District Guide for Creating Early Warning System Indicators

Prepared by AIR for REL West at WestEd

Yibing Li
Jenny Scala
Dean Gerdeman
David Blumenthal

About REL West

The Regional Educational Laboratory West (REL West) at WestEd, serving Arizona, California, Nevada, and Utah, is part of a national network of 10 RELs whose mission is to provide research, analytic support, and resources that increase the use of high-quality data and evidence in education decision-making. Most REL work is carried out in partnership with educators — from state and local decision-makers to district and school support providers and practitioners — through eight regional research alliances. For more information, visit http://relwest.WestEd.org.

About WestEd

WestEd is a nonpartisan, nonprofit research, development, and service agency that partners with education and other communities throughout the United States and abroad to promote excellence, achieve equity, and improve learning for children, youth, and adults. WestEd has more than a dozen offices nationwide, from Massachusetts, Vermont and Georgia, to Illinois, Arizona and California, with headquarters in San Francisco. For more information, visit http://www.WestEd.org; call 415.565.3000 or, toll-free, (877) 4-WestEd; or write: WestEd / 730 Harrison Street / San Francisco, CA 94107-1242.

This resource was prepared for the Institute of Education Sciences (IES) under Contract ED-IES-12-C-0002 by Regional Educational Laboratory (REL) West at WestEd. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Appendix A. Early Warning Indicator Documentation Template (from section 1)

Drawing on information shared in section 1, use this template (table A1) to organize the early warning indicators used in the research into "ABC" categories and identify the variables your district is interested in developing into early warning indicators. District staff can group the indicators into broad categories, including attendance, behavior, course performance, and composite/other. When documenting the early warning indicators, please record the following:

- **Grade range:** An individual grade or range of grades (for example, grades 9–12). Indicators may have different cut points at different grade ranges. In this case, list indicators with unique grade range and cut point combinations on separate rows.
- **Time frame:** The time(s) during the school year at which an existing indicator was applied may differ by indicator. Some indicators might have been collected every grading period, whereas others may only be available at the end of a year. In many instances, the indicators may become available on multiple occasions.
- **Cut point:** The threshold, based on research, at which students are flagged in early warning indicators for being at risk. Be sure to indicate the polarity of the cut point (above or below).

Note: An editable version of this table is available in District Guide for Selecting Early Warning Indicators to Identify Students At Risk of Not Achieving Desired Graduation Outcomes: Editable Planning Forms.

Table A1. Template for Documenting Indicators Used in the Research

Indicator	Grade Range	Time Frame	Cut Point	Interested in Exploring?
Attendance				
Behavior				
Course Performance				
Composite Indicators/Others				

Appendix B. Map of Data Availability
(from section 2)

Use this template (table B1) to identify the specific data elements comprising the indicators and variables that will used to create the early warning indicators. For each data element, note the number of years the data have been collected and specific location of the files. Use separate rows as necessary to list the data elements that are located in different files or locations. *Note: An editable version of this table is available in* District Guide for Selecting Early Warning Indicators to Identify Students At Risk of Not Achieving Desired Graduation Outcomes: Editable Planning Forms.

Table B1. Template: Data Elements, Data Collection Schedules, and Locations

Indicator or Variable	Data Elements	Frequency of Collection	Location of Files
Demographics			
Attendance			
Behavior			
Course Performance			
Composite Indicators/Other			

Note. Demographic variables will not be used to create indicators. We recommend gathering them so they can inspect patterns in outcome and predictors between student subgroups based on demographic characteristics (such as gender and racial group).

Appendix C. Data Quality Checklist
(from Section 2)

Step 2.3 of the guide describes the process of conducting data quality checks. Once the requisite data to be analyzed have been merged into a single dataset, conduct a few quality checks to ensure accuracy and availability. Quality checks will identify missing data for students, for particular grading periods and school years, and for all early warning indicators and outcome variables. Select the checkbox next to each quality check when the process has been completed.

- ☐ All student data being analyzed for all selected years and grading periods are available (based on appendix B: Map of Data Availability).
- ☐ All student records have a matching Student ID.
- ☐ All variables use a consistent format (for example, all dates use the MM/DD/YYYY format).
- ☐ All student records include all variables being analyzed; the student records do not include gaps in any variable being analyzed.
- ☐ Student records are continuously present across grading periods and school years.

Appendix D. Documenting the Distribution of the Outcome Variable
(from section 3)

Use this template (table D1) to record the distribution of the outcome variable. Replace the column headings with any outcome variable, such as four- or five-year graduation rates, that your district has selected based on the guidance from section 3. *Note: An editable version of each of the following tables is available in* District Guide for Selecting Early Warning Indicators to Identify Students At Risk of Not Achieving Desired Graduation Outcomes: Editable Planning Forms.

Table D1. Distribution of Outcome Variable: High School Graduation and Dropout Rates

Cohort	Graduation Rate	Four-Year Dropout Rate	Five-Year Dropout Rate
Example: Class of 2015	*97.5%*	*3.3%*	*2.7%*
Class of 20____			
Class of 20____			
Class of 20____			
Class of 20____			
Average			

Use the frequency table (table D2) to display the distribution of predictor variables (step 3.2) and create a line or bar chart with the percentage of students grouped by the predictor variable ranges you have selected (step 3.3)

Table D2. Frequency Table of Students by Predictor Variable

Predictor Variable Range	Frequency	Percentage of Students
Example: 0–4 days absent	*375*	*47.8%*

Appendix E. Graduation Outcomes Tracker (from section 4)

Use table E1 to document the relationship between the selected binary predictors and graduation outcomes. Replicate the table for each indicator you will explore for a relationship to graduation or dropout. See step 4.1 for additional guidance. *Note: An editable version of this table is available in* District Guide for Selecting Early Warning Indicators to Identify Students At Risk of Not Achieving Desired Graduation Outcomes: Editable Planning Forms.

Table E1. Graduation Outcomes Tracker

Indicator: Attendance

Binary Predictor	Graduated On Time Number	Graduated On Time Percentage	Did Not Graduate Number	Did Not Graduate Percentage	Total
		%		%	
		%		%	
Total		%		%	

Indicator: Behavior

Binary Predictor	Graduated On Time Number	Graduated On Time Percentage	Did Not Graduate Number	Did Not Graduate Percentage	Total
		%		%	
		%		%	
Total		%		%	

Indicator: Course Performance

Binary Predictor	Graduated On Time Number	Graduated On Time Percentage	Did Not Graduate Number	Did Not Graduate Percentage	Total
		%		%	
		%		%	
Total		%		%	

Indicator: Composite/Other

Binary Predictor	Graduated On Time Number	Graduated On Time Percentage	Did Not Graduate Number	Did Not Graduate Percentage	Total
		%		%	
		%		%	
Total		%		%	

Appendix F. Template for Calculating Performance Measures of Indicators and Identifying Optimal Cut Points Through Receiver Operating Characteristic Analysis (from section 5)

This appendix includes two tools for districts to compare indicators. Table F1 is the template for calculating performance measures of indicators created based on the different cut points. The cut points could be actual values from district variables or selected values that are at certain percentiles. An introduction to Receiver Operating Characteristic (ROC) analysis that districts with statistical capacity can apply in identifying optimal cut points follows table F1. *Note: An editable version of this table is available in* District Guide for Selecting Early Warning Indicators to Identify Students At Risk of Not Achieving Desired Graduation Outcomes: Editable Planning Forms.

Table F1. Template for Calculating Performance Measures of Indicators Created Based on Different Cut Points

Potential Cut Points for Number of Course Failures	True Positive	False Negative	True Negative	False Positive	Sensitivity	Specificity	Youden Index
Cut point 1	TP	FN	TN	FP	TP/(TP + FP)	TN/(TN + FN)	Sensitivity + 1— Specificity
Cut point 2							

Identifying Optimal Cut Points Using ROC Curve Analysis

Districts with statistical capacity can identify optimal cut points for continuous variables through ROC curve analysis. ROC analysis can be performed by conducting logistic regression usingstatistical software packages such as SPSS, SAS, and Stata. Logistic regression reports the likelihood of achieving a designated outcome in a log-odds format.

The ROC curve analysis determines a cut point for a continuous variable that optimizes the classification quality of that variable on a certain outcome of interest (Eng, 2005). The optimal cut points, once identified, can be used to create binary indicators that would classify students into graduation "likely" and "not likely" categories. Cross-tabulations between the predicted likelihood and a particular outcome (for example, dropout) would generate four categories: those who were classified as at risk of dropping out and actually dropped out (true positive), those who were classified as not at risk and actually did not drop out (true negative), those who were classified as at risk but did not drop out (false positive), and those who were classified as not at risk but dropped out (false negative). To determine the optimal cut points, you need to calculate true positive and true negative rates. The true positive rate is the proportion of students who dropped out of high school who were successfully identified as at risk of dropping out. The true negative rate is the proportion of graduates correctly identified as not at risk. Ideally, you would want to identify the optimal cut point that has a true positive and true negative as close to 1 as possible.

The ROC curve analysis is driven by a set of true positive and true negative rates for all the possible cut points that exist in the data. The choice of cut points affects cell counts in the true positive, true negative, false positive, and false negative groups, which in turn affects true positive and true negative rates. All pairs of true positive and (1 – true negative) are plotted in a space defined by an *X* axis and a *Y* axis. The upper-left corner is the perfect point that signifies student likelihood of dropping out. The goal is to find a cut point that maximizes the power of the predictor variable in predicting the dichotomous outcome of dropping out (or other graduation outcomes). When all possible pairs of the true positive rate and (1 – true negative rate) are plotted on an *X*-*Y* graph, a curved line emerges. In the ROC curve analysis, quantitative summary measures of the ROC curve such as the area under the curve (AUC) are often reported. District staff can compare the AUC between indicators. The larger the AUC is, the more effective an indicator is.

Appendix G. Indicator Criterion Decision Matrix *(from section 6)*

Unlike the comparisons done to select cut points, the comparisons between indicators should be based on criteria aligned with local priorities. Use this decision matrix (figure G1) to compare your district's priorities against an indicator criterion. See step 6.1 for examples.

Figure G1. Indicator Criterion Decision Matrix

Priority 1: Target only students who are at the highest risk of dropout	**Indicator criterion:** • Highest sensitivity OR • Identify smallest proportion of students as at risk
Priority 2: Identify and provide services to all grade 9 students at risk of dropout	**Indicator criterion:** • Highly specific OR • Identify a large number of students as at risk
Priority 3: Balance both targeting students at highest risk of dropout AND identifying and providing services to all grade 9 students at risk of dropout	**Indicator criterion:** • Balance measure (for example, Youden index)

Consider ranking the priorities for your district before selecting the indicator criterion that most closely matches that priority. Supposing that priority 1 is most aligned with the district goals, then sensitivity should be used to compare and rank the indicators. Table G1 can be used in the comparison and ranking process. *Note: An editable version of this table is available in* District Guide for Selecting Early Warning Indicators to Identify Students At Risk of Not Achieving Desired Graduation Outcomes: Editable Planning Forms.

Table G1. Template for Calculating Performance Measures of Indicators and Indicator Ranking

	True Positive	False Negative	True Negative	False Positive	Sensitivity	Rank of indicators (sensitivity)	Specificity	Rank of indicators (specificity)	Youden Index	Rank of indicators (Youden Index)
Name of indicators	TP	FN	TN	FP	TP/(TP + FP)	Rank	TN/(TN + FN)	Rank	Sensitivity + 1— Specificity	Rank
Indicator 1 (for example attendance less than 90%)										
Indicator 2 (for example, failed a core course)										
Indicator 3 (for example, GPA below 2.0)										
Indicator 4 (for example, being suspended)										
Indicator 5 (for example, failed math)										
….										