The Student Persistence Index:
A New Measure of Student Persistence

by
Miles M. Hamby, *Ph.D.*

Background. Post-secondary institutions, and especially for-profit institutions, are particularly concerned with student persistence as it affects their profit and funding as well as staffing. Higher education, in general, is concerned about graduation rates and especially about time to graduation. Although time to graduation is a typical measure, it does not directly indicate how well the student is on track to graduate. The Student Persistence Index (SPI) offers a direct indication of whether the student is on schedule, ahead of schedule, or behind in completion of his or her program and can also be readily used in a multiple regression to predict persistence.

Common Measures of Persistence. Heretofore, the most common measure of persistence has been by cohort and measured according to whether or not a student enrolls in the subsequent term or the time it takes students in that cohort to complete a program. Some common definitions of persistence used in Institutional Research include:

- ". . . the percentage of all students enrolled … who enrolled in any course the following term."
- "Fall 2000 students who enrolled in Spring 2001 would constitute the 2000 rate . . ." (Peralta Community College, retrieved April 8, 2009, from http://www.peralta.edu/indev/factbook/SOPER.PDF)
- ". . . the intent to enroll in the same institution the following year." (Ziskin, et al, retrieved April 3, 2009, from http://pas.indiana.edu/cb/docs/ZiskinGrossHossler_ASHE06.pdf)
- Completion of the four-year degree as exemplified by "…many students who begin a four year institution fail to complete their four-year degree." (Tinto, Enhancing Student Persistence: Connecting the Dots, Syracuse, University, retrieved April 5, 2009, from http://aiea.syr.edu/vtinto/Files/Connecting%20the%20Dots.pdf.

The drawback to these forms of measure is that they do not really measure persistence. I define academic persistence as the active pursuit of a degree – the operable word being 'active'. This can also be viewed as enrollment behavior. These above measures are short term, or single circumstance, that is, they measure only whether the student enrolled in the next term, or the next school year, or simply completed. Simply completing a program is not really a measure of persistence – it's a 'yes' or 'no' measure and gives no indication of variation, or how actively the student pursued completion. Furthermore, these measures are usually tied to a cohort and do not account for other students enrolled in the same term but at different points in their program, or students who begin there program in an off-term (that is, winter, spring or summer.) But most significantly, they tell nothing about the nature of progress toward completion.

Conditions of Persistence. There are essentially three conditions of persistence a student can face: behind schedule, on schedule, or ahead of schedule. The schedule is defined by the institution's suggested curriculum, that is, the suggested number of credits taken in the suggested

number of terms. My new measure of student persistence behavior I call Hamby's Student Persistence Index (SPI), and is a measure of these three conditions.

The SPI. Essentially, the SPI is the product of the ratio of the number of credit hours achieved to the number of credit hours available (not to exceed total credit hours in the respective program) at the time and the ratio of the number of terms enrolled to the number of terms available at the time. Essentially, this measure becomes Credit-Terms. An SPI of 1 indicates the student is on schedule. An SPI of less than 1 indicates the student is behind schedule while an SPI of greater than 1 indicates the student is ahead of schedule.

Thus,

$$SPI = \frac{CAc}{CAv} \times \frac{TE}{TAv}$$

where CAc is the number of credits achieved, and CAv is the number of credits available at that point not to exceed total credit hours in the respective program, and where TE is the number of terms (quarter, semester, etc.) in which the student has enrolled, and TAv is the number of terms available at that point,.

The first term, ranges from 0 to greater than 1. The second term can range from 0 to 1. The following examples demonstrate the SPI.

> Example 1: Program is 120 credits and normally scheduled for 8 terms. the student has achieved 30 credits out of 60 identified as a normal load up to that point, resulting in a ratio of .5 and thus indicating the student is behind schedule in credits achieved to that point in the program. The student has also enrolled in 4 terms out of 4 possible up to that point, resulting in a ratio of 1 and thus indicating the student is on schedule in terms enrolled for that point in the program.

$$SPI = \frac{30}{60} \times \frac{4}{4} = .5$$

The resulting SPI, then, can range from 0 to greater than 1. In the above example, .5 x 1 = .5, thus indicating that, although the student has been continually enrolled, the student is behind schedule for completion, being short on credits.

> Example 2: Program is 8 terms and 120 credits. Student has enrolled in 8 out of 8 terms, but has completed only 80 of the 120 credits, and is therefore behind schedule for completion. The resulting SPI =.67 indicates student is behind schedule. The function indicates the delay is due to light load.

$$SPI = \frac{80}{120} \times \frac{8}{8} = .67$$

Example 3: Program is 8 terms and 120 credits. Student has enrolled in 6 out of 8 terms and has completed 120 of the 120 credits, and, although has completed the program (number of terms), completed it behind schedule (SPI = . 75), thus indicating erratic persistence. The function indicates this was due to breaks in enrollment.

$$SPI = \frac{120}{120} \times \frac{6}{8} = .75$$

Example 4: Program is 8 terms and 120 credits. Student has enrolled in 8 out of 8 terms and has completed 120 of the 120 credits. The student has completed the program on schedule (SPI = 1), indicating consistent persistence.

$$SPI = \frac{120}{120} \times \frac{8}{8} = 1$$

Example 5: Program is 8 terms and 120 credits. Student has enrolled in 4 out of 4 terms possible thus far and has completed 75 of the 60 credits normally scheduled by overloading.. The student is ahead of schedule (SPI = 1.25), indicating hyperactive persistence.

$$SPI = \frac{75}{60} \times \frac{4}{4} = 1.25$$

Example 6: Program is 8 terms and 120 credits. Student has enrolled in 6 out of 12 terms possible since initial enrollment and has completed 75 of the 120 credits normally scheduled to be achieved in 8 terms. The very low SPI (.325) suggests that not only is the student behind schedule, but also demonstrates erratic enrollment behavior and low course loading.

$$SPI = \frac{75}{120} \times \frac{6}{12} = .325$$

Example 7: Program is 8 terms and 120 credits. Student has enrolled in 10 out of 14 terms possible since initial enrollment and has completed 75 of the 120 credits normally scheduled to be achieved in 8 terms. The low SPI (.446) indicates the student is behind schedule. However, the slightly higher SPI of .446, when compared to .325 from Example 5, indicates that this student's enrollment behavior is less erratic (or more consistent), and, therefore, course loading is more even.

$$SPI = \frac{75}{120} \times \frac{10}{14} = .446$$

Advantages. A major advantage of the SPI is that it is free of cohorts. All students may be included in the measure of persistence for any desired time period, independent of any cohort or point of entry into a program.

The real usefulness of the SPI is as intimated earlier – the variation that occurs in enrollment behavior, i.e., persistence. Completion is either a 'yes' or 'no' measure, and is thus a nominal variable that is limited in its treatment, usually in a logistic regression or ANOVA. It can only predict the likelihood of completing a program, not indicate persistence behavior. The SPI is a continuous variable and can thus be used in a multiple regression offering prediction of how well the student adheres to a normal schedule of completion, that is, whether the student is likely to be ahead of schedule, on schedule, or behind schedule.

Factoring in Completion. The SPI, purposely, does not attempt to factor in completion as completion would also be a dependent variable. Once a student has completed, the case is closed, so to speak. That is, the student's case enters the past tense and prediction is irrelevant. However, it would be useful to learn how well the student persisted in achieving completion and what variables accounted for this. In this case, the SPI becomes a meaningful dependent variable in a regression when the sample is restricted to only those who have completed. Simply using the dichotomous 'yes' or 'no' variable of completion in a logistic regression, t-test, or ANOVA, gives only whether there was a difference, but reveals little about the strength of the effect each independent variable had on completion. By selecting a sample of only those who have completed and running a regression with the SPI as dependent variable, the result shows what variables have an effect on actual persistence, or how actively the student pursued completion, not just simply whether or not the student completed.

However, SPI can be related to completion to reveal what effect persistence (or the degree of active pursuit) had on the odds, or probability, of completion. In this context, the SPI score would be used as an independent, continuous variable regressed against the dichotomous nominal variable of completion, thus revealing how significant a factor adherence to schedule was in completing. The regression equation, then, would be something like:

$$\text{Log-odds of COMPLETION} = \text{Constant} + (B_{SPI} * \text{SPI})$$

where Log-odds of COMPLETION represents the log-odds of completion (i.e., the odds of scoring a '1', not the actual score) of the student's respective program, Constant = the log-odds of completion without regard to SPI, and B_{SPI} = the log-odds of increase (or decrease if negative) away from '0' odds, given the respective student's SPI.

> Example 8: Program is 8 terms and 120 credits. 100 students completed over various lengths of times and 250 students have not completed, all beginning their programs at different terms. Each of the 350 student SPIs would be entered into a logistic (LOGIT) regression as the dependent variable and their respective COMPLETION code ('0' or '1') entered as the dependent variable. The resulting statistic could take several forms depending on the procedure used, but in essence would be the effect that a unit increase in SPI would have on the probability of the respective student receiving a '1'for COMPLETION. For example, given the following LOGIT regression result table:

Variable	B	Std. Error	Wald	df	Sig.
Constant	2.100	.988	23.789	0	.004
SPI	5.804	.332	5.872	1	.015

The log-odds of completion (without regard to SPI) would be 2.1, but with each unit increase in SPI (i.e., an SPI of 1.1 represents 1.1 units), a student's log-odds of completion would increase .804. Thus, the log-odds of completion for a student with an SPI of .750 would be:

$$\text{Log-odds of COMPLETION} = \text{Constant} + (B_{SPI} * \text{SPI})$$

$$\text{Log-odds of COMPLETION} = 2.1 + (5.804 * .950)$$

$$6.453 = 2.1 + (5.804 * .750)$$

Compared to a student whose SPI is 1.1, then

$$8.484 = 2.1 + (5.804 * 1.1)$$

Thus, the log-odds of this student for completing has increased 2.03 (8.484 – 6.453) beyond the student with an SPI of .750. Keep in mind that the log-odds form of the LOGIT statistic is a logarithmic curve and not a linear expression of direct probability.

Conclusion and Further Study. Perhaps the greatest benefit to using the SPI as a dependent variable is in how persistence can be measured to suggest specific intervention. As SPI is measured over time, indices, or benchmarks, begin to form for that institution that would suggest specific interventions. For example, an institution might find that students with an SPI of more than .7 eventually complete on their own after six years, and students with an SPI or .4 to .7 benefit from intervention, and intervention for students with an SPI of less than .4 has no effect and never complete. Knowing this could help the institution allocate resources more efficiently. With these results, further research should be done on linking SPI to the effectiveness of specific intervention programs at other institutions of higher education. Other college and university officials should replicate this study to establish their own benchmarks for the SPI.

References

Peralta Community College (2009), retrieved April 8, 2009, from http://www.peralta.edu/indev/factbook/SOPER.PDF

SPSS Data Analysis Examples: Logit Regression (2010), retrieved from http://www.ats.ucla.edu/stat/spss/dae/logit.htm, August 2009.

Tinto (2009). Enhancing Student Persistence: Connecting the Dots, Syracuse, University, retrieved April 5, 2009, from http://aiea.syr.edu/vtinto/Files/Connecting%20the%20Dots.pdf.

Ziskin, et al, retrieved April 3, 2009, retrieved April 8, 2009, from http://pas.indiana.edu/cb/docs/ZiskinGrossHossler_ASHE06.pdf

