

Articles

USING ACHIEVEMENT TEST SCORES TO PREDICT STUDENT SUCCESS IN ADULT BASIC EDUCATION

Cindy L. James

Thompson Rivers University

Leslee Francis-Pelton

University of Victoria

Abstract

Numerous prediction models of student success/nonsuccess in Adult Basic Education (ABE) have been designed and tested. Some of these studies indicate there is a significant relationship between the academic ability of ABE participants (as measured by some assessment tool) and their success/nonsuccess. This two-year study involving 153 participants was conducted to determine if student success or nonsuccess in an ABE mathematics course could be predicted by student scores on the Canadian Achievement Test – 2nd edition (CAT/2). A logistic regression model based on CAT/2 scores achieved by the Year 1 student cohort was moderately successful at predicting success/nonsuccess in that same group of students (70%), when students with modeled success probabilities of ≥ 0.5 (the “cutoff value”) were predicted to eventually succeed in the mathematics course. However, when the same model was tested against students in Year 2 of the study, the percentage of students accurately predicted to succeed or not succeed was slightly lower (65%). Thus, had the modeled probabilities of success been used to limit admission into the mathematics course, a significant number of students destined to succeed in the course would have been excluded. Lowering the cutoff value would have reduced this potential error, but at the expense of allowing large numbers of “non-success” students into the course.

*The Canadian Journal for the Study of Adult Education/
La Revue canadienne pour l'étude de l'éducation des adultes
19,2 November/novembre 2005 14-33*

ISSN 0835-4944 © Canadian Association for the Study of Adult Education/
L'Association canadienne pour l'étude de l'éducation des adultes

Résumé

De nombreux modèles prévisionnels liés à la réussite scolaire en éducation des adultes ont été mis au point et évalués. Certains d'entre eux ont fait ressortir un lien significatif entre les aptitudes des apprenants adultes (mesurées par un outil d'évaluation quelconque) et leur degré de réussite (ou de non réussite). Le but de cette recherche, qui s'est échelonnée sur deux ans et ayant impliqué 153 participants, était de voir s'il était possible de prédire la réussite ou l'échec d'un apprenant adulte à un cours de mathématique selon ses résultats à l'Épreuve canadienne de rendement pour adultes, version anglaise deuxième édition (CAT/2). Un modèle de régression logistique, élaboré en fonction des résultats de la cohorte d'apprenants de première année à l'Épreuve canadienne de rendement pour adultes, a dans une certaine mesure permis de prédire la réussite de ce même groupe d'apprenants (à 70 pourcent), stipulant que les apprenants avec un indice de probabilité de $\geq 0,5$ (la valeur limite) réussiraient éventuellement le cours de mathématiques. Toutefois les résultats ont été un peu moins probants avec les apprenants de deuxième année, puisque le taux d'exactitude des prédictions n'était que de 65 pourcent. Par conséquent, si les résultats à l'Épreuve canadienne de rendement pour adultes avaient été utilisés pour limiter l'admission au cours de mathématiques, bon nombre d'apprenants susceptibles de réussir le cours auraient été exclus. Il aurait été possible de restreindre la marge d'erreur en abaissant la valeur limite, mais cela aurait eu pour effet d'augmenter le nombre d'apprenants susceptibles d'échouer le cours.

Background

High incompleteness and failure rates directly affect the efficiency ratings of adult basic education (ABE) programs, not to mention their impacts in human and economic terms. It therefore is imperative that educational institutes increase their ABE student success rates by providing students with the optimum conditions for succeeding in their post-secondary education. Critical to student success is the careful diagnosis of educational needs of individual learners (Crandall, Lerche, & Marchilonis, 1984; Gravenberg & Rivers, 1987). Often mature students (students out of the educational system for at least two years) are not required to provide educational transcripts to enter ABE programs nor do they have to undergo educational assessment. Unfortunately, without the transcript or some type of educational assessment, students may register in ABE courses for which they are academically unprepared. Such inaccurate placements may be one of the reasons many students withdraw from or fail ABE courses. One obvious means of avoiding

unsuitable placement is to provide educational assessment testing for potential ABE students.

Some of the most widely-used assessment tests for ABE programs in North America are the Adult Basic Learning Examination (ABLE), the Canadian Achievement Test (CAT), the Canadian Adult Achievement Test (CAAT), the Differential Aptitude Test (DAT), the Tests of Adult Basic Education (TABE), the Wide Range Achievement Test (WRAT), and domestic tests designed by the individual institutes (Brand, 1995; Crandall et. al. 1984; Ehringhaus, 1991; Nurss, 1989; Stricht, 1990; Venezky, Bristow & Sabatini, 1997). Factors that should be considered when selecting a suitable assessment tool are test content, administrative time, cost, reliability and validity. If the tool is to be used specifically for placement, it is imperative that users conduct studies to evaluate the tool's *predictive validity*. The predictive validity of a test refers to how well a test predicts some future behavior of learners (Stricht, 1990). Standardized tests such as the ABLE, CAAT, CAT, DAT, and TABE have been the subject of reviews and critiques that analyze their strengths and weaknesses, and studies that compare the various testing tools or evaluate instruction and/or learning using pre-test and post-test scores (Farr, Moon & Williams, 1986; Frager, 1991; Stricht, 1990; Taylor, 1990). However, published research on their predictive validity is limited.

Results from the few studies that have been published confirm the existence of a relationship between performance on assessment tests and student success in ABE programs or courses. For example, Grulick (1987) examined the predictive validity of the TEC-MAT, a domestic entrance exam at Florence-Darlington Technical College, and compared it with that of the Scholastic Aptitude Test - Math (SAT-M), the Career Planning Program (CPP) test, and the Test of Adult Basic Education (TABE). The results from his study showed a significant correlation between the TEC-MAT scores and the students' first quarter GPAs ($r = 0.369$, $p_s < 0.05$) and mathematics course grades at the end of their first quarter ($r = 0.5011$, $p_s < 0.05$). Grulick also found a moderate to high correlation for the TEC-MAT with the CPP ($r = 0.548$, $p_s < 0.01$), SAT-MATH ($r = 0.580$, $p_s < 0.01$) and TABE instruments ($r = 0.840$, $p_s < 0.01$).

Other studies have shown a significant relationship exists between student success and TABE scores. For instance, Dirks and Jha (1994) tried to differentiate between students who completed and did not complete a course by testing two prediction models using demographic data and the TABE reading and mathematics scores. They found that a prediction model that

utilized the participant's age and their TABE reading and mathematics scores could successfully predict course completers 70% of the time and continuing students 58% of the time, but was not successful in predicting noncontinuing students. Another study by Venezky, Bristow and Sabatini (1997) used a variety of measures to evaluate how the TABE and several other literacy tests predicted actual placement in ABE and GED classes. They found that the TABE locator test was a more effective predictor of placement than any of the full TABE tests or the other literacy tests, and at least as effective as the TABE Total Reading. Since the TABE locator test only requires 37 minutes to administer and the full TABE test battery requires three hours to administer, Venezky et al. concluded that lengthy testing procedures were not necessary for placement.

The purpose of this study was to determine if the Canadian Achievement Test—2nd edition (CAT/2) would be a useful assessment tool for ABE placement. Specifically we wanted to assess the predictive validity of the mathematics subtests of the CAT/2 as measured by student performance in an adult basic education mathematics course.

Methods

This two-year study took place at Thompson Rivers University (TRU) in Kamloops, British Columbia, Canada. TRU offers a variety of university, college and technical programs including comprehensive Adult Basic Education programs. The primary data sources for this study were achievement test scores on the two mathematics sections of the Canadian Achievement Test—2nd edition and the final percentages in the ABE mathematics courses.

During the first year of the study (1999), data were collected from three classes of Math 050, a course that approximates the former British Columbia High School Introductory Mathematics 11. Two of the classes (labeled Class One and Class Two) had the same instructor and received their instruction over a standard 4-month academic semester. However, Class Two participated in a test-taking tutorial prior to writing the achievement test while Class One did not. A third class (Class Three) had the same course provided over a 7-week period by a different instructor and did not participate in a test-taking tutorial. Both instructors for these classes were experienced full-time faculty members of the institute's College Preparation program and employed similar teaching practices consisting of traditional lecture mixed with students-instructor question/answer sessions.

During the subsequent year (2000), data were collected from the same three classes of Math 050, namely two classes with the four month schedule and one class with the condensed time format (7 weeks). The test-taking tutorial was not administered to any of the students in this year. The instructor who taught Class One and Two in the first year of the study also taught all three classes in the second year.

Project Participants and Assessment Tools

One hundred and fifty-three adults participated in this study: seventy-one in the first year and eighty-two in the second year. The adults in the year-one cohort had an average age of 24.7 (sd = 7.9) with an even distribution of males (48%) to females (52%). The demographics of the year-two cohort were very similar to the year-one cohort with an average age of 24.6 (sd = 8.4) and a gender distribution of 49% males and 51% males. Since there was no placement procedure for Math 050, students in both cohorts self-selected their course and level. The main reason most of the students were enrolled in the ABE mathematics course was to upgrade their education in order to apply for post-secondary programs and/or to obtain credit towards the ABE Provincial Diploma.

Although the Canadian Achievement Test–2nd edition was not designed as a placement tool, it does measure the kind of literacy and numeracy skills expected in ABE programs. This was one of the main reasons for selecting the CAT/2 as a possible placement test for ABE programs at TRU and hence the assessment tool for this study. The other reasons were that the test questions focused on Canadian content, it was normed on a Canadian population, it was being used by other vocational programs at TRU as well as other educational institutions, it was relatively easy to administer and score, and it was fairly inexpensive.

There are eight levels of the CAT/2 related to grade ranges; the CAT/2-level 18, which is related to the grade ranges of 8.0 through 10.2, was used for this study (Canadian Test Centre, 1992). The CAT/2-level 18 contains eight tests in five content areas: reading, spelling, language, study skills and mathematics. This study focused on the mathematics content area and utilized the associated tests: Test 7-Mathematics Concepts and Applications and Test 8-Mathematics Computation. Test 7 measures a student's ability to apply mathematical concepts related to numeration, number theory, data interpretation, basic algebra, measurement, logical reasoning and basic geometry. It consists of 45 multiple choice questions which the students have 35 minutes to complete. Test 8 measures a student's ability to add, subtract, multiply, and divide whole numbers, decimals, fractions and integers, and to

solve problems involving percents, exponents and algebraic operations. Test 8 consists of 40 multiple choice questions which the students have 30 minutes to complete.

Unit tests, assignments, mid-term exams and final exams created by the ABE instructors were the other educational material used in this study to evaluate the students' achievement in Math 050.

Data Collection and Analysis

For all six groups, the CAT/2 mathematics subtests were administered to the participants during the initial week of classes. For the statistical analysis, student scores for correct answers on the CAT/2 Tests 7 and 8 were collected and converted to scale-scores to calculate a combined mathematics scale-score. Conversion tables used to convert the number-correct scores to scale-scores were obtained from the Canadian Test Centre's Technical Bulletin (Canadian Test Centre, 1992). In addition, each student's achievement in Math 050 was classified as a success (passed the course) or nonsuccess (failed or did not complete the course).

Using data from the year-one cohort, logistic regression analysis was employed to develop three individual models for each class as well as a "comprehensive" model for all of the classes combined. Logistic regression is similar to linear regression except the outcome variable is dichotomous, not continuous. In logistic regression, the probability of an event occurring is determined by one or more predictor variables. For analysis that involves only one predictor variable, the probability values are obtained using the equation $prob = e^{B_0 + B_1x} / (1 + e^{B_0 + B_1x})$ where B_0 and B_1 are coefficients estimated from the data, x is the independent or predictor variable, and e is the base of the natural logarithms (Hosmer & Lemeshow, 1989; Montgomery & Peck, 1992). These generated probability values then may be used to classify the cases into the two different outcomes using a specific cutoff value. Cases with predicted values that equal or exceed the cutoff value are classified as positive, while those with predicted values smaller than the cutoff are classified as negative. One of the most common ways to assess how well a logistic regression model fits the data is to compare the model's predictions to the observed outcomes. Another way of assessing the goodness of fit of the model is to examine the Nagelkerke R Square value which is similar to the R^2 in a linear regression model, in that it quantifies the proportion of "variation" explained in the prediction model (Nagelkerke, 1991).

For each of the models, the dichotomous response variable (success versus nonsuccess in course) was regressed against the predictor variable (the CAT/2 scale-scores). Since the CAT/2 would be administered to students before they enrolled in a specific class, the comprehensive model was preferred to individual models for each class. However, before selecting the comprehensive model for further testing, its goodness of fit was compared to the others to determine if any of the individual models out performed the comprehensive model (see *Results*). Following this, the efficacy of the comprehensive model was tested by applying it to the CAT/2 scaled-scores from the year-two cohort. An initial cut-off value of 0.5 was used to classify whether students were predicted to succeed or not succeed in the course (i.e., students with CAT/2 scores that generated a probability value of ≥ 0.5 were predicted to successfully complete the course). Following completion of the course by the year-two cohort, the actual performance of each student was then compared to their predicted performance. The effects of making predictions using cut-off values other than 0.5 also were explored.

Results

Of the 71 students in the first cohort, only 39 successfully completed the courses, compared to 55 of the 82 students for the second-year cohort (Table 1). For both cohorts, the majority of the nonsuccessful students were noncompleters (Table 1).

Table 1: Success/NonSuccess Rates of Cohorts

	# of Successful Students	# of NonSuccessful Students		
		Failed	Did Not Complete	Total
Year-One Cohort	39 (55%)	2	30	32 (45%)
Year-Two Cohort	55 (67%)	4	23	27 (33%)

Logistic Regression Prediction Models

The goodness of fit for the comprehensive logistic regression model was much the same as that for each of the individual models (Table 2). However, the relationship between the response variable (success or nonsuccess) and the CAT/2 combined mathematics scale-scores was not significant for the Class Three model (Table 2).

Table 2: Logistic Regression Analysis of CAT/2 Achievement and Success versus NonSuccess for the Year-One Cohort^a

Model	Observed	n	Predicted			Overall % Correct	p-value	Nagelkerke R Square
			Success	Non-success	% Correct			
Class One	Success	10	6	4	60.0%	70.8%	0.010	0.323
	Nonsuccess	14	3	11	78.6%			
Class Two	Success	12	10	2	83.3%	75.0%	0.029	0.288
	Nonsuccess	8	3	5	62.5%			
Class Three	Success	17	15	2	88.2%	70.4%	0.079	0.148
	Nonsuccess	10	6	4	40.0%			
Comprehensive	Success	39	30	9	76.9%	70.4%	<0.001	0.260
	Nonsuccess	32	12	20	62.5%			

^a The cutoff value = 0.50

Testing the Comprehensive Logistic Regression Model with data from the year-two cohort.

When the CAT/2 scale-scores from the year-two cohort were entered into the equation for the comprehensive model

[$prob = e^{-9.613+0.017x} / (1 + e^{-9.613+0.017x})$] and the cutoff value was 0.5, approximately 73% (40/55) of “successful” students from the second year of the study were correctly predicted as being successful by the comprehensive model. At the same time, only 48% (13/27) of the “non-success” students were correctly predicted not to succeed. Combined, the model correctly predicted success/nonsuccess for approximately 65% of the students in the year-two cohort, using the 0.5 cut-off (Figure 1).

The comprehensive model’s overall ability to correctly predict success and non-success for the year-two students changed only slightly when the cutoff value was varied from 0.3 to 0.7 (Figure 1). However, more significant effects of altering the cutoff value were revealed when the predictions for successful and nonsuccessful students were considered separately. For example, when the cutoff value of 0.3 was used, the model correctly predicted success for nearly 90% of the successful students, but when the cutoff off value was increased to 0.7, the model only predicted success for 35% of the successful students (Figure 1). It was the reverse situation for predicting nonsuccess. At a cutoff value of 0.3, the model only correctly predicted nonsuccess for 22% of the nonsuccessful students, but at 0.7 the model’s accuracy for predicting nonsuccess was increased to 81% (Figure 1).

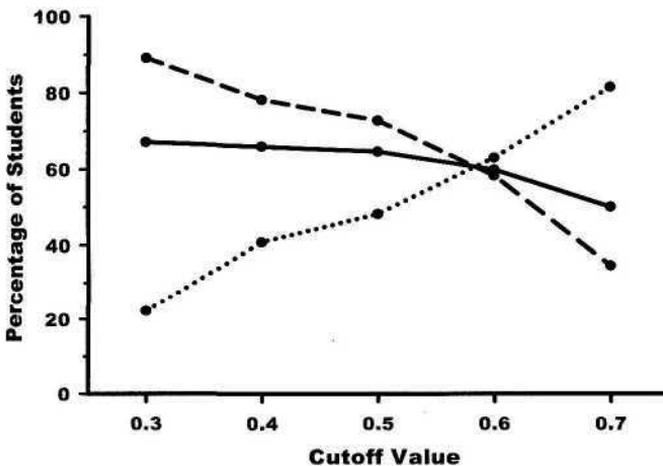


Figure 1: Impact of Varying the Cutoff Values on the Comprehensive Model’s Ability to Correctly Predict Success and Nonsuccess for the Year-Two Cohort

The solid line represents the percentage of students who were correctly predicted to succeed and not succeed ($n = 82$). The dashed line represents only the percentage of successful students who were correctly predicted to

succeed ($n = 55$), and the dotted line represents the corresponding percentage of unsuccessful students who were correctly predicted not to succeed ($n = 27$).

Discussion

Our study suggests that the eventual success or lack of success by students entering the ABE program is not easily predicted. Although initially the comprehensive logistic regression model appeared reasonably useful, it proved only moderately successful at predicting success and nonsuccess in the year-two cohort of students regardless of the cutoff value used.

Since the achievement test may be used as an entrance assessment tool, it is extremely important to consider the number of students who were successful but would have been excluded from this course based solely on their test scores. Of 82 year-two students, 15 students who were predicted to fail the course went on to successful completion when the cutoff score was 0.5. These results illustrate the major dilemma associated with using assessment tools as placement tests. Placement test scores may exclude those who are destined to fail, but at the expense of excluding a considerable portion of students who are capable of passing.

It is equally important to recognize the number of students who did not successfully complete the course, even though their test scores indicated success. With a cutoff value of 0.5, 14 out of the 82 students in the subsequent cohort who were predicted to succeed did not. These students seemed to have the academic abilities necessary to successfully complete the course, yet they did not do so. It is worth noting that 13 of these students were noncompleters. Therefore, it is possible that other factors not measured by the model may have contributed to their nonsuccess in the mathematics course.

Although adjusting the cutoff value has little impact on the overall predictive power of the model, it dramatically alters the model's ability to separately predict success versus nonsuccess. Specifically, as the cutoff value decreased, the model's ability to predict success improved, but its ability to predict nonsuccess diminished. Increasing the cutoff value had the opposite effect. These results are significant if you consider them in the context of using the model for determining placement. With a cutoff value of 0.3, the model would have admitted the vast majority of students destined to succeed in the course; however, it also would have admitted a significant number of students destined for nonsuccess. At the other end of the spectrum, with a cutoff value of 0.7, the model would have not admitted the majority of the

nonsuccessful students, but it also would have excluded more of the successful students. Deciding which cutoff value to adopt would depend upon the institution's preferred outcome. If the main purpose of the assessment is to ensure success then a lower cutoff value should be considered. If the purpose of the assessment is to limit nonsuccess, then a higher cutoff value would be a better option.

Conclusion

This study supports the premise that academic ability, as measured by achievement or aptitude test scores, is a credible predictor of student success and nonsuccess. However, given that 35.4% of the students in the year-two dataset would have been "misplaced" using the conventional cutoff value of 0.5, it is apparent that ABE cohorts can and do vary from year to year and that student success and nonsuccess in adult education endeavors is not easily predicted. Hence, to design a more accurate student success model that minimizes the variation within annual cohorts, we recommend implementing a longitudinal study that spans multiple years. The predictive ability of models calculated from the data collected during the early years could then be tested against subsequent cohorts of students and continually refined. We also recommend testing predictive models with a variety of cutoff scores to ascertain the value most suitable for the institution's goals. Until more accurate models are obtained, educational institutions should use the information from assessment testing prediction models primarily in terms of advising or as a part of a placement portfolio that would include other measurements such as motivation, maturity, financial security, level of interest, course load, and learning styles.

Additional research is necessary to support, refute or expand upon our conclusions. Specifically, there is a need to identify other factors impacting student success and nonsuccess—such as family obligations, employment status, financial and time constraints, and lack of confidence—and to incorporate those factors along with academic ability into a multinomial regression model. Also desirable would be comparison studies involving various statistical methods used to construct prediction models, such as discriminant analysis and logistic regression. Lastly, all of the assessment tools used by adult educators need to be investigated in terms of their predictive value and suitability for ABE programs and courses.

Acknowledgement: This paper was adapted from the senior author's Master's thesis (James, 2000). The research project was approved by the University of Victoria's Human Ethics Committee (project #359-98) and Thompson Rivers University's

12 James and Francis-Pelton, "Using Achievement Test Scores"

Human Ethics Committee (permit #98-99/07/A & #99-00/02/A). A sincere thank you to Dr. K. Larsen, Associate Professor, Thompson Rivers University for his suggestions and editing contributions.

References

- Brand, J. (1995). *Adult basic education: An assessment resource guide*. A report on Assessment Research in Literacy/Fundamental ABE programs in British Columbia. Victoria, BC: Ministry of Skills, Training and Labour.
- Canadian Test Centre. (1992). *Canadian Achievement Test, Second Edition, Technical Bulletin*. Markham, ON: CTC/Canadian Test Centre Inc.
- Crandall, D., Lerche, R., & Marchilonis, B. (1984). *Guidebook for effective literacy practice 1983-1984*. San Francisco, CA: National Adult Literacy Project, ED 253 776.
- Dirkx, J. M. & Jha, L. R. (1994). Completion and attrition in adult basic education: Test of two pragmatic prediction models. *Adult Education Quarterly*, 45 (1), 269-285.
- Ehringhaus, C. (1991). Testing in adult basic education. *Adult Basic Education*, 1(1), 12-26.
- Farr, C. W., Moon, C. E. & Williams, A. (1986). Correlating the test of adult basic education and the test of general education development. *Lifelong Learning*, 9, 17-19.
- Fragar, A. M. (1991). Adult literacy assessment: Existing tools and promising developments. *Journal of Reading*, 35 (3), 256-259.
- Gravenberg, E. V. & Rivers, J. H. (1987). Learning assistance programs. In L. Noel, R. Levitz, D. Saluri and Associates, (Eds.). *Increasing student retention* (pp. 264-282). San Francisco: Jossey-Bass.
- Grulick, L. E., (1987). *A study of various admission tests for use in the placement of entry-level students at Florence-Darlington Technical College*. Master's Thesis, University of South Carolina: ED 283 539.
- Hosmer, D. W. & Lemeshow, S. (1989). *Applied logistic regression*. Toronto: John Wiley & Sons, Inc.
- James, C. L. (2000). *Assessment testing as a predictor of student success in adult basic education mathematics courses*. Unpublished Masters Thesis, University of Victoria, BC.
- Montgomery, D. C. & Peck, E. A. (1992). *Introduction to linear regression analysis*. Toronto: John Wiley & Sons, Inc.
- Nagelkerke, M. J. D. (1991). A note on general definition of the coefficient of determination. *Biometrika*, 78: 691-692.
- Nurss, J. R. (1989). *Assessment models and instruments: Adult populations*. Adapted from a paper presented at the Assessment Issues Forum, Macon, GA: ED 313 572.

- Stricht, T. G. (1990). *Testing and assessment in adult basic education and English as a second language programs*. San Diego: Applied Behavioral and Cognitive Sciences, ED 317 867.
- Taylor, M. C. (1990). *Workplace literacy assessment tools*. Paper presented at the Annual Conference of the Canadian Association for the Study of Adult Education, Vancouver, BC: ED 330 885.
- Venezky, R. L., Bristow, P. S., & Sabatini, J. P. (1997). When less is more: Methods for placing students in adult literacy classes. *Adult Basic Education*, 7(1), 3-22.