out within two to three years of identification. Furthermore, a program designed to have enough slots to serve all future dropouts in a school would end up serving well under half the actual dropouts if it were to use

these risk factors to identify participants. Even programs with highly effective dropout prevention services are unlikely to reduce the dropout rate in a school, because the programs are unlikely to identify the students who are likely to drop out; that is, these programs would serve students who would remain in school regardless of the intervention.

B. PAST RESEARCH HAS IDENTIFIED RISK FACTORS

A well-established literature in education, sociology, and economics has examined factors associated with dropping out. The studies have examined the relationships between dropping out and all or a combination of five types of variables: (1) demographic characteristics and family background, (2) past school performance, (3) personal/psychological characteristics, (4) adult responsibilities, and (5) school or neighborhood characteristics. The studies agree on the factors that are related to dropping out, even though they use different data sources, cover different time periods, and differ in the extent to which they control for other factors before measuring these relationships.¹

- With respect to *demographic characteristics*, many researchers found that black and Hispanic youths are more likely than white youths to drop out of high school (for example, Natriello et al. 1990; and Eckstrom et al. 1986). This relationship disappears or is reversed after controlling for other factors (Clark 1992; Wojtkiewicz 1993; and Haveman 1991). Similarly, males are more likely than females to drop out until other factors are controlled (Wojtkiewicz 1993).
- Studies examining *family background* found that family income, socioeconomic status, and parents' educational attainment are related to dropping out (Rumberger 1983; National Center for Education Statistics 1990 and 1992; Mare 1980; and Barro and Kolstad 1987). In addition, studies have found that dropping out is more likely to occur among students with limited English proficiency (Natriello et al. 1990; and National Center for Education Statistics 1990), students whose families receive welfare (Haveman et al. 1991; and Mare 1980), and students with a sibling who has already dropped out (National Center for Education Statistics 1990) than among counterparts without these characteristics.

¹The relationships of a few characteristics with dropping out change after controlling for other factors. These cases are noted below.

- Researchers consistently found that *past school performance* is related to dropping out. Pallas (1987) states that “poor academic performance is the best predictor of who drops out of school.” Important indicators of poor academic performance that are associated with dropping out include low grades, poor test scores, and being on a nonacademic track (U.S. Department of Education 1983; National Center for Education Statistics 1992; Eckstrom et al. 1987; and Barro and Kolstad 1987). Other school performance measures related to dropping out include being overage for grade, having disciplinary problems in school, being truant, and spending little time on homework (National Center for Education Statistics 1992; Eckstrom et al. 1987; Barro and Kolstad 1987; and Pallas 1984).
- Studies have found that students with *adult responsibilities*, such as being employed or having to take care of a child, are more likely to drop out than are their counterparts without these responsibilities. Pallas (1987), Rumberger (1983), and Barro and Kolstad (1987) found that having a child is positively associated with a student’s likelihood of dropping out. Pallas (1984) and D’Amico (1984) identified a positive association between working more than 20 hours a week and dropping out.
- *Personal/psychological characteristics* related to dropping out include factors that reflect students’ commitment to schooling and ability to follow through on this commitment. Research by Eckstrom et al. (1987) and Rumberger (1983) showed that students with low self-esteem and an external locus of control (that is, a feeling that they have little control over their own destinies) are more likely to drop out. Rumberger (1983) found that students with low educational expectations or plans are more likely to drop out. Barro and Kolstad (1987) found that students’ religiosity is related to dropping out.
- Dropping out may be influenced not only by students’ (or their families’) characteristics, but also by *characteristics of their schools or neighborhoods*. Various researchers have shown that dropout rates are higher among students in urban schools than in rural or suburban schools (for example, Pallas 1987). Crane (1991), Clark (1992), and Vartanian and Gleason (forthcoming) show that students living in neighborhoods whose residents are poor are more likely to drop out than those in neighborhoods with wealthier residents, even after controlling for family income.

These studies provide clear evidence of risk factors associated with dropping out, but they do not address the key issue in this paper--how well risk factors predict school dropout. In particular, none of the studies attempted to assess how risk factors would fare if used to select participants in dropout prevention programs. In a school with a given number of students who would drop out in

the absence of program services, how many of these students would a particular risk factor identify?
How many students who would not drop out would the risk factor identify?

A few studies examined the predictive efficiency or validity of risk factors. For example, Lloyd (1978) looked at how well a set of student characteristics describing third graders predicts whether these students would drop out after entering high school. These results can be used to determine the extent to which a risk factor based on these characteristics would reach the future dropouts in a school. The study found that risk factors identify a group of students who are more likely than other third graders subsequently to drop out. However, the risk factors also identify a substantial number of students who do not ultimately drop out and fail to identify a substantial number of students who ultimately do so. Weber (1989) reviewed several studies of other dropout prediction scales and reached similar conclusions. In particular, these scales predict dropping out better than do usual informal procedures for identifying future dropouts but are not very accurate overall. For example, Weber (1989) suggests that an average dropout prediction scale identifies only about one-third of actual dropouts.

Finally, McKee et al. (1998) examined two dropout prediction scales, one based on information available from student records and a second based on information drawn from a student survey. The administrative records scale predicted dropping out well among a sample of 10th-grade students. However, the analysis of this scale was based on a small sample (N = 49) and predicted students' dropout status over a short period. In particular, the dropout prediction scale was measured in October and November of a school year, and students' dropout status was measured at the end of the school year (June). It is not clear whether the administrative records scale would predict dropping out as well over a longer period.

C. LONGITUDINAL DATA ALLOW RISK FACTORS TO BE ASSESSED

From the perspective of a dropout prevention program, an efficient risk factor identifies as many students as possible who are likely to drop out. If the program serves most of these students, referred to as *future dropouts*, the program will have the greatest effect on reducing the dropout rate in a school or larger community.²

The approach used here to measure efficiency is to select a group of students at a point in time, measure the students' risk factors, and follow up the students to determine whether they have dropped out of school.³ The efficiency of various risk factors is then assessed by calculating the dropout rate for students with different risk factors. The higher this dropout rate, the more efficient a risk factor. In particular, the higher the dropout rate for a given risk factor, the larger the number of future dropouts that a dropout prevention program using this risk factor will be able to serve (assuming it has a given number of program slots).⁴

²Hereafter, we will assume that the community served by a dropout prevention program is limited to a single school, although the underlying issues involved in measuring the efficiency of risk factors are the same if a program serves a larger group of students attending a cluster of schools or a school district.

³Because of a limited follow-up period to measure students' ultimate dropout status, the analysis in this paper treats students' dropout status at a point in time two to three years after their risk characteristics initially were measured. This "point-in-time" dropout status may differ from the students' ultimate dropout status, as students who drop out in one year may return to school in another, and students who remain in school in one year may drop out in the next. However, the point-in-time dropout status and ultimate dropout status are likely to be highly correlated, which is the assumption made in this analysis.

⁴One might argue that even if a risk factor does not identify future dropouts, it is likely to identify students who are struggling in school and who will benefit from the type of services provided by dropout prevention programs. The extent to which this argument is true will be discussed later. However, dropout prevention programs can better accomplish their ultimate goal of preventing dropout if they serve students who would drop out unless they were to participate in the program.

1. Longitudinal Data Are from Middle Schools and High Schools in Four Cities

Our data are from the evaluation of the School Dropout Demonstration Assistance Program (SDDAP), sponsored by the U.S. Department of Education. The SDDAP provided assistance to 85 local dropout prevention programs operating in sites throughout the United States, mostly between September 1991 and May 1995. The analysis in this paper makes use of extensive data from restructuring schools and comparison schools in four sites: (1) Dallas, Texas; (2) Grand Rapids, Michigan; (3) Phoenix, Arizona; and (4) Santa Ana, California.⁵ The sites are not nationally representative, but the data provide ample information about samples of students who are disadvantaged, diverse with respect to race/ethnicity, and fairly likely to drop out.⁶

As part of the evaluation of the SDDAP, we collected information on samples of middle and high school students at a subset of schools in these sites that were undergoing schoolwide reform, as well as from similar schools that were not undergoing reform. The middle school sample includes data from three of the four sites (all but Phoenix), where students were sampled and interviewed at the beginning of their seventh-grade year. The baseline interview asked questions about students' characteristics and attitudes and their experiences in school during the previous year (when most were sixth graders). The high school sample includes data from all four sites. In three of the sites, the baseline interview took place at the beginning of the students' 10th-grade year; in Phoenix, this

⁵These sites were chosen because they involved schoolwide restructuring initiatives rather than student services targeted to at-risk students. Thus, dropout prevention activities were designed to affect all students, not simply those whom the program defined as at risk in some respect. Thus, our analysis of risk factor efficiency could not be influenced or biased by SDDAP activities affecting outcomes only for students the programs defined as at risk. Our sample includes students from schools involved in the SDDAP restructuring initiative and those from comparison group schools not engaging in SDDAP restructuring activities.

⁶Gleason and Dynarski (1994) provide more information on the characteristics of students in these sites.

interview took place at the beginning of their 9th-grade year. All students in both the middle and high school samples were enrolled at the time of the baseline interview.

In all four sites, baseline data were collected from one cohort in fall 1992 and from another cohort in fall 1993. Students were interviewed again in spring 1995, nearly three years after baseline for cohort 1 and two years after baseline for cohort 2. Baseline and follow-up data were obtained for 2,672 middle school students from three sites and from 2,808 high school students in four sites.

2. Single Risk Factors

Risk factors are defined using data from the baseline survey. Dropout status is measured two to three years later, in what should be the end of students' 8th or 9th grade among middle school students and 11th or 12th grade among high school students (10th or 11th grade in Phoenix). Dropping out is defined as not being enrolled or on summer vacation and not having earned a high school degree or GED certificate at the end of this school year (as reported by students). After the risk factors and dropout status are defined, two measures can be calculated to assess the efficiency of a risk factor: (1) the proportion of all students with the risk factor, and (2) the proportion of students with the risk factor who are future dropouts.

The set of risk factors captures characteristics associated with dropping out derived from the research literature. Each risk factor can take on one of two values--1 if the student has the characteristic and 0 if the student does not. Some characteristics on which risk factors are based are continuous, so cutoffs were chosen to create a 0/1 factor.⁷ The set of risk factors includes:

⁷The choice of this cutoff point was somewhat arbitrary, but we conducted robustness analysis whenever possible to determine whether the selection of a different cutoff point would have changed the results.

Family Background

- Is in single-parent family
- Family receives public assistance
- English is not primary language
- Has sibling who dropped out of school
- Mother did not graduate from high school

Previous School Experiences

- Was absent 20 or more times during the school year
- Is overage for grade level by at least one year
- Is overage by two or more years
- Ever previously dropped out of school (high school sample only)
- Received low grades (Cs and Ds or below)
- Had disciplinary problems during school year
- Attended five or more schools during lifetime (six or more schools for the high school sample)
- Spends less than one hour per week on homework
- Spends no time each week reading for fun

Personal/Psychological Characteristics

- Has external locus of control
- Has low self-esteem
- Has parents who do not talk to student about things studied in school
- Is not “very sure” of graduating from high school
- Watches television more than five hours a night

Adult Responsibilities

- Has child (high school sample only)

3. Risk Factors Based on Multiple Characteristics

A program operator could combine characteristics and treat a student as being at risk if the student had a combination of characteristics. We address this possibility by constructing two risk factors based on a combination of characteristics: (1) a “composite risk factor,” and (2) a “propensity score risk factor.”

The composite risk factor indicates that a particular student is at risk if he or she has some minimum number of single risk factors. Here, the composite risk factor is based on eight single risk factors from the list.⁸ We analyzed alternative versions of the composite risk factor based on whether students had at least two, three, or four of eight single risk factors.

The propensity score risk factor is based on the same underlying concept as is the composite risk factor but uses a more complicated approach to combine single factors (see Appendix A for a more detailed description of the propensity score risk factor). We estimated a logistic regression model that relates 40 risk factors and other student characteristics to dropping out, and then we used the estimated model to calculate a “propensity score” for each student as his or her predicted probability of dropping out.⁹ Students with higher propensity scores are more likely to drop out than are those with lower scores, according to the model. We then defined students to be at risk--that is, to have the propensity score risk factor--if their propensity score exceeded a predetermined threshold. We set the threshold so that the proportion of at-risk students equaled the proportion of dropouts in the sample. Middle school students were coded as being at risk if their propensity score was above the 94th percentile of the middle school distribution (in other words, in the top 6 percent), and high

⁸The eight contributing risk factors were selected because they are the most predictive individual risk factors. For the middle school sample, they are (1) public assistance receipt, (2) high absenteeism, (3) being overage for grade, (4) low grades, (5) disciplinary problems, (6) spending no time reading, (7) having attended five or more schools, and (8) having parents who do not talk to the student about things studied in school. For the high school sample, the eight contributing risk factors are (1) high absenteeism, (2) being overage for grade, (3) low grades, (4) having a child, (5) having a sibling who dropped out, (6) having previously dropped out, (7) being unsure of graduating from high school, and (8) spending less than an hour a week on homework.

⁹For statistical background on propensity scores, see Rosenbaum and Rubin (1984). We estimated logistic regression models to generate the propensity score that both included and excluded a binary variable indicating whether a student was “treatment group” student in a restructuring school or a “control group” student in a comparison school. This variable did not significantly affect dropping out, nor did it affect our results.

school students were coded as being at risk if their score was above the 85th percentile of the high school distribution (the top 15 percent).

The idea behind the propensity score risk factor is to use as much information as possible about a student to determine whether he or she is at risk. In this sense, the propensity score risk factor is the “best” risk factor possible, in that it makes the most use of the data. The performance of the propensity score risk factor relative to the single risk factors provides a sense of whether using additional data and more sophisticated statistical techniques helps create a better risk indicator.

D. MOST RISK FACTORS ARE INEFFICIENT

None of the single risk factors is efficient in identifying future dropouts among middle school students. The dropout rate among students with a risk factor is low for each of the risk factors we examined. Table 1 shows the results for middle school students. In this sample of 2,568 students who were in the seventh grade, six percent had dropped out by the end of the eighth or ninth grade.¹⁰ The analysis described here examines the efficiency of risk factors for a dropout prevention program designed to serve six percent of the students in a school (for example, a program with 15 slots in a school that has 250 seventh graders).¹¹ The table lists (1) the single risk factors and the risk factors based on multiple characteristics; (2) the proportion of all students who are at risk, according to the

¹⁰Among students in both cohorts, the dropout rate after two years (by the end of the eighth grade) is 5.5 percent. Among students in cohort 1 with three years of follow-up data, the dropout rate at three years (by the end of the ninth grade) is 7.5 percent.

¹¹All but two of the individual risk factors we examine--high absenteeism and overage by two or more years--identify at least six percent of the sample as being at risk (and thus would fill all the program slots). We created another risk factor that defines students as at risk if they either have high absenteeism or are overage by two or more years. In addition, the composite risk factor requiring sample members to have four of eight individual risk factors also fails to identify at least six percent of the sample.

TABLE 1
MIDDLE SCHOOL RISK FACTOR EFFICIENCY

| Risk Factor | Percentage of All Students with Risk Factor | Dropout Rate Among Students with Risk Factor (Percentage) |
|---|---|---|
| Family Characteristics | | |
| Single Parent | 34 | 8 |
| Public Assistance | 24 | 8 |
| Mother Is High School Dropout | 24 | 5 |
| Sibling Has Dropped Out | 22 | 7 |
| English Not Primary Language | 18 | 5 |
| Previous School Experiences | | |
| Does Little Homework | 42 | 6 |
| Overage for Grade | 30 | 9 |
| Disciplinary Problems | 30 | 9 |
| Does Not Read for Fun | 18 | 9 |
| Has Attended 5 or More Schools | 17 | 10 |
| Low Grades | 12 | 8 |
| High Absenteeism | 6 | 15 |
| Overage by 2 or More Years | 5 | 16 |
| Personal/Psychological Characteristics | | |
| External Locus of Control | 46 | 7 |
| Not Sure of High School Graduation | 37 | 8 |
| Low Self-Esteem | 33 | 7 |
| Watches Much Television | 30 | 7 |
| Parents Do Not Talk to Student About School | 17 | 11 |
| Multiple Characteristics | | |
| Composite Risk Factor (3 of 8) | 15 | 14 |
| High Absenteeism or Overage by 2 or More Years | 11 | 15 |
| Composite Risk Factor (4 of 8) | 6 | 19 |
| Propensity Score Risk Factor | 6 | 23 |

SOURCE: SDDAP baseline and follow-up survey data.

NOTE: The sample size is 2,568. Among this sample, six percent dropped out within two to three years.

factor; and (3) the dropout rate among these students. A risk factor is efficient if students who have the factor also have a high dropout rate.

The factors associated with the highest dropout rates are high absenteeism (students with that factor have a dropout rate of 15 percent) and overage by two or more years (students with that factor have a dropout rate of 16 percent). For students with the combined risk factor of having high absenteeism or being overage by two or more years, the dropout rate is 15 percent. These two risk factors are more efficient than other risk factors, which all have dropout rates below 11 percent among at-risk students.

Despite the fact that high absenteeism and being overage by two or more years are more efficient than other single risk factors, dropout prevention programs using these risk factors would not reach most of their target populations. Indeed, only 15 percent of students with these two factors drop out. Programs using these risk factors to identify participants would end up providing dropout prevention services primarily to students who never would have dropped out.

Composite risk factors do only a little better than the most efficient single risk factors (high absenteeism and overage by two or more years). The dropout rate for students with the composite factor defined by whether students have four of eight single factors is 18 percent, only slightly higher than the dropout rates for students with the high absenteeism and overage risk factors (Table 1). More than four out of five students identified using this composite risk factor would not be future dropouts.

The propensity score risk factor predicts dropping out most efficiently. The dropout rate among students defined as at risk by the propensity score risk factor is 23 percent, substantially higher than any other risk factor and almost four times as high as the dropout rate for all students. However, even the propensity score risk factor does not identify most dropouts. More than three-fourths of

students identified as at risk of dropping out according to the propensity score risk factor do not drop out within two to three years.

Thus, even though the propensity score risk factor predicts dropping out among middle school students better than do any of the individual risk factors, it would be an inefficient risk factor for three reasons. First, most of the students the factor identifies for the program do not need the program (they would stay in school even if the program did not serve them).¹² Second, most of the students in the school who need services would be passed over by the program because they would not have the risk factor. Third, even if the program's services were effective for participants, the program would have only a small impact on the dropout problem in the school.

A numerical example helps to underscore the program effects of using an inefficient risk factor. Suppose that a middle school has 1,000 students beginning seventh grade. On the basis of past experience, the school expects 60 of these students to drop out by the end of what would be their eighth- or ninth-grade year, so it establishes a dropout prevention program to serve 60 students. If the program uses the propensity score risk factor to identify 60 at-risk participants, the program will end up serving only 14 students who would have dropped out (about 23 percent of 60). The other 46 students served by the program would not need the services, as they will not drop out. The flip side of this example is that the propensity score risk factor will incorrectly code 46 future dropouts in the school as not at risk, so these students will not be served by the program. The program will

¹²A somewhat larger proportion of students who have this risk factor would need the services if we consider a broader definition of educational failure. In particular, if we consider an individual to have failed if he or she drops out, is absent daily or almost daily, skips school more than once a week, receives mainly Ds and Fs in school, or has been arrested, slightly less than half the students with the propensity score risk factor will have failed (overall, 22 percent of the middle school sample falls into this category). Thus, even with this broader definition of educational failure, more than half the students who are served by the dropout prevention program do not “need” the services.

reach only 23 percent of the school's seventh graders who need the services, the students who are future dropouts.

The implication of using inefficient risk factors is that even if the program provides highly effective dropout prevention services, it will not reduce the school dropout rate by much. For example, if the program described in the example prevented half the dropouts it served from dropping out, it would reduce the number of dropouts in the school from 60 to 53 (from 6.0 percent to 5.3 percent), an overall impact of only 12 percent. Although the program serves 60 students, it prevents only 7 students who would have dropped out in the absence of services from doing so. The cost of using inefficient risk factors is that the program's maximum feasible impact of 50 percent is reduced to only 12 percent.

The high school analysis yields results analogous to those of the middle school analysis (Table 2). The high school sample consists of 2,615 students, most beginning the 10th grade. Within two to three years, 14.6 percent had dropped out.¹³ The analysis here examines the effectiveness of risk factors for a dropout prevention program designed to serve 14.6 percent of the students in a school (for example, a program with 73 slots in a school that has 500 10th graders). This dropout prevention program is larger than the program in the middle school analysis because the scale of the dropout problem is larger in high school.

The high school risk factors include having ever dropped out or having a child, as well as the factors used for middle school students. Definitions of some risk factors also differ for the high school sample--a student has to have attended six or more schools (rather than five or more) to be considered at risk, and the two composite variables require individuals to have two out of eight and

¹³Among students in both cohorts, the dropout rate after two years for the high school sample (by the end of the 11th grade) is 13.1 percent. Among students in cohort 1 with three years of follow-up data, the dropout rate after three years (by the end of the 12th grade) is 17.0 percent.

TABLE 2
HIGH SCHOOL RISK FACTOR EFFICIENCY

| Risk Factor | Percentage of Students with Risk Factor | Dropout Rate Amount Students with Risk Factor (Percentage) |
|---|--|--|
| Family Characteristics | | |
| Single Parent | 35 | 18 |
| Public Assistance | 15 | 19 |
| Mother Is High School Dropout | 26 | 18 |
| Sibling Has Dropped Out | 22 | 21 |
| English Not Primary Language | 19 | 17 |
| Previous School Experiences | | |
| High Absenteeism | 9 | 27 |
| Overage for Grade | 31 | 21 |
| Overage by 2 or More Years | 9 | 28 |
| Low Grades | 14 | 27 |
| Disciplinary Problems | 19 | 23 |
| Ever Dropped out | 7 | 13 |
| Does Little Homework | 31 | 21 |
| Does Not Read for Fun | 16 | 17 |
| Has Attended 6 or More Schools | 18 | 17 |
| Personal/Psychological Characteristics | | |
| External Locus of Control | 38 | 18 |
| Low Self-Esteem | 23 | 19 |
| Not Sure of High School Graduation | 23 | 25 |
| Parents Do Not Talk to Student About School | 22 | 20 |
| Watches Much Television | 15 | 17 |
| Adult Responsibilities | | |
| Has Child | 3 | 32 |
| Multiple Characteristics | | |
| High Absenteeism or Overage by 2 or More Years | 17 | 27 |
| Low Grades or Has Child | 18 | 28 |
| High Absenteeism or Has Child or Ever Dropped Out | 17 | 27 |
| Composite Risk Factor (2 of 8) | 30 | 25 |
| Composite Risk Factor (3 of 8) | 14 | 34 |
| Propensity Score Risk Factor | 15 | 42 |

SOURCE: SDDAP baseline and follow-up survey data.

NOTE: The sample size is 2,615. Among this sample, 14.6 percent dropped out within two to three years.

three out of eight single risk factors to be considered at risk (rather than three and four out of eight). Several of the single risk factors do not identify a sufficient number of students to fill the slots of a program serving 14.6 percent of the school population. For these risk factors, new risk factors were constructed that defined students as being at risk if they have either of two or three single risk factors.¹⁴ This type of combination risk factor identifies a larger number of students while preserving the general character of the single risk factors.

As in the middle school analysis, none of the single risk factors efficiently identifies dropouts. Previous school experiences generally do best among single risk factors. For example, 31 percent of high school students were overage for grade, and 21 percent of those who were overage dropped out. However, even using the best single risk factors, students classified as being at risk are much more likely to remain in school during the next two to three years than to drop out.

As in the middle school analysis, a dropout prevention program could more efficiently identify dropouts by using risk factors based on multiple characteristics, rather than single risk factors. The composite risk factor based on whether students have at least three of eight specific risk factors was associated with a dropout rate of 34 percent among students it identified, higher than the dropout rate of 28 percent for the top single risk factor (Table 2). The propensity score risk factor performs best, with a dropout rate among students it identifies of 42 percent.

Nonetheless, the propensity score risk factor is wrong more often than it is right. In particular, 58 percent of the students identified as likely to drop out by the propensity score risk factor do not

¹⁴In particular, the risk factors for high absenteeism, being overage by two or more years, having low grades, and having a child identify fewer than 14.6 percent of students. Three constructed risk factors were created to increase the number of students identified: (1) either high absenteeism or being overage, (2) either having low grades or having a child, and (3) either high absenteeism, having a child, or having previously dropped out.

drop out.¹⁵ Even a propensity score risk factor based on 40 student characteristics and measures of past school performance cannot identify at-risk students in a way that allows a dropout prevention program to serve a large proportion of the students who really need the services.

A numerical example illustrates the implications of using an inefficient risk factor to identify high school dropouts. Suppose that a high school expects 146 of its 1,000 entering 10th graders to drop out and establishes a dropout prevention program to serve 146 students. If the school uses the propensity score risk factor to identify program participants, about 61 of the participants would drop out in the future. Thus, the program would serve more students who do not need dropout prevention services than students who do need them. Of the 146 future dropouts in the school, 85 will not be coded as at risk and will not be served by the program. Ultimately, even if the program provided services preventing 31 of the 61 future dropouts it served from dropping out, it would reduce the overall school dropout rate from 14.6 percent to only 11.5 percent.

E. CAN OTHER FACTORS PREDICT DROPPING OUT BETTER?

The literature on the dropout problem describes a number of student characteristics and measures of past performance in school related to the chances of dropping out of school. These results suggest that a dropout prevention program should be able to identify students who will drop out.

The analysis in this paper suggests that the task of predicting who drops out is far more difficult than simply taking into account known risk factors. A number of student characteristics and measures of past school performance are related to dropping out, but the relationships are not strong

¹⁵Using the broader definition of educational failure described previously, 33 percent of the high school sample and 65 percent of those with the propensity score risk factor have failed. Thus, more than one-third of the students served by a dropout prevention program would not need these services, according to this measure. Among students in a school classified in this way as having failed educationally, the program would serve only 23 percent.

enough to allow a dropout prevention program to identify efficiently students who will drop out. Even for the most efficient risk factor, only a relatively small proportion of the students served by a program will become dropouts. Many of those served do not need the services. Many of those in need will not be served. And no matter how effective the program's dropout prevention services for participants, it will have a small impact on the school's dropout rate.

One could argue that it is not surprising that characteristics measured when students are beginning the seventh grade do not predict accurately whether they will drop out. After all, measures of school performance at this grade level may be unreliable, or the factors leading to the dropout decision may become evident only later. However, the fact that neither risk factors measured in 7th grade nor those measured in 10th grade reliably predict dropping out suggests that the problem is not simply the point when risk factors are measured.

The risk factors analyzed in this paper do not predict dropping out precisely. What characteristics or factors not examined here could more efficiently identify students who will drop out? Might a dropout prevention program operator use other risk factors to identify participants? Four possible sets of factors include (1) ecological characteristics, (2) unobserved psychological factors, (3) measures of the persistence of specific characteristics over time, and (4) transitory events that occur after the point at which baseline risk factors are measured.

Ecological factors that could influence dropping out but that were not measured here include neighborhood characteristics and peer group characteristics. Students growing up in very poor neighborhoods and surrounded by many dropouts may be more likely to drop out than students living in areas where staying in school is the norm. Being able to measure neighborhood or peer group characteristics carefully may improve the ability of a propensity score risk factor to predict dropping out.

Unobserved psychological factors may be important determinants of whether individuals drop out. This paper examines the performance of risk factors based on students' self-esteem, locus of control, and educational expectations and finds that they do not predict dropping out. These risk factors may be measuring the underlying psychological constructs imprecisely, or other psychological factors may be more important in predicting dropping out. For example, the degree to which an individual perseveres through difficult times or is optimistic about the future may better predict his or her future success in school. These constructs are difficult to measure through a survey, but teachers or counselors who work closely with students may be able to assess the degree to which particular students have these qualities.

The two other explanations involve viewing risk in a more dynamic way. The risk factors used here measure student characteristics at a single point in time and performance in school in a single academic year. Characteristics or factors that lead to dropping out may have a more cumulative effect. Performing poorly in school one year may lead to temporary disappointment and an increased resolve to do better; performing poorly for several years in a row may lead students to become detached from school and to believe that they are failures, eventually leading them to drop out. Risk factors that measure student performance over several school years may be better predictors of dropping out than those that measure performance over a single year.

Some students may drop out not because of the cumulative effects of poor academic performance but because of an unexpected event that severely affects them. A student may have had a child, been arrested, started using drugs, or had serious personal problems at home. Despite previous success in school, the event may lead the student to decide abruptly to leave school. Again, the likelihood of these events occurring is difficult to quantify using a survey. However, teachers

or counselors who know the students may be aware of such events and get students into dropout prevention programs to mitigate any negative effects.

A program operator working within a single school or community and with substantial input from teachers and counselors may be able to develop admission criteria (or risk factors) that predict dropping out more accurately than do the risk factors examined here. Two caveats limit the likelihood of such a scenario, however. First, the unobserved factors described above would be difficult to quantify. For example, although teachers may have some sense of students' degree of perseverance through hard times, it would be difficult to determine the point at which the lack of this quality puts a student at risk of dropping out. Moreover, different teachers may have different standards for assessing perseverance. A program operator could not easily define objective standards for admitting students to a dropout prevention program, and using subjective admission criteria might be viewed as unfair. Second, although it seems plausible that these unmeasured risk factors would predict dropping out accurately, we have no evidence that this is true. As was the case with the wide range of risk factors measured in this paper, these additional risk factors may slightly improve our ability to predict dropping out but still result in dropout prevention programs that serve many students who do not need dropout prevention services and fail to serve many students who do need them.

The results here do not suggest that common risk factors should no longer be used to identify dropouts. Risk factors do better than using no information at all and allow programs to codify who will be served. The problem is that risk factors do not do well in identifying dropouts. It will be difficult for programs to make much headway in reducing the dropout rate until more is understood about why students drop out.

REFERENCES

- Barro, Stephen M., and Andrew Kolstad. "Who Drops Out of High School? Findings from High School and Beyond." Report no. CS 87-397c. Washington, DC: U.S. Department of Education, May 1987.
- Clark, Rebecca L. "Neighborhood Effects on Dropping Out of School Among Teenage Boys." Population Studies Center Discussion Paper Series, UI-DSC-UI-13. Washington, DC: The Urban Institute, December 1992.
- Crane, Jonathan. "The Epidemic Theory of Ghettos and Neighborhood Effects on Dropping Out and Teenage Childbearing." *American Journal of Sociology*, vol. 96, no. 5, March 1991, pp. 1226-1259.
- D'Amico, R. "Does Employment During High School Impair Academic Progress?" *Sociology of Education*, vol. 57, 1984, pp. 152-164.
- Eckstrom, Ruth B., et al. "Who Drops Out of High School and Why? Findings from a National Study." In *School Dropouts: Patterns and Policies*, edited by Gary Natriello. New York: Teachers College Press, 1986.
- Gleason, Philip M., and Mark Dynarski. "Falling Behind: Characteristics of Students in Federally Funded Dropout Prevention Programs." Report submitted to the U.S. Department of Education. Princeton, NJ: Mathematica Policy Research, Inc., 1994.
- Haveman, Robert, Barbara Wolfe, and James Spaulding. "Childhood Events and Circumstances Influencing High School Completion." *Demography*, vol. 28, February 1991, pp. 133-157.
- Lloyd, Dee Norman. "Prediction of School Failure from Third-Grade Data." *Educational and Psychological Measurement*, vol. 38, 1978, pp. 1193-1200.
- Maddala, G.S. *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press, 1983.
- Mare, Robert D. "Social Background and School Continuation Decisions." *Journal of the American Statistical Association*, vol. 75, June 1980, pp. 295-305.
- McKee, John M., Kenneth B. Melvin, Victor Ditoro, and Susan P. McKee. "SARIS: Student At-Risk Identification Scale." *Journal of At-Risk Issues*, vol. 4, no. 2, winter/spring 1998, pp. 24-33.
- National Center for Education Statistics. *Characteristics of At-Risk Students in NELS:88*. Washington, DC: U.S. Department of Education, August 1992.

- National Center for Education Statistics. *A Profile of the American Eighth Grader*. Washington, DC: U.S. Department of Education, June 1990.
- Natriello, Gary, Edward L. McDill, and Aaron M. Pallas. *Schooling Disadvantaged Children: Racing Against Catastrophe*. New York: Teachers College Press, 1990.
- Pallas, Aaron M. "School Dropouts in the United States." Washington, DC: U.S. Department of Education, 1987.
- Pallas, Aaron M. "The Determinants of High School Dropout." Ph.D. dissertation, Department of Sociology, The Johns Hopkins University, 1984.
- Rosenbaum, Paul, and Donald Rubin. "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score." *Journal of the American Statistical Association*, vol. 79, 1984, pp. 516-524.
- Rumberger, Russell W. "High School Dropouts: A Review of Issues and Evidence." *Review of Educational Research*, vol. 20, summer 1983, pp. 101-121.
- U.S. Department of Education. *High School Dropouts: Descriptive Information from High School and Beyond*. Publication no. NCES 83-221b. Washington, DC: National Center for Education Statistics, 1983.
- Vartanian, Thomas P., and Philip M. Gleason. "Do Neighborhood Conditions Affect High School Dropout and College Graduation Rates?" Forthcoming in *Journal of Socioeconomics*.
- Weber, J.M. *Identifying Potential Dropouts: A Compilation and Evaluation of Selected Procedures*. Columbus, OH: Ohio State University, 1989.

APPENDIX A

CREATING THE PROPENSITY SCORE RISK FACTOR

The propensity score risk factor is based on the estimation results from a logistic regression model of students' dropout status. This model can be represented mathematically as:

$$D_i^* = X_i\beta + e_i$$

$$D_i = 1 \quad \text{if } D_i^* \geq 0$$

$$= 0 \quad \text{if } D_i^* < 0,$$

where:

D_i^* = student i 's propensity to drop out

D_i = student i 's ultimate dropout status

X_i = vector of student i 's individual and family characteristics.

In the model, we observe D_i , a binary variable indicating whether or not student i drops out. However, we do not observe the student's actual dropout propensity, D_i^* . Thus, to estimate the parameters (β) of the model, we assume that the random error term (e_i) has a logistic distribution, which allows us to estimate a logit model using maximum likelihood estimation techniques (Madalla 1983). Table A.1 presents coefficient estimates from this model for the SDDAP middle school and high school samples.

The ultimate goal of this analysis is to estimate the value of D_i^* , student i 's propensity to drop out. This estimate is referred to as the student's dropout propensity score. We cannot directly observe or calculate D_i^* because e_i is, by definition, unobserved, and we can only estimate the value of β . However, because we can estimate β and because we know that the expected value of e_i is zero, we can estimate D_i^* . In particular:

$$\hat{D}_i^* = X_i\hat{\beta}.$$

TABLE A.1
DROPOUT MODEL RESULTS

| Independent Variable | Middle School | | High School | |
|---|---------------|----------------|-------------|----------------|
| | Coefficient | Standard Error | Coefficient | Standard Error |
| Race/Ethnicity (White Excluded) | | | | |
| Hispanic | -0.33 | 0.32 | -0.12 | 0.23 |
| Black | -0.32 | 0.28 | -0.30 | 0.20 |
| Other | -0.14 | 0.41 | -0.30 | 0.32 |
| Gender (Female Excluded) | | | | |
| Male | -0.08 | 0.20 | 0.03 | 0.14 |
| Test Score (10 NCE Points) | | | | |
| Reading | -0.04 | 0.07 | 0.12 ** | 0.06 |
| Math | 0.10 | 0.06 | -0.04 | 0.06 |
| Number of Absences (0 to 5 Days Excluded) | | | | |
| 5 to 10 days | 0.51 ** | 0.23 | 0.42 ** | 0.17 |
| 11 to 15 days | 0.93 ** | 0.30 | 0.42 ** | 0.21 |
| 16 to 20 days | 0.84 ** | 0.39 | 0.56 ** | 0.24 |
| 21 to 30 days | 0.85 ** | 0.43 | 0.26 | 0.28 |
| More than 30 days | 1.12 ** | 0.41 | 0.44 | 0.30 |
| Age | 0.24 * | 0.14 | 0.19 ** | 0.08 |
| Single-Parent Family | 0.21 | 0.19 | 0.31 ** | 0.14 |
| Receives Public Assistance | 0.24 | 0.23 | 0.03 | 0.19 |
| Nonnative English Speaker | 0.07 | 0.29 | 0.19 | 0.20 |
| Has Own Child | -- | -- | 0.55 | 0.34 |
| Ever Dropped Out | -- | -- | 0.29 | 0.23 |
| Locus of Control (Average Excluded) | | | | |
| External | 0.13 | 0.22 | -0.17 | 0.16 |
| Internal | -0.00 | 0.27 | -0.31 * | 0.17 |
| Parental Disciplinary Index | -0.33 | 0.27 | -0.55 ** | 0.22 |
| Educational Expectations (Less than High School Excluded) | | | | |
| High school only | -0.36 | 0.34 | -0.21 | 0.23 |
| College degree | -0.07 | 0.31 | -0.46 * | 0.24 |
| Graduate school | -0.20 | 0.31 | -0.05 | 0.24 |
| Number of Household Possessions | | | | |
| Education-related | -0.10 | 0.12 | -0.15 | 0.08 |
| Total | 0.01 | 0.04 | 0.00 | 0.03 |
| Grades (As and Bs Excluded) | | | | |
| Bs and Cs | 0.04 | 0.22 | 0.15 | 0.18 |
| Cs and Ds | 0.22 | 0.29 | 0.77 ** | 0.21 |
| Ds and Fs | -0.16 | 0.45 | 0.96 ** | 0.31 |
| Time Spent on Homework per Week (None Excluded) | | | | |
| Less than 1 hour | -0.39 | 0.36 | 0.03 | 0.29 |
| 1 to 3 hours | -0.16 | 0.38 | -0.35 | 0.30 |
| 4 to 5 hours | 0.03 | 0.42 | -0.23 | 0.33 |
| More than 5 hours | 0.14 | 0.46 | -0.81 ** | 0.37 |

TABLE A.1 (continued)

| Independent Variable | Middle School | | High School | |
|--|---------------|----------------|-------------|----------------|
| | Coefficient | Standard Error | Coefficient | Standard Error |
| Self-Esteem (Medium Excluded) | | | | |
| Low | -0.13 | 0.22 | 0.04 | 0.17 |
| High | -0.17 | 0.24 | -0.14 | 0.16 |
| Sure of High School Graduation | -0.18 | 0.20 | -0.19 | 0.15 |
| Mother's Education (High School Only Excluded) | | | | |
| High school dropout | -0.05 | 0.29 | 0.20 | 0.20 |
| Attended college | 0.04 | 0.29 | -0.12 | 0.19 |
| Number of Siblings | 0.08 | 0.05 | 0.13 ** | 0.04 |
| At Least One Parent Works | 0.09 | 0.22 | 0.07 | 0.18 |
| School Climate Index | -0.16 | 0.18 | -0.09 | 0.12 |
| Has Sibling Who Has Dropped Out | -0.10 | 0.22 | 0.12 | 0.15 |
| Number of Hours Watched TV per Day (None Excluded) | | | | |
| 1 to 3 hours | -0.38 | 0.33 | 0.03 | 0.20 |
| 3 to 4 hours | -0.07 | 0.37 | -0.22 | 0.24 |
| More than 4 hours | -0.18 | 0.32 | 0.04 | 0.22 |
| Number of Schools Attended Since First Grade | 0.12 ** | 0.05 | -0.02 | 0.03 |
| How Frequently Student Skips School (1 to 4 Scale) | -0.00 | 0.14 | -0.02 | 0.08 |
| How Frequently Student Is Late for School (1 to 4 Scale) | -0.01 | 0.10 | -0.07 | 0.07 |
| Number of Athletic Activities | -0.01 | 0.10 | 0.09 | 0.07 |
| Number of Extracurricular Activities | -0.06 | 0.09 | 0.02 | 0.07 |
| How Often Student Talks with Parents About School (1 to 3 Scale) | -0.28 ** | 0.13 | -0.10 | 0.09 |

SOURCE: SDDAP Baseline and Follow-up Survey and School Records Data.

NOTE: All independent variables measured at baseline. The dependent variable is the student's dropout status two to three years after the baseline year. Maximum likelihood techniques were used to estimate this logit model. The model also included the following independent variables: 10 dummy variables representing students' disciplinary problems, 4 dummy variables representing students' parents' involvement in school activities, the amount of time students spend reading for fun per week, the number of different in-school activities, students' attitudes about the importance of attending college, a binary variable indicating students' perception of their relationship with their parents, binary variables indicating the SDDAP demonstration site, and binary variables indicating missing values for several variables (gender, number of absences, public assistance, mother's educational attainment, parents' school involvement, and whether sibling has dropped out).

NCE = Normal Curve Equivalent

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

In other words, each student's propensity score is a weighted sum of their characteristics, X_i , where the weights are the coefficient estimates from the logit model.

The interpretation of the propensity score is straightforward. The higher the value of the propensity score, the more likely the student is to drop out. To define the propensity score risk factor, we simply define all students whose propensity score exceeds some threshold as at risk, and all students whose propensity score is below the threshold as not at risk. The value of the threshold can vary according to how many students we wish to define as at risk.

In this paper, we set out to define as at risk a number of students equal to the number of dropouts in the sample. Because about six percent of the middle school sample dropped out, we set the propensity score risk factor to 1 for the six percent of the sample with the highest propensity scores (in other words, those at or above the 94th percentile of the propensity score distribution). The propensity score risk factor was set to 0 for those whose propensity score was below this threshold. The high school propensity score risk factor was created analogously, so as to define 14 percent of the sample as at risk.