An Introduction to Logistic Regression Analysis and Reporting

CHAO-YING JOANNE PENG
KUK LIDA LEE
GARY M. INGERSOLL
Indiana University-Bloomington

ABSTRACT **The purpose of this article is to provide researchers, editors, and readers with a set of guidelines for what to expect in an article using logistic regression techniques. Tables, figures, and charts that should be included to comprehensively assess the results and assumptions to be verified are discussed. This article demonstrates the preferred pattern for the application of logistic methods with an illustration of logistic regression applied to a data set in testing a research hypothesis. Recommendations are also offered for appropriate reporting formats of logistic regression results and the minimum observation-to-predictor ratio. The authors evaluated the use and interpretation of logistic regression presented in 8 articles published in *The Journal of Educational Research* between 1990 and 2000. They found that all 8 studies met or exceeded recommended criteria.**

Key words: binary data analysis, categorical variables, dichotomous outcome, logistic modeling, logistic regression

Many educational research problems call for the analysis and prediction of a dichotomous outcome: whether a student will succeed in college, whether a child should be classified as learning disabled (LD), whether a teenager is prone to engage in risky behaviors, and so on. Traditionally, these research questions were addressed by either ordinary least squares (OLS) regression or linear discriminant function analysis. Both techniques were subsequently found to be less than ideal for handling dichotomous outcomes due to their strict statistical assumptions, i.e., linearity, normality, and continuity for OLS regression and multivariate normality with equal variances and covariances for discriminant analysis (Cabrera, 1994; Cleary & Angel, 1984; Cox & Snell, 1989; Efron, 1975; Lei & Koehly, 2000; Press & Wilson, 1978; Tabachnick & Fidell, 2001, p. 521). Logistic regression was proposed as an alternative in the late 1960s and early 1970s (Cabrera, 1994), and it became routinely available in statistical packages in the early 1980s.

Since that time, the use of logistic regression has increased in the social sciences (e.g., Chuang, 1997; Janik & Kravitz, 1994; Tolman & Weisz, 1995) and in educational research—especially in higher education (Austin, Yaffee, & Hinkle, 1992; Cabrera, 1994; Peng & So, 2002a; Peng, So, Stage, & St. John, 2002. With the wide availability of sophisticated statistical software for high-speed computers, the use of logistic regression is increasing. This expanded use demands that researchers, editors, and readers be attuned to what to expect in an article that uses logistic regression techniques. What tables, figures, or charts should be included to comprehensibly assess the results? What assumptions should be verified? In this article, we address these questions with an illustration of logistic regression applied to a data set in testing a research hypothesis. Recommendations are also offered for appropriate reporting formats of logistic regression results and the minimum observation-to-predictor ratio. The remainder of this article is divided into five sections: (1) Logistic Regression Models, (2) Illustration of Logistic Regression Analysis and Reporting, (3) Guidelines and Recommendations, (4) Evaluations of Eight Articles Using Logistic Regression, and (5) Summary.

Logistic Regression Models

The central mathematical concept that underlies logistic regression is the logit—the natural logarithm of an odds ratio. The simplest example of a logit derives from a 2×2 contingency table. Consider an instance in which the distribution of a dichotomous outcome variable (a child from an inner city school who is recommended for remedial reading classes) is paired with a dichotomous predictor variable (gender). Example data are included in Table 1. A test of independence using chi-square could be applied. The results yield $\chi^2(1) = 3.43$. Alternatively, one might prefer to assess

Address correspondence to Chao-Ying Joanne Peng, Department of Counseling and Educational Psychology, School of Education, Room 4050, 201 N. Rose Ave., Indiana University, Bloomington, IN 47405–1006. (E-mail: peng@indiana.edu)

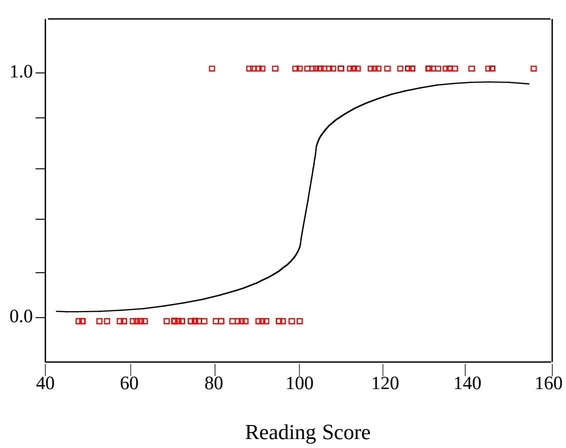
a boy's odds of being recommended for remedial reading instruction relative to a girl's odds. The result is an odds ratio of 2.33, which suggests that boys are 2.33 times more likely, than not, to be recommended for remedial reading classes compared with girls. The odds ratio is derived from two odds (73/23 for boys and 15/11 for girls); its natural logarithm [i.e., ln(2.33)] is a logit, which equals 0.85. The value of 0.85 would be the regression coefficient of the gender predictor if logistic regression were used to model the two outcomes of a remedial recommendation as it relates to gender.

Generally, logistic regression is well suited for describing and testing hypotheses about relationships between a categorical outcome variable and one or more categorical or continuous predictor variables. In the simplest case of linear regression for one continuous predictor *X* (a child's reading score on a standardized test) and one dichotomous outcome variable *Y* (the child being recommended for remedial reading classes), the plot of such data results in two parallel lines, each corresponding to a value of the dichotomous outcome (Figure 1). Because the two parallel lines are difficult to be described with an ordinary least squares regression equation due to the dichotomy of outcomes, one may instead create categories for the predictor and compute the mean of the outcome variable for the respective categories. The resultant plot of categories' means will appear linear in the middle, much like what one would expect to see on an ordinary scatter plot, but curved at the ends (Figure 1, the S-shaped curve). Such a shape, often referred to as sigmoidal or S-shaped, is difficult to describe with a linear equation for two reasons. First, the extremes do not follow a linear trend. Second, the errors are neither normally distributed nor constant across the entire range of data (Peng, Manz, & Keck, 2001). Logistic regression solves these problems by applying the logit transformation to the dependent variable. In essence, the logistic model predicts the logit of *Y* from *X*. As stated earlier, the logit is the natural logarithm (ln) of odds of *Y*, and odds are ratios of probabilities (π) of *Y* happening (i.e., a student is recommended for remedial reading instruction) to probabilities (1 – π) of *Y* not happening (i.e., a student is not recommended for remedial reading instruction). Although logistic regression can accommodate categorical outcomes that are polytomous, in this article we focus on dichotomous outcomes only. The illustration presented in this article can be extended easily to polytomous variables with ordered (i.e., ordinal-scaled) or unordered (i.e., nominal-scaled) outcomes.

Table 1.—Sample Data for Gender and Recommendation for Remedial Reading Instruction

Remedial reading instruction	Gender: Boys	Gender: Girls	Total
Recommended (coded as 1)	73	15	88
Not recommended (coded as 0)	23	11	34
Total	96	26	122

Figure 1. Relationship of a Dichotomous Outcome Variable, *Y* (1 = Remedial Reading Recommended, 0 = Remedial Reading Not Recommended) With a Continuous Predictor, Reading Scores



The simple logistic model has the form

$$\text{logit}(Y) = natural\,\log(odds) = \ln\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta X. \qquad (1)$$

For the data in Table 1, the regression coefficient (β) is the logit (0.85) previously explained. Taking the antilog of Equation 1 on both sides, one derives an equation to predict the probability of the occurrence of the outcome of interest as follows:

$$\pi = Probability(Y = \text{outcome of interest} \mid X = x,\ a\ specific\ value\ of\ X) = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}, \qquad (2)$$

where π is the probability of the outcome of interest or "event," such as a child's referral for remedial reading classes, α is the *Y* intercept, β is the regression coefficient, and *e* = 2.71828 is the base of the system of natural logarithms. *X* can be categorical or continuous, but *Y* is always categorical. According to Equation 1, the relationship between logit (*Y*) and *X* is linear. Yet, according to Equation 2, the relationship between the probability of *Y* and *X* is nonlinear. For this reason, the natural log transformation of the odds in Equation 1 is necessary to make the relationship between a categorical outcome variable and its predictor(s) linear.

The value of the coefficient β determines the direction of the relationship between *X* and the logit of *Y*. When β is greater than zero, larger (or smaller) *X* values are associated with larger (or smaller) logits of *Y*. Conversely, if β is less than zero, larger (or smaller) *X* values are associated with smaller (or larger) logits of *Y*. Within the framework of inferential statistics, the null hypothesis states that β equals zero, or there is no linear relationship in the population. Rejecting such a null hypothesis implies that a linear relationship exists between *X* and the logit of *Y*. If a predictor is binary, as in the Table 1 example, then the odds ratio is equal to *e*, the natural logarithm base, raised to the exponent of the slope β (e^{β}).

Extending the logic of the simple logistic regression to multiple predictors (say X_1 = reading score and X_2 = gender), one can construct a complex logistic regression for Y (recommendation for remedial reading programs) as follows:

$$\text{logit}(Y) = \ln\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta_1 X_1 + \beta_2 X_2. \quad (3)$$

Therefore,

$$\pi = \text{Probability } (Y = \textit{outcome of interest} \mid X_1 = x_1, X_2 = x_2$$
$$= \frac{e^{\alpha+\beta_1 X_1+\beta_2 X_2}}{1+e^{\alpha+\beta_1 X_1+\beta_2 X_2}}, \quad (4)$$

where π is once again the probability of the event, α is the Y intercept, βs are regression coefficients, and Xs are a set of predictors. α and βs are typically estimated by the maximum likelihood (ML) method, which is preferred over the weighted least squares approach by several authors, such as Haberman (1978) and Schlesselman (1982). The ML method is designed to maximize the likelihood of reproducing the data given the parameter estimates. Data are entered into the analysis as 0 or 1 coding for the dichotomous outcome, continuous values for continuous predictors, and dummy codings (e.g., 0 or 1) for categorical predictors.

The null hypothesis underlying the overall model states that all βs equal zero. A rejection of this null hypothesis implies that at least one β does not equal zero in the population, which means that the logistic regression equation predicts the probability of the outcome better than the mean of the dependent variable Y. The interpretation of results is rendered using the odds ratio for both categorical and continuous predictors.

Illustration of Logistic Regression Analysis and Reporting

For the sake of illustration, we constructed a hypothetical data set to which logistic regression was applied, and we interpreted its results. The hypothetical data consisted of reading scores and genders of 189 inner city school children (Appendix A). Of these children, 59 (31.22%) were recommended for remedial reading classes and 130 (68.78%) were not. A legitimate research hypothesis posed to the data was that “the likelihood that an inner city school child is recommended for remedial reading instruction is related to both his/her reading score and gender.” Thus, the outcome variable, remedial, was students being recommended for remedial reading instruction (1 = yes, 0 = no), and the two predictors were students’ reading score on a standardized test (X_1 = the reading variable) and gender (X_2 = gender). The reading scores ranged from 40 to 125 points, with a mean of 64.91 points and standard deviation of 15.29 points (Table 2). The gender predictor was coded as 1 = boy and 0 = girl. The gender distribution was nearly even with 49.21% (n = 93) boys and 50.79% (n = 96) girls.

Table 2.—Description of a Hypothetical Data Set for Logistic Regression

Remedial reading recommended?	Total sample (N)	Boys (n_1)	Girls (n_2)	Reading score M	Reading score SD
Yes	59	36	23	61.07	13.28
No	130	57	73	66.65	15.86
Summary	189	93	96	64.91	15.29

Logistic Regression Analysis

A two-predictor logistic model was fitted to the data to test the research hypothesis regarding the relationship between the likelihood that an inner city child is recommended for remedial reading instruction and his or her reading score and gender. The logistic regression analysis was carried out by the Logistic procedure in SAS® version 8 (SAS Institute Inc., 1999) in the Windows 2000 environment (SAS programming codes are found in Table 3). The result showed that

Predicted logit of (REMEDIAL) = 0.5340
+ (–0.0261)*READING + (0.6477)*GENDER. (5)

According to the model, the log of the odds of a child being recommended for remedial reading instruction was negatively related to reading scores ($p < .05$) and positively related to gender ($p < .05$; Table 3). In other words, the higher the reading score, the less likely it is that a child would be recommended for remedial reading classes. Given the same reading score, boys were more likely to be recommended for remedial reading classes than girls because boys were coded to be 1 and girls 0. In fact, the odds of a boy being recommended for remedial reading programs were 1.9111 (= $e^{0.6477}$; Table 3) times greater than the odds for a girl.

The differences between boys and girls are depicted in Figure 2, in which predicted probabilities of recommendations are plotted for each gender group against various reading scores. From this figure, it may be inferred that for a given score on the reading test (e.g., 60 points), the probability of a boy being recommended for remedial reading programs is higher than that of a girl. This statement is also confirmed by the positive coefficient (0.6477) associated with the gender predictor.

Evaluations of the Logistic Regression Model

How effective is the model expressed in Equation 5? How can an educational researcher assess the soundness of a logistic regression model? To answer these questions, one must attend to (a) overall model evaluation, (b) statistical tests of individual predictors, (c) goodness-of-fit statistics, and (d) validations of predicted probabilities. These evaluations are illustrated below for the model based on Equation 5, also referred to as Model 5.

Overall model evaluation. A logistic model is said to provide a better fit to the data if it demonstrates an improvement over the intercept-only model (also called the null model). An

Table 3.—Logistic Regression Analysis of 189 Children's Referrals for Remedial Reading Programs by SAS PROC LOGISTIC (Version 8)

Predictor	β	*SE* β	Wald's χ^2	*df*	*p*	e^{β} (odds ratio)
Constant	0.5340	0.8109	0.4337	1	.5102	NA
Reading	–0.0261	0.0122	4.5648	1	.0326	0.9742
Gender (1 = boys, 0 = girls)	0.6477	0.3248	3.9759	1	.0462	1.9111
Test			χ^2	*df*	*p*	
Overall model evaluation						
Likelihood ratio test			10.0195	2	.0067	
Score test			9.5177	2	.0086	
Wald test			9.0626	2	.0108	
Goodness-of-fit test						
Hosmer & Lemeshow			7.7646	8	.4568	

Note. SAS programming codes: [PROC LOGISTIC; MODEL REMEDIAL=READING GENDER/CTABLE PPROB=(0.1 TO 1.0 BY 0.1) LACKFIT RSQ;]. Cox and Snell R^2 = .0516. Nagelkerke R^2 (Max rescaled R^2) = .0726. Kendall's Tau-*a* = .1180. Goodman-Kruskal Gamma = .2760. Somers's D_{xy} = .2730. *c*-statistic = 63.60%. All statistics reported herein use 4 decimal places in order to maintain statistical precision. NA = not applicable.

intercept-only model serves as a good baseline because it contains no predictors. Consequently, according to this model, all observations would be predicted to belong in the largest outcome category. An improvement over this baseline is examined by using three inferential statistical tests: the likelihood ratio, score, and Wald tests. All three tests yield similar conclusions for the present data (Table 3), namely, that the logistic Model 5 was more effective than the null model. For other data sets, these three tests may not lead to similar conclusions. When this happens, readers are advised to rely on the likelihood ratio and score tests only (Menard, 1995).

Statistical tests of individual predictors. The statistical significance of individual regression coefficients (i.e., βs) is tested using the Wald chi-square statistic (Table 3). According to Table 3, both reading score and gender were significant predictors of inner city school children's referrals for remedial reading programs ($p < .05$). The test of the intercept (i.e., the constant in Table 3) merely suggests whether an intercept should be included in the model. For the present data set, the test result ($p > .05$) suggested that an alternative model without the intercept might be applied to the data.

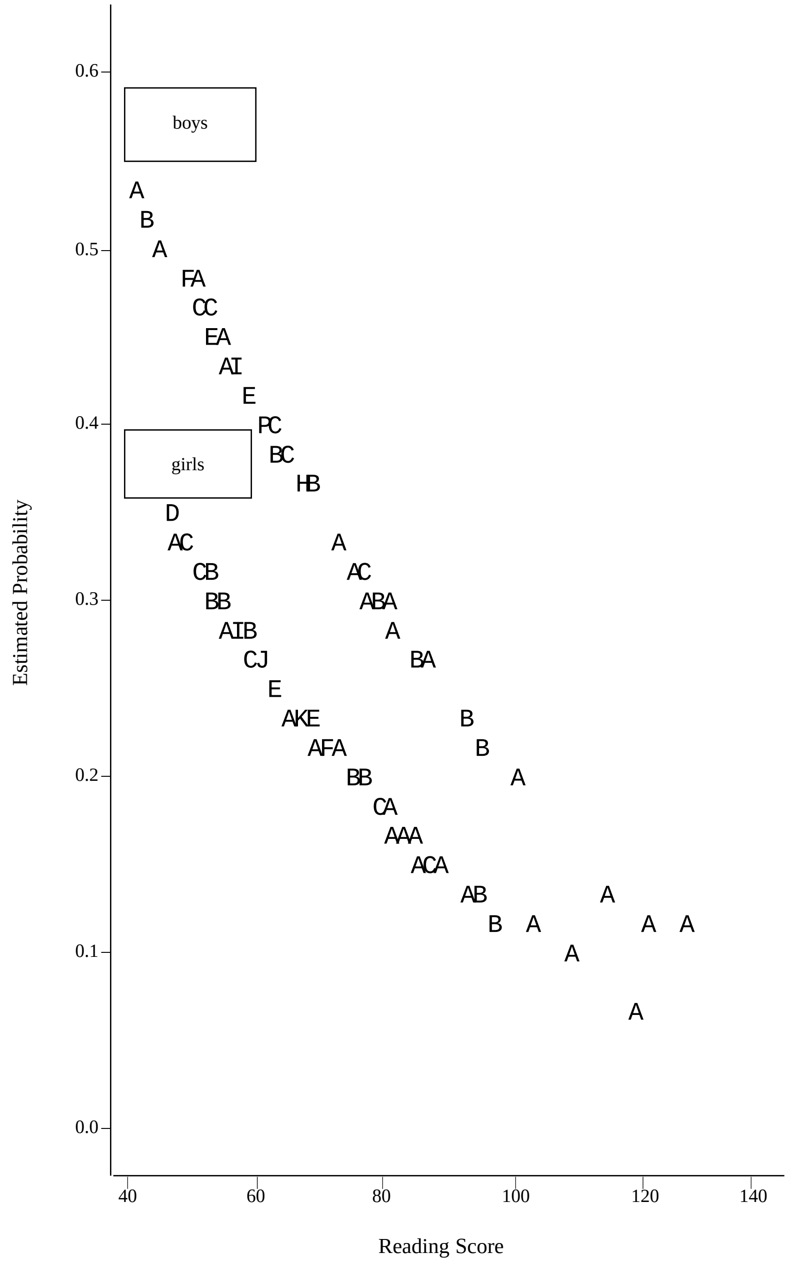
Goodness-of-fit statistics. Goodness-of-fit statistics assess the fit of a logistic model against actual outcomes (i.e., whether a referral is made for remedial reading programs). One inferential test and two descriptive measures are presented in Table 3. The inferential goodness-of-fit test is the Hosmer–Lemeshow (H–L) test that yielded a $\chi^2(8)$ of 7.7646 and was insignificant ($p > .05$), suggesting that the model was fit to the data well. In other words, the null hypothesis of a good model fit to data was tenable.

The H–L statistic is a Pearson chi-square statistic, calculated from a $2 \times g$ table of observed and estimated expected frequencies, where g is the number of groups formed from the estimated probabilities. Ideally, each group should have an equal number of observations, the number of groups should exceed 5, and expected frequencies should be at least 5. For the present data, the number of observations in each group was mostly 19 (3 groups) or 20 (5 groups); 1 group had 21 observations and another had 11 observations. The number of groups was 10, and the expected frequencies were at or exceeded 5 in 90% of cells. Thus, it was concluded that the conditions were met for reporting the HL test statistic.

Two additional descriptive measures of goodness-of-fit presented in Table 3 are R^2 indices, defined by Cox and Snell (1989) and Nagelkerke (1991), respectively. These indices are variations of the R^2 concept defined for the OLS regression model. In linear regression, R^2 has a clear definition: It is the proportion of the variation in the dependent variable that can be explained by predictors in the model. Attempts have been devised to yield an equivalent of this concept for the logistic model. None, however, renders the meaning of variance explained (Long, 1997, pp. 104–109; Menard, 2000). Furthermore, none corresponds to predictive efficiency or can be tested in an inferential framework (Menard). For these reasons, a researcher can treat these two R^2 indices as supplementary to other, more useful evaluative indices, such as the overall evaluation of the model, tests of individual regression coefficients, and the goodness-of-fit test statistic.

Validations of predicted probabilities. As we explained earlier, logistic regression predicts the logit of an event outcome from a set of predictors. Because the logit is the natural log of the odds (or probability/[1–probability]), it can be transformed back to the probability scale. The resultant predicted probabilities can then be revalidated with the actual outcome to determine if high probabilities are indeed associated with events and low probabilities with nonevents. The degree to which predicted probabilities agree with actual outcomes is expressed as either a measure of association or a classification table. There are four measures

Figure 2. Predicted Probability of Being Referred for Remedial Reading Instructions Versus Reading Scores



Note. Plotting symbols A = 1 observation, B = 2 observations, C = 3 observations, and so forth.

of association and one classification table that are provided by SAS (Version 8).

The four measures of association are Kendall's Tau-*a*, Goodman-Kruskal's Gamma, Somers's *D* statistic, and the *c* statistic (Table 3). The Tau-*a* statistic is Kendall's rank-order correlation coefficient without adjustments for ties. The Gamma statistic is based on Kendall's coefficient but adjusts for ties. Gamma is more useful and appropriate than Tau-*a* when there are ties on both outcomes and predicted probabilities, as was the case with the present data (see Appendix A). The Gamma statistic for Model 5 is 0.2760 (Table 3). It is interpreted as 27.60% fewer errors made in predicting which of two children would be recommended for remedial reading programs by using the estimated probabilities than by chance alone (Demaris, 1992). Some caution is advised in using the Gamma statistic because (a) it has a tendency to overstate the strength of association between estimated probabilities and outcomes (Demaris), and (b) a value of zero does not necessarily imply independence when the data structure exceeds a 2 × 2 format (Siegel & Castellan, 1988).

Somers's *D* is a preferred extension of Gamma whereby one variable is designated as the dependent variable and the other the independent variable (Siegel & Castellan, 1988). There are two asymmetric forms of Somers's *D* statistic: D_{xy} and D_{yx}. Only D_{yx} correctly represents the degree of association between the outcome (*y*), designated as the dependent variable, and the estimated probability (*x*), designated as the independent variable (Demaris, 1992). Unfortunately, SAS computes only D_{xy} (Table 3), although this index can be corrected to D_{yx} in SAS (Peng & So, 1998).

The *c* statistic represents the proportion of student pairs with different observed outcomes for which the model correctly predicts a higher probability for observations with the event outcome than the probability for nonevent observations. For the present model, the *c* statistic is 0.6360 (Table 3). This means that for 63.60% of all possible pairs of children—one recommended for remedial reading programs and the other not—the model correctly assigned a higher probability to those who were recommended. The *c* statistic ranges from 0.5 to 1. A 0.5 value means that the model is no better than assigning observations randomly into outcome categories. A value of 1 means that the model assigns higher probabilities to all observations with the event outcome, compared with nonevent observations. If several models were fitted to the same data set, the model chosen as the best model should be associated with the highest *c* statistic. Thus, the *c* statistic provides a basis for comparing different models fitted to the same data or the same model fitted to different data sets.

In addition to these measures of association, SAS output includes a classification table that documents the validity of predicted probabilities (Table 4). The first two rows in Table 4 represent the two possible outcomes, and the two columns under the heading "Predicted" are for high and low probabilities, based on a cutoff point. The cutoff point may be specified by researchers or set at 0.5 by SAS. According to Table 4, with the cutoff set at 0.5, the prediction for children who were not recommended for remedial reading programs was more accurate than that for those who were. This observation was also supported by the magnitude of sensitivity (3.39%) compared to that of specificity (99.23%). Sensitivity measures the proportion of correctly classified events (i.e., those recommended for remedial reading programs), whereas specificity measures the proportion of correctly classified nonevents (those not recommended for remedial reading programs). Both false positive and false negative rates were a little more than 30%. The false positive rate measures the proportion of observations misclassified as events over all of those classified as events. The false negative therefore measures the proportion of observations misclassified as nonevents over all of those classified as nonevents. The overall correction prediction was 69.31%, an improvement over the chance level. In the opinion of Hosmer and Lemeshow (2000, p. 160), "the classification table is most appropriate when classification is a stated goal of the analysis; otherwise it should only supplement more rigorous methods of assessment of fit."

Table 4.—The Observed and the Predicted Frequencies for Remedial Reading Instructions by Logistic Regression With the Cutoff of 0.50

	Predicted		
Observed	Yes	No	% Correct
Yes	2	57	3.39
No	1	129	99.23
Overall % correct			69.31

Note. Sensitivity = 2/(2+57)% = 3.39%. Specificity = 129/(1+129)% = 99.23%. False positive = 1/(1+2)% = 33.33%. False negative = 57/(57+129)% = 30.65%.

Table 4 was prepared with SAS using a reduced-bias algorithm. The algorithm minimizes the bias of using the same observations both for model fitting and for predicting probabilities (SAS Institute Inc., 1999). According to a recent comparative study of six statistical packages that can be used for logistic regression (Peng & So, 2002b), SAS is the only package that uses this algorithm. Thus, entries in Table 4 would be slightly different if other software (such as SPSS) was used to prepare it.

Reporting and Interpreting Logistic Regression Results

In addition to the data presented in Tables 3 and 4 and Figure 2, it is helpful to demonstrate the relationship between the predicted outcome and certain characteristics found in observations. For the present data, this relationship is demonstrated in Table 5 for four cases (1–4) extracted from Appendix A, as well as for four observations (5–8) for whom reading scores were hypothesized at two levels for both genders. For the first four cases, the predicted probabilities of referrals for remedial reading programs were calculated using Equation 5. Even though these four cases were not perfectly predicted, the correct prediction rate was better than chance.

The last four hypothetical cases show the descending predicted probabilities of referrals for remedial reading programs as the reading scores increase for children of both genders. For each point increase on the reading score, the odds of being recommended for remedial reading programs decrease from 1.0 to 0.9742 (= $e^{-0.0261}$; Table 3). If the increase on the reading score was 10 points, the odds decreased from 1.0 to 0.7703 (= $e^{10*[-0.0261]}$). However, when the reading score was held as a constant, boys were predicted to be referred for remedial reading instructions with greater probability than girls. The differences between boys and girls are graphically shown in Figure 2 and confirmed previously by the positive coefficient (0.6477) of the gender predictor in Equation 5.

Table 5.—Predicted Probability of Being Referred for Remedial Reading Instructions for 8 Children

Case number	Reading score β = –0.0261	Gender β = 0.6477	Intercept = 0.5340	Predicted probability of being referred for remedial reading program	Actual outcome 1 = Yes, 0 = No
1	52.5	Boy	0.5340	0.4530	1
2	85	Boy	0.5340	0.2618	0
3	75	Girl	0.5340	0.1941	1
4	92	Girl	0.5340	0.1250	0
5	60	Boy	0.5340	0.4051	—
6	60	Girl	0.5340	0.2627	—
7	100	Boy	0.5340	0.1934	—
8	100.5	Girl	0.5340	0.1115	—

The odds of a boy being recommended for remedial reading programs were 1.9111 (= $e^{0.6477}$; Table 3) times greater than the odds for a girl.

In terms of the research hypothesis posed earlier to the hypothetical data—"the likelihood that an inner city school child is recommended for remedial reading instruction is related to both his/her reading score and gender"—logistic regression results supported this proposition. Specifically, the likelihood of a child being recommended for remedial reading instruction was negatively related to his or her reading scores. However, given the same reading score, boys were more likely to be recommended for remedial reading classes than girls. We reached this conclusion with multiple evidences: the significant test result of the logistic model, statistically significant test results of both predictors, insignificant HL test of goodness-of-fit, and several descriptive measures of associations between predicted probabilities and data.

Guidelines and Recommendations

What Tables, Figures, or Charts Should Be Included to Comprehensively Assess the Result?

In presenting the assessment of logistic regression results, researchers should include sufficient information to address the following:

- an overall evaluation of the logistic model
- statistical tests of individual predictors
- goodness-of-fit statistics
- an assessment of the predicted probabilities

Table 3 illustrates the presentation of the first three types of information and Table 4 the fourth. To illustrate the impact of a statistically significant categorical predictor (e.g., gender in our example) on the dichotomous dependent variable (e.g., recommendation for remedial reading programs), it is helpful to include a figure such as Figure 2. It is our recommendation that logistic regression results be reported, similar to those in Tables 3 and 4 and Figure 2, to help communicate findings to readers.

A model's adequacy should be justified by multiple indicators, including an overall test of all parameters, a statistical significance test of each predictor, the goodness-of-fit statistics, the predictive power of the model, and the interpretability of the model. Furthermore, researchers should pay attention to mathematical definitions of statistics (such as D_{xy}) generated by the statistical package of choice. Among the packages that perform logistic regression, none was found to be error free (Peng & So, 2002b). A reference to the software should inform readers of programming mistakes and limitations, and help researchers verify results with another statistical package. A recent review of six statistical software programs, conducted by Peng and So (2002b, pp. 55–56) for performing logistic regression, concluded that

> The versatile SAS logisitic and BMDP LR [were recommended] for researchers experienced with logistic regression techniques and programming. . . . Several unique goodness-of-fit indices and selection methods are provided in SAS. Its ability to fit a broad class of binary response models, plus its provision to correct for over-sampling, over-dispersion, and bias introduced into predicted probabilities, sets it apart from the other five. . . . If either SPSS LOGISTIC REGRESSION or SYSTAT LOGIT is the only package available, researchers must be aware that both compute the goodness-of-fit and diagnostic statistics from individual observations. Consequently, these statistics are inappropriate for statistical tests. With dazzling graphic interfaces, both packages are user-friendly.
>
> MINITAB BLOGISTIC is the simplest to use. It adopts the hierarchical modeling restriction in direct modeling. . . . A substantial number of goodness-of-fit indices are available including the unique Brown statistic. However, the absence of predictor selection methods may make it less appealing to some researchers. . . . STATA LOGISTIC provides the most detailed information on parameter estimates, yet its goodness-of-fit indices are limited. We recommend MINITAB and STATA for beginners, although experienced researchers may also employ them for logistic regression.

What Assumptions Should Be Verified?

Unlike discriminant function analysis, logistic regression does not assume that predictor variables are distributed as a multivariate normal distribution with equal covariance matrix. Instead, it assumes that the binomial distribution describes the distribution of the errors that equal the actual *Y* minus the predicted *Y*. The binomial distribution is also the assumed distribution for the conditional mean of the dichotomous outcome. This assumption implies that the same probability is maintained across the range of predictor values. The binomial assumption may be tested by the normal z test (Siegel & Castellan, 1988) or may be taken to be robust as long as the sample is random; thus, observations are independent from each other.

Recommended Reporting Formats of Logistic Regression

In terms of reporting logistic regression results, we recommend presenting the complete logistic regression model including the Y-intercept (similar to Equation 5), odds ratios, and a table such as Table 5 to illustrate the relationship between outcomes and observations with profiles of certain characteristics. Odds ratios are directly derived from regression coefficients in a logistic model. If β_j represents the regression coefficient for predictor X_j, then exponentiating β_j yields the odds ratio. When all other predictors are held at a constant, the odds ratio means the change in the odds of Y given a unit change in X_j. It is one of three epidemiological measures of effect that have been recently recommended by psychologists for informing public policy makers (Scott, Mason, & Chapman, 1999). Three conditions must be met before odds ratios can be interpreted sensibly: (a) the predictor X_j must not interact with another predictor; (b) the predictor X_j must be represented by a single term in the model; and (c) a one-unit change in the predictor X_j must be meaningful and relevant. It is worth noting that odds ratios and odds are two different concepts. They are related but not in a linear fashion. Likewise, the relationship between the predicted probability and odds, though positive, is not linear either.

Recommended Minimum Observation-to-Predictor Ratio

In terms of the adequacy of sample sizes, the literature has not offered specific rules applicable to logistic regression (Peng et al., 2002). However, several authors on multivariate statistics (Lawley & Maxwell, 1971; Marascuilo & Levin, 1983; Tabachnick & Fidell, 1996, 2001) have recommended a minimum ratio of 10 to 1, with a minimum sample size of 100 or 50, plus a variable number that is a function of the number of predictors.

Evaluations of Eight Articles Using Logistic Regression

To help understand how logistic regression has been applied by authors of articles published in *The Journal of Educational Research* (JER), we reviewed articles that used this technique between 1990 and 2000. During this period, eight articles were found to have used logistic regression. The criterion used in selecting articles was simple: at least one empirical analysis in the article must have been conducted to derive the logistic model and its regression coefficients. This criterion excluded any article that relied on others' work to derive the model or merely performed a logarithm or logit transformation of the dependent or the independent variable. A complete list of these eight articles is found in Appendix B.

A breakdown of the articles by year showed that, prior to 1993, there was no article that used logistic regression. In 1993, 1994, 1996, and 1997, one article per year applied logistic regression; in 1998 and 2000, there were two per year. This trend mirrors the pattern that was found in higher education journals (Peng et al., 2002, except that the rise of logistic regression began a year earlier, in 1992, in higher education journals.

The research questions addressed in the eight articles included American Indian adolescents' educational commitment (Trusty, 2000), school performance and activities (Alexander, Dauber, & Entwisle, 1996; McNeal, 1998; Smith, 1997), students at-risk (Meisels & Liaw, 1993; Rush & Vitale, 1994), family connectedness (Machamer & Gruber, 1998), and parents' conceptions of kindergarten readiness (Diamond, Reagan, & Bandyk, 2000). One central theme shared by all was education-related adjustment and performance. The dependent variable was dichotomous, whether it was retention in school, dropping-out from high school, or readiness for kindergarten. The predictors typically included a combination of demographic characteristics (such as age, gender, and ethnicity) and cognitive, affective, or personality-related measures. The objective of each study was to predict or to distinguish the outcome categories on the basis of predictors.

To test pertinent research hypotheses, the authors of these eight articles used three modeling approaches: direct, sequential, and stepwise modeling. Of these three, only direct and sequential models were controlled and implemented by researchers (Peng & So, 2002a). Three studies investigated interactions among predictors (Alexander, Dauber, & Entwisle, 1996; Meisels & Liaw, 1993; Trusty, 2000); the others did not. Though not all prior studies have always followed the guidelines and recommendations outlined in the previous section, all authors are credited for making substantive contributions as well as for introducing logistic regression into the field of educational research.

The Assessment of Logistic Regression Results

Four groups of authors (Alexander, Dauber, & Entwisle, 1996; Diamond, Reagan, & Bandyk, 2000; McNeal, 1998; Rush & Vitale, 1994) evaluated the overall logistic model; all reported tests of individual predictors, such as those shown in Table 3. Evidence of the goodness-of-fit of logistic models was provided by the R^2 index for either the entire model or for each predictor (Alexander, Dauber, & Entwisle, 1996; Diamond, Reagan, & Bandyk, 2000; Rush & Vitale, 1994; Trusty, 2000). None reported the HL test. Only one study (Rush & Vitale, 1994) validated predicted probabilities against data in the Table 4 format. Our review, however, uncovered two minor discrepancies in Rush and Vitale's (1994) classification table (Table 5, p. 331). In Table 5, the hit rate was reported to be 90.6%, and misclassifications were 223 for at-risk children and 112 for non-at-risk children. The text on page 332 reported a hit rate of

90.71%, and the misclassifications were 223 versus 115, written on page 329. None reported measures of association such as Kendall's Tau-*a*, Goodman-Kruskal's Gamma, Somers's *D* statistic, or the *c* statistic. None mentioned the statistical package that performed the logistic analysis, although Rush and Vitale (1994) used SPSS-X to perform factor analysis, and those results were subsequently incorporated into logistic regression.

Verification of the Binomial Assumption

As stated earlier, logistic regression has only one assumption: The binomial distribution is the assumed distribution for the conditional mean of the dichotomous outcome. This assumption implies that the same probability is maintained across the range of predictor values. Though none of the eight studies verified or tested this assumption, the binomial assumption is known to be robust as long as the sample is random; thus, observations are independent from each other. Samples used in the eight studies did not appear to be nonrandom, nor did they have inherent dependence among observations. Thus, the binomial assumption appeared to be robust underlying all logistic analyses conducted by these eight studies.

Reporting Formats of Logistic Regression Results

Five of the articles (Alexander, Dauber, & Entwisle, 1996; Diamond, Reagan, & Bandyk, 2000; Machamer & Gruber, 1998) did present the logistic model. Of those five, three (Meisels & Liaw, 1993; Smith, 1997; Trusty, 2000) did not include intercepts in the logistic model. Odds ratios were reported in three studies (McNeal, 1998; Meisels & Liaw, 1993; Rush & Vitale, 1994), and odds were reported in one (Trusty, 2000).

One study presented results in terms of marginal probabilities (McNeal, 1998). The use of marginal probabilities has been criticized by Long (1997, pp. 74–75) and Peng et al. (2002) because marginal probabilities do not correspond to a fixed change in the predicted probabilities that will occur if there is a discrete change in one predictor (e.g., reading), while other predictors are realized at a constant. In other words, the marginal probability corresponding to a change in reading from 50 points to 60 points is different from that associated with another 10-point change from, say, 60 to 70 points. Furthermore, if other predictors (e.g., age) are held at their respective means, the corresponding marginal probability for reading is different from that computed at other values (e.g., the mode). One study did not explain how a categorical predictor was coded in the data (Diamond, Reagan, & Bandyk, 2000). These reporting formats create difficulties for readers to verify results with another sample or at another time or place.

One study (Trusty, 2000) coded a dichotomous predictor as 1 (do not have a computer in the home) and 2 (do have a computer), instead of the recommended 0 and 1, or –1/2 and +1/2 (Peng & So, 2002b). This practice is not necessarily incorrect; it simply makes the interpretation of the regression coefficient awkward and less direct.

Observation to Predictor Ratio

As stated earlier, the literature has not offered specific rules that are applicable to logistic regression (Peng et al., 2002). On the basis of the general rule of a minimum ratio of 10 to 1, with a minimum sample size of 100, all eight studies met and even exceeded our recommendation. Therefore, the results reported in these studies were considered stable.

Summary

In this paper, we demonstrate that logistic regression can be a powerful analytical technique for use when the outcome variable is dichotomous. The effectiveness of the logistic model was shown to be supported by (a) significance tests of the model against the null model, (b) the significance test of each predictor, (c) descriptive and inferential goodness-of-fit indices, (d) and predicted probabilities.

During the last decade, logistic regression has been gaining popularity. The trend is evident in the JER and higher education journals. Such popularity can be attributed to researchers' easy access to sophisticated statistical software that performs comprehensive analyses of this technique. It is anticipated that the application of the logistic regression technique is likely to increase. This potential expanded usage demands that researchers, editors, and readers be coached in what to expect from an article that uses the logistic regression technique. What tables, charts, or figures should be included? What assumptions should be verified? And how comprehensive should the presentation of logistic regression results be? It is hoped that this article has answered these questions with an illustration of logistic regression applied to a data set and with guidelines and recommendations offered on a preferred pattern of application of logistic methods.

ACKNOWLEDGMENTS

We wish to thank James D. Raths and one anonymous consulting editor for their very helpful comments on earlier drafts of this article.

REFERENCES

Austin, J. T., Yaffee, R. A., & Hinkle, D. E. (1992). Logistic regression for research in higher education. *Higher Education: Handbook of Theory and Research, 8,* 379–410.

Cabrera, A. F. (1994). Logistic regression analysis in higher education: An applied perspective. *Higher Education: Handbook of Theory and Research, Vol. 10,* 225–256.

Chuang, H. L. (1997). High school youth's dropout and re-enrollment behavior. *Economics of Education Review, 16*(2), 171–186.

Cleary, P. D., & Angel, R. (1984). The analysis of relationships involving dichotomous dependent variables. *Journal of Health and Social Behavior, 25,* 334–348.

Cox, D. R., & Snell, E. J. (1989). *The analysis of binary data* (2nd ed.). London: Chapman and Hall.

Demaris, A. (1992). *Logit modeling: Practical applications.* Newbury Park, CA: Sage.

Efron, B. (1975). The efficiency of logistic regression compared to normal discriminant analysis. *Journal of the American Statistical Association, 70,* 892–898.

Haberman, S. (1978). *Analysis of qualitative data* (Vol. 1). New York: Academic Press.
Hosmer, D. W., Jr., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: Wiley.
Janik, J., & Kravitz, H. M. (1994). Linking work and domestic problems with police suicide. *Suicide and Life Threatening Behavior, 24*(3), 267–274.
Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method*. London: Butterworth & Co.
Lei, P.-W., & Koehly, L. M. (2000, April). *Linear discriminant analysis versus logistic regression: A comparison of classification errors.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
Long, J. S. (1997). *Regression models for categorical and limited dependent variables.* Thousand Oaks, CA: Sage.
Marascuilo, L. A., & Levin, J. R. (1983). *Multivariate statistics in the social sciences: A researcher's guide*. Monterey, CA: Brooks/Cole.
Menard, S. (1995). *Applied logistic regression analysis* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–106). Thousand Oaks, CA: Sage.
Menard, S. (2000). Coefficients of determination for multiple logistic regression analysis. *The American Statistician, 54*(1), 17–24.
Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78*, 691–692.
Peng, C. Y., Manz, B. D., & Keck, J. (2001). Modeling categorical variables by logistic regression. *American Journal of Health Behavior, 25*(3), 278–284.
Peng, C. Y., & So, T. S. (1998). If there is a will, there is a way: Getting around defaults of PROC LOGISTIC in SAS. *Proceedings of the MidWest SAS Users Group 1998 Conference* (pp. 243–252). Retrieved from http://php.indiana.edu/~tso/articles/mwsug98.pdf
Peng, C. Y., & So, T. S. H. (2002a). Modeling strategies in logistic regression. *Journal of Modern Applied Statistical Methods, 14*, 147–156.
Peng, C. Y., & So, T. S. H. (2002b). Logistic regression analysis and reporting: A primer. *Understanding Statistics, 1*(1), 31–70.
Peng, C. Y., So, T. S., Stage, F. K., & St. John, E. P. (2002). The use and interpretation of logistic regression in higher education journals: 1988–1999. *Research in Higher Education, 43*, 259–293.
Peterson, T. (1984). A comment on presenting results from logit and probit models. *American Sociological Review, 50*(1), 130–131.
Press, S. J., & Wilson, S. (1978). Choosing between logistic regression and discriminant analysis. *Journal of the American Statistical Association, 73*, 699–705.
Ryan, T. P. (1997). *Modern regression methods.* New York: Wiley.
SAS Institute Inc. (1999). *SAS/STAT® user's guide* (Version 8, Vol. 2). Cary, NC: Author.
Schlesselman, J. J. (1982). *Case control studies: Design, control, analysis.* New York: Oxford University Press.
Scott, K. G., Mason, C. A., & Chapman, D. A. (1999). The use of epidemiological methodology as a means of influencing public policy. *Child Development, 70*(5), 1263–1272.
Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral science* (2nd ed.). New York: McGraw-Hill.
Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: Harper Collins.
Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Needham Heights, MA: Allyn & Bacon.
Tolman, R. M., & Weisz, A. (1995). Coordinated community intervention for domestic violence: The effects of arrest and prosecution on recidivism of woman abuse perpetrators. *Crime and Delinquency, 41*(4), 481–495.

APPENDIX A
Hypothetical Data for Logistic Regression

ID	Gender	Reading score	Remedial reading recommended?
1	Boy	91.0	No
2	Boy	77.5	No
3	Girl	52.5	No
4	Girl	54.0	No
5	Girl	53.5	No
6	Boy	62.0	No
7	Girl	59.0	No
8	Boy	51.5	No
9	Girl	61.5	No
10	Girl	56.5	No
11	Boy	47.5	No
12	Boy	75.0	No
13	Boy	47.5	No
14	Boy	53.5	No
15	Girl	50.0	No
16	Girl	50.0	No
17	Boy	49.0	No
18	Girl	59.0	No
19	Boy	60.0	No
20	Girl	60.0	No
21	Boy	60.5	No
22	Girl	50.0	No
23	Girl	101.0	No
24	Boy	60.0	No
25	Boy	60.0	No
26	Girl	83.5	No
27	Girl	61.0	No
28	Girl	75.0	No
29	Boy	84.0	No
30	Boy	56.5	No
31	Boy	56.5	No
32	Girl	45.0	No
33	Boy	60.5	No
34	Girl	77.5	No
35	Boy	62.5	No
36	Girl	70.0	No
37	Girl	69.0	No
38	Girl	62.0	No
39	Girl	107.5	No
40	Girl	54.5	No
41	Boy	92.5	No
42	Girl	94.5	No
43	Boy	65.0	No
44	Girl	80.0	No
45	Girl	45.0	No
46	Girl	45.0	No
47	Girl	66.0	No
48	Boy	66.0	No
49	Girl	57.5	No
50	Boy	42.5	No
51	Girl	60.0	No
52	Boy	64.0	No
53	Girl	65.0	No
54	Girl	47.5	No
55	Boy	57.5	No
56	Boy	55.0	No
57	Boy	55.0	No
58	Boy	76.5	No
59	Boy	51.5	No
60	Boy	59.5	No
61	Boy	59.5	No
62	Boy	59.5	No

(Appendix continues)

APPENDIX A—continued

ID	Gender	Reading score	Remedial reading recommended?
63	Boy	55.0	No
64	Girl	70.0	No
65	Boy	66.5	No
66	Boy	84.5	No
67	Boy	57.5	No
68	Boy	125.0	No
69	Girl	70.5	No
70	Boy	79.0	No
71	Girl	56.0	No
72	Boy	75.0	No
73	Boy	57.5	No
74	Boy	56.0	No
75	Girl	67.5	No
76	Boy	114.5	No
77	Girl	70.0	No
78	Girl	67.0	No
79	Boy	60.5	No
80	Girl	95.0	No
81	Girl	65.5	No
82	Girl	85.0	No
83	Boy	55.0	No
84	Boy	63.5	No
85	Boy	61.5	No
86	Boy	60.0	No
87	Boy	52.5	No
88	Girl	65.0	No
89	Girl	87.5	No
90	Girl	62.5	No
91	Girl	66.5	No
92	Boy	67.0	No
93	Girl	117.5	No
94	Girl	47.5	No
95	Girl	67.5	No
96	Girl	67.5	No
97	Girl	77.0	No
98	Girl	73.5	No
99	Girl	73.5	No
100	Girl	68.5	No
101	Girl	55.0	No
102	Girl	92.0	No
103	Boy	55.0	No
104	Girl	55.0	No
105	Boy	60.0	No
106	Boy	120.5	No
107	Girl	56.0	No
108	Girl	84.5	No
109	Girl	60.0	No
110	Boy	85.0	No
111	Girl	93.0	No
112	Boy	60.0	No
113	Girl	65.0	No
114	Girl	58.5	No
115	Girl	85.0	No
116	Boy	67.0	No
117	Girl	67.5	No
118	Boy	65.0	No
119	Girl	60.0	No
120	Boy	47.5	No
121	Girl	79.0	No

ID	Gender	Reading score	Remedial reading recommended?
122	Boy	80.0	No
123	Girl	57.5	No
124	Girl	64.5	No
125	Girl	65.0	No
126	Girl	60.0	No
127	Girl	85.0	No
128	Girl	60.0	No
129	Girl	58.0	No
130	Girl	61.5	No
131	Boy	60.0	Yes
132	Girl	65.0	Yes
133	Boy	93.5	Yes
134	Boy	52.5	Yes
135	Boy	42.5	Yes
136	Boy	75.0	Yes
137	Boy	48.5	Yes
138	Boy	64.0	Yes
139	Boy	66.0	Yes
140	Girl	82.5	Yes
141	Girl	52.5	Yes
142	Girl	45.5	Yes
143	Boy	57.5	Yes
144	Boy	65.0	Yes
145	Girl	46.0	Yes
146	Girl	75.0	Yes
147	Boy	100.0	Yes
148	Girl	77.5	Yes
149	Boy	51.5	Yes
150	Boy	62.5	Yes
151	Boy	44.5	Yes
152	Girl	51.0	Yes
153	Girl	56.0	Yes
154	Girl	58.5	Yes
155	Girl	69.0	Yes
156	Boy	65.0	Yes
157	Boy	60.0	Yes
158	Girl	65.0	Yes
159	Boy	65.0	Yes
160	Boy	40.0	Yes
161	Girl	55.0	Yes
162	Boy	52.5	Yes
163	Boy	54.5	Yes
164	Boy	74.0	Yes
165	Boy	55.0	Yes
166	Girl	60.5	Yes
167	Boy	50.0	Yes
168	Boy	48.0	Yes
169	Girl	51.0	Yes
170	Girl	55.0	Yes
171	Boy	93.5	Yes
172	Boy	61.0	Yes
173	Boy	52.5	Yes
174	Boy	57.5	Yes
175	Boy	60.0	Yes
176	Girl	71.0	Yes
177	Girl	65.0	Yes
178	Girl	60.0	Yes
179	Girl	55.0	Yes
180	Boy	60.0	Yes

(Appendix continues)

APPENDIX A—continued

ID	Gender	Reading score	Remedial reading recommended?
181	Boy	77.0	Yes
182	Boy	52.5	Yes
183	Girl	95.0	Yes
184	Boy	50.0	Yes
185	Girl	47.5	Yes
186	Boy	50.0	Yes
187	Boy	47.0	Yes
188	Boy	71.0	Yes
189	Girl	65.0	Yes

APPENDIX B
List of JER Articles Reviewed

1. Alexander, K. L., Dauber, S. L., & Entwisle, D. R. (1996). Children in motion: School transfers and elementary school performance. *The Journal of Educational Research*, *90*(1), 3–11.

2. Diamond, K. E., Reagan, A J., & Bandyk, J. E. (2000). Parents' conceptions of kindergarten readiness: Relationships with race, ethnicity, and development. *The Journal of Educational Research, 94*(2), 93–100.

3. Machamer, A. M., & Gruber, E. (1998). Secondary school, family, and educational risk: Comparing American Indian adolescents and their peers. *The Journal of Educational Research, 91*(6), 357–369.

4. McNeal, R. B., Jr. (1998). High school extracurricular activities: Closed structures and stratifying patterns of participation. *The Journal of Educational Research, 91*(3), 183–191.

5. Meisels, S. J., & Liaw, F.-R. (1993). Failure in grade: Do retained students catch up? *The Journal of Educational Research, 87*(2), 69–77.

6. Rush, S., & Vitale, P. A. (1994). Analysis for determining factors that place elementary students at risk. *The Journal of Educational Research, 87*(6), 325–333.

7. Smith, J. B. (1997). Effects of eighth-grade transition programs on high school retention and experiences. *The Journal of Educational Research, 90*(3), 144–152.

8. Trusty, J. (2000). High educational expectations and low achievement: Stability of educational goals across adolescence. *The Journal of Educational Research, 93*(6), 356–365.