Association for Institutional Research

PROFESSIONAL FILES | SPRING 2013 VOLUME

Supporting quality data and decisions for higher education.

Letter from the Editor

AIR is pleased to reintroduce *Professional Files*. This volume contains three articles recommended for inclusion by AIR members through a blind peer-review process. Each article puts forth ideas and analysis for our consideration as institutional research professionals. Most salient for me in my role as volunteer coordinating editor is that the articles frame the work of IR as more than pure technical analysis through acknowledgment of the importance of context.

Association for Institutional Research

Zan, Yoon, Khasawneh, and Srihari unpack differences in enrollment projections, with a particular focus on the influence of the unit of analysis on the appropriateness of the model. Saupe and Eimers examine the under-appreciated area of grade point average reliability; something that will be of greater importance as policy dialogues move to multiple measures of college readiness and adjusted metrics of student outcomes. Crisp, Palacios, and Kaulfus empower IR professions in efforts to better understand student engagement efforts on campus.

It has been an honor to be a part of this volume of Professional Files and to strive to elevate the work of peers for the benefit of the field. We look forward to receiving your submission next.

Sincerely,

Christopher M. Mullin

IN THIS ISSUE...

EDITORS

ISSN 2155-7535

PROFESSIONAL FILE

ARTICLE 129

A COMPARISON STUDY OF RETURN RATIO–BASED ACADEMIC ENROLLMENT FORECASTING MODELS

Xinxing Anna Zan, Sang Won Yoon, Mohammad Khasawneh, Krishnaswami Srihari

ACKNOWLEDGEMENTS

This study was supported by the Watson Institute for Systems Excellence at the State University of New York at Binghamton. The authors wish to thank Sandra Starke, Tania Das, and Jill Heneghan from the Institutional Research and Assessment Office for providing us with all the data for this research. We also thank anonymous reviewers who have given valuable comments to improve this paper.

About the Authors

Xinxing Anna Zan is a master of science student, Sang Won Yoon is assistant professor, Mohammad Khasawneh is associate professor, and Krishnaswami Srihari is distinguished professor. All are with the Department of Systems Science and Industrial Engineering, State University of New York at Binghamton, Binghamton, NY 13902.

The corresponding author, Sang Won Yoon, can be reached at the following: Phone: +1-607-777-5935; Fax: +1-607-777-4094, Email: yoons@binghamton.edu.

Abstract

In an effort to develop a low-cost and user-friendly forecasting model to minimize forecasting error, we have applied average and exponentially weighted return ratios to project undergraduate student enrollment. We tested the proposed forecasting models with different sets of historical enrollment data, such as university-, school-, and division-level enrollment. The numerical results indicate that the proposed models perform better when the school-level and university-level analyses are applied, as compared to the division-level analysis. We also observed that the forecasting error is lower when the most recent enrollment data sets are considered than when we consider all past enrollment data. In addition, when forecasting for spring semesters, the 1-year average return ratio method, using the school-level analysis, yields the lowest forecasting error of 0.40%. When forecasting for fall semesters, the average return ratio method, using the university-level analysis, yields the lowest forecasting error of 0.81%.

INTRODUCTION

Undergraduate student enrollment patterns require an accurate forecasting model to assist in college and university strategic planning efforts. Forecasting is the projection, estimation, or prediction of future event occurrences that are uncertain in nature (Tersine, 1994). Accurate forecasting can help people plan wisely for an organization's future. Over the years, many forecasting models and techniques have been applied in business organizations, government agencies, educational systems, and public services (Armstrong, 2001). However, it is necessary to use suitable and accurate forecasting techniques. As a result, identifying the best enrollment forecasting model for a college or university is critical to effective decision making; such a forecasting model relates to the following services from a strategic planning perspective (Desjardins et al., 2006; Glover, 1986; Hossler & Bean, 1990; Norton, 1998):

- Improve the accuracy of student enrollment and income or budget forecasts

- Given the limitations of existing resources, offer high-quality academic programs as well as great campus experiences to meet students' needs
- Project campus housing assignments, necessary building or classroom planning, staffing allocation, and course scheduling
- Plan for total budgeting and allocate or balance income and expense at realistic levels for both academic values and student enrollment demands
- Seek ways to describe, analyze, predict, and improve student retention percentages
- Improve student and staff satisfaction
- Link and coordinate activities of recruitment, admissions, financial aid, and career planning

Therefore, the objective of this research is to develop a low-cost forecasting model to minimize forecasting error and to provide a less computationally intensive method that can be used to forecast undergraduate student enrollment. The research questions include the following:

1. How can we accurately forecast total undergraduate student enrollment in a college or university at a specific semester to reduce forecasting error?
2. How can we minimize the forecasting error when projecting undergraduate student enrollment (i.e., division-, school-, or university-level analyses)?
3. When projecting undergraduate student enrollment, should we apply all available past enrollment data or the most recent enrollment data for better forecasting accuracy?

One limitation of this paper is that our research does not consider the projection of freshman student enrollment, since it is typically given. Also, we make several forecasting assumptions to project student enrollment and improve forecasting accuracy.

The remainder of the paper is organized as follows. First, we present the background associated with this research. Then, we address the detailed research methodologies. Next, we explain and discuss numerical results of forecasting models. Finally, we summarize conclusions and opportunities for future work.

RESEARCH BACKGROUND

Many approaches and methods have been studied and proposed to forecast student enrollment, with each forecasting model generating different forecasting errors (Guo, 2002). Among other purposes, probability forecasting methods are used to calculate probabilities of an uncertain event happening in the future. A linear probability model was proposed where transition probabilities are used to calculate student enrollment (Marshall & Oliver, 1970). In their model, Marshall and Oliver considered students' total work to be done, based on probabilities of students attending, vacationing or interning, and dropping out. Marshall and Oliver applied this to forecast student enrollment at the University of California, Berkeley. Similar to the linear probability model, logit or probit models are used occasionally to forecast student enrollment when the outcome or dependent variables are known. However, logit or probit models are usually used to analyze more-complex educational behaviors (Porter, 1999).

Another category of probability forecasting methods is known as the "ratio forecasting method." For example, student retention rate has been utilized to forecast campus student enrollment for the University of Wisconsin-Madison (Beck, 2009). His results showed that the students' retention rate generally decreases as they approach their graduation year. Their results also showed that the decline of the students' retention rate decreased dramatically from 4–5 school years to 5–6 school years, largely due to graduation. We draw several important features from the ratio analysis (Beck, 2009):

- Models should be analyzed separately, especially when the retention ratios or percentages are significantly different.
- Retention patterns of students should be examined carefully based on historical data points and previous student enrollment data.
- Models should be tested against recent or past actual enrollment data.

The ratio method has also been applied to forecast student enrollment at the University of Washington (Schmid & Shanley, 1952) using a three-step procedure. First, they derived a series of estimates for the entire population for which all or at least a major component of the student enrollment is drawn. Therefore, they took the data for this category directly from the population forecast from the State of Washington. Second, it was important to learn the enrollment trends for all institutions of higher education in the State of Washington as a whole. This was necessary to determine the relationship between student enrollment and the age group between 18 to 21 years in the entire population during the past 30 years and during the next 10 years. Third, they determined the trend in the ratio of student enrollment to the total population of the age group between 18 and 21 years for the forecasting period. Based on detailed analysis of the historical data, they calculated the ratio between student enrollment and total population age group, and utilize that ratio for a future student enrollment projection. As a result, the proposed ratio method showed several advantages:

- Institutional researchers need expend less time and labor in performing student enrollment forecasts.
- Institutional researchers are able to use historical data to forecast.
- Institutional researchers do not need to define parameters or variables.

Furthermore, the exponential weighted moving average method is one type of moving average method that does not require a large amount of historical data points or records (Dobbs, 2001). The exponential weighted moving average methods vary, depending on single or multiple exponential smoothing approaches. The idea of the exponential weighted moving average method is that it smooths out variations in a time-series model by applying more weights on the more-recent data than on previous data (Tersine, 1994). Exponential weight forecasting methods have been used widely in operations research and economics (Muth, 1960). Many researchers used this forecasting method to predict short-term sales in inventory control (Brown, 1959; Magee, 1958). Nowadays, many other fields have started using the exponential weighted average forecasting method to predict different forecasts, because they have seen the success of this method in forecasting sales.

For instance, the double exponential smoothing method was studied as a pattern-based method to apply and adapt to a number of circumstances (Gardner, 1981). In this research, Gardner compared different forecasting methods, including correlation analysis, intention survey, and professional judgment methods to predict student enrollment where double exponential smoothing has the most reasonable and most consistent results. According to the literature (Dobbs, 2001; Holt, 2004; Snyder, 1988), the exponential weighted moving average has the following forecasting advantages:

- It includes all previous data to represent the entire history of data.
- It is easy to compute and provides better forecasting results for short-term projections.
- It does not require a large amount of historical data to implement.
- It provides flexibility in forecasting with seasonal behaviors and trends.
- In this research, therefore, the probability or ratio forecasting method, combined with the exponential weighted moving average method, is used in predicting university student enrollment.

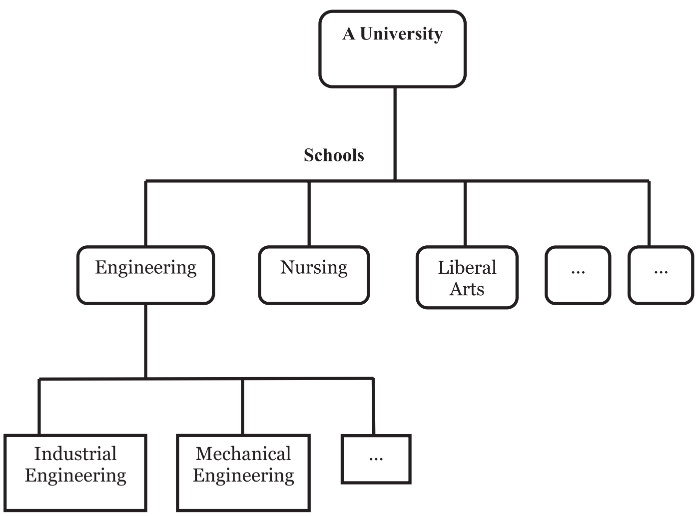
RESEARCH METHODOLOGIES

Different forecasting models are developed to forecast the number of undergraduate student enrollment; we exclude first-semester freshman students in this research. Based on a detailed comparison and the analysis of various forecasting models, this research endeavor is expected to identify the most suitable forecasting model for the projection of undergraduate student enrollment at a university. We cannot assume this method will be best for all institutions, but put it forth for consideration by the reader. It is important to highlight the principal model-related assumptions made in this research:

1. Student enrollment patterns that have happened in the past are considered likely to occur in the future for enrollment forecasting purposes.
2. Available historical data for analysis can be assumed to represent the entire historical pattern that can be used to predict future student enrollment.
3. Student return ratios are different from fall to spring semesters.

Forecasting undergraduate student enrollment is primarily based on historical data. Three levels of analysis—university, school, and division levels—are analyzed to forecast undergraduate student enrollment each semester, including university analyses. In this research, a university is assumed to consist of several schools (e.g., engineering or nursing), each having multiple divisions (e.g., industrial engineering or mechanical engineering) as shown in Figure 1.

Figure 1. Overview of a University, Schools, and Divisions



University-level analysis forecasts student enrollment based on return ratio calculations of total student enrollment each semester, which is similar to many proposed methods in the literature that have shown high accuracy in projecting student enrollment as discussed in the previous section of the paper. School- and division-level analyses, on the other hand, have not been mentioned in the literature we reviewed since these are the two new units of analysis proposed in this research. School-level analysis forecasts student enrollment based on return ratio calculations of each school's student enrollment to calculate total student enrollment each semester. Division-level analysis forecasts student enrollment, based on return ratio calculations of each division's student enrollment, to calculate total student enrollment each semester.

All three forecasting models used two types of return ratio (RR) calculations: average return ratio (ARR) and exponential weighted return ratio (EWRR). We based all calculations for forecasting a campus's student enrollment on calculating the ratio between student enrollment in a specific semester and the previous semester. The mathematical equation of each RR can be expressed as

$$RR_j = \frac{E_t}{E_{t-1}} \tag{1}$$

where E_t is the undergraduate student enrollment at semester t.

University-Level Analysis

University-level analysis forecasts undergraduate student enrollment, based on ARR and EWRR calculations of a university's student enrollment each semester from the previous semester. When calculating ARR and EWRR, fall semesters and spring semesters are separated to increase the forecasting accuracy. The ARR of university-level analysis is expressed as

$$ARR_t = \frac{1}{n}\sum_{t=1}^{n} RR_t \tag{2}$$

where n is the total number of semesters. The EWRR of university-level analysis can also be calculated by

$$EWRR_t = \alpha[RR_{t-1} + (1-\alpha)RR_{t-2} + (1-\alpha)^2 RR_{t-3} + \ldots] + (1-\alpha)^{t-1} RR_1 \tag{3}$$

where α is a weighting factor ($0 < \alpha < 1$); as α increases, more weight is given to recent data. Based on Equations (1) and (2), we can obtain undergraduate student enrollment projections, using university-level analysis, by

$$E_t = ARR_t \times E_{t-1} \tag{4}$$

$$E_t = EWRR_t \times E_{t-1} \tag{5}$$

School-Level Analysis

It is possible that university-level analysis may not capture different students' retention rates in different schools. As a result, we propose school-level forecasting analysis to (potentially) increase the forecasting accuracy. In this analysis, we project undergraduate student enrollment, based on each school's student enrollment, and calculate the total student enrollment, considering the differences between the fall and spring semesters' student RR. The ARR of school-level analysis is calculated by

$$ARR_{t,j} = \frac{1}{n}\sum_{t=1}^{n} RR_t \tag{6}$$

where $E_{t,j}$ represents the undergraduate student enrollment in school j at semester t. The EWRR for the school-level analysis is calculated by

$$EWRR_{t,j} = \alpha[RR_{t-1} + (1-\alpha)RR_{t-2} + (1-\alpha)^2 RR_{t-3} + \ldots] + (1-\alpha)^{t-1} RR_1 \,. \tag{7}$$

Therefore, the total student enrollment, using school-level analysis, can be obtained by

$$E_{t,j} = ARR_{t,j} \times E_{t-1,j} \tag{8}$$

$$E_{t,j} = EWRR_{t,j} \times E_{t-1,j} \,. \tag{9}$$

Division-Level Analysis

In this approach, different schools are broken down into different divisions to see if forecasting accuracy can be further improved. After we obtain student enrollment from each division, we can calculate the total student enrollment. When calculating ARR and EWRR, we also separate fall semesters and spring semesters. The ARR for the division-level analysis is calculated by

$$ARR_{t,k} = \frac{1}{n}\sum_{t=1}^{n} RR_t \tag{10}$$

where $E_{t,k}$ is the undergraduate student enrollment in division k at semester t. The EWRR for division-level analysis is calculated by

$$EWRR_{t,k} = \alpha[RR_{t-1} + (1-\alpha)RR_{t-2} + (1-\alpha)^2 RR_{t-3} + \ldots] + (1-\alpha)^{t-1} RR_1 \,. \tag{11}$$

Then, the total student enrollment can be obtained as follows:

$$E_{t,k} = ARR_{t,k} \times E_{t-1,k} \tag{12}$$

$$E_{t,k} = EWRR_{t,k} \times E_{t-1,k} \,. \tag{13}$$

In this study, the forecasting errors are measured to compare the performance of different forecasting models, and the forecasting error, ε (%), is defined as

$$\varepsilon = \frac{E^A_{total} - E^P_{total}}{E^A_{total}} \times 100 \tag{14}$$

where E^A_{total} and E^P_{total} are the total numbers of actual and projected student enrollment, respectively. An example of forecasting analysis is given below.

Numerical Results: Forecasting Model Illustration

Table 1 shows sample data from the fall and spring semesters of 2000 to 2010 at the State University of New York (SUNY) at Binghamton, where Y_{00f} represents the fall semester of 2000, Y_{01s} represents the spring semester of 2001, s_1 represents first semester, s_2 represents second semester, and so on. Undergraduate students typically take 4 years (or eight semesters) to graduate; there are some students, however, who take more than 4 years, which is why the data include student enrollment for up to 6 years.

Table 1. Sample Data of Undergraduate Student Enrollment (2000–2010)

Academic Year	Semester					
	s_1	s_2	…	s_{11}	s_{12}	Total
Y_{00f}	879					
Y_{01s}	95	830				
⋮			⋮			⋮
Y_{10s}	135	900	...	7	8	3,690
Y_{10f}	1,089	124	...	9	2	3,796

To calculate the forecasting error, using the university-level analysis, the RR should be calculated by Equation (1), as shown in Table 2. For instance, the RR of first-semester students becoming second-semester students from the fall semester of 2000 to the spring semester of 2001 is calculated as

$$RR_{Y_{00f},Y_{01s}}(s_{1,2}) = \frac{E_2(Y_{01s})}{E_1(Y_{00f})} = \frac{830}{879} = 0.94. \tag{15}$$

The RR of the 11th-semester students becoming 12th-semester students from the spring semester of 2010 to the fall semester of 2010 is calculated as

$$RR_{Y_{108s},Y_{10f}}(s_{11,12}) = \frac{E_8(Y_{10f})}{E_7(Y_{10s})} = \frac{2}{7} = 0.29. \tag{16}$$

Table 2. Summary of RRs for Each Semester

Academic Year	Semester				
	s_1	s_2	…	s_{11}	s_{12}
$Y_{00f,01s}$	0.94				
$Y_{01s,01f}$	0.85	0.91			
⋮			⋮		
$Y_{08f,09s}$	0.96	1.03	...	1.29	0.42
$Y_{09s,09f}$	0.98	0.98	...	0.22	0.56

After all the RRs are calculated, we can apply the two forecasting methods, ARR and EWRR. The ARRs between each semester are calculated by Equation (2) as shown in Table 3, where $ARR_{F,S}$ represents the ARR from fall to spring semesters and $ARR_{S,F}$ represents the ARR from spring to fall semesters. For instance, the $ARR_{F,S}(s_{1,2})$ is equal to the average of all the RRs from fall to spring semesters in the $s_{1,2}$ column in Table 2.

$$ARR_{F,S}(s_{1,2}) = \frac{1}{9}(0.94 + 0.95 + ... + 0.96) = 0.96. \qquad (17)$$

The $ARR_{S,F}(s_{5,6})$ is equal to the average of all the RRs from spring to fall semesters between the fifth and the sixth semesters.

$$ARR_{S,F}(s_{5,6}) = \frac{1}{7}(0.67 + 0.91 + ... + 0.89) = 0.83. \qquad (18)$$

Based on ARRs, student enrollment can be projected for Y_{10s} and Y_{10f} using Equation (4), as shown in Table 4. When projecting for Y_{10s}, we use the $ARR_{F,S}$ since it projects for a spring semester. When projecting for Y_{10f}, however, we use the $ARR_{S,F}$ since it projects for a fall semester. For instance, the 2nd- and 10th-semesters' student enrollment at Y_{10s} and Y_{10f} can be expressed as

$$E_{10s}(s_2) = E_{09f}(s_1) \times ARR_{F,S}(s_{1,2}) = 930 \times 0.96 = 892.8 \qquad (19)$$

$$E_{10f}(s_{10}) = E_{10s}(s_9) \times ARR_{S,F}(s_{9,10}) = 21 \times 0.32 = 7. \qquad (20)$$

We calculate the forecasting errors projecting for the spring and fall semesters of 2010, using university-level analysis with the ARR, by Equation (14) as shown in Table 5. The detailed calculations are illustrated as

$$\varepsilon(Y_{10s}) = \frac{E^A_{total}(Y_{10s}) - E^P_{total}(Y_{10s})}{E^A_{total}(Y_{10s})} \times 100 = \frac{3555 - 3532}{3555} \times 100 = 0.66\% \qquad (21)$$

$$\varepsilon(Y_{10f}) = \frac{E^A_{total}(Y_{10f}) - E^P_{total}(Y_{10f})}{E^A_{total}(Y_{10f})} \times 100 = \frac{2707 - 2718}{2707} \times 100 = -0.37\% . \qquad (22)$$

Table 5. Summary of ε for Y_{10s} and Y_{10f} Using ARR

	E^A_{total}	E^P_{total}	ε
Y_{10s}	3555	3532	0.66
Y_{10f}	2707	2718	–0.37

Table 3. Summary of ARRs for Each Semester

	$s_{1,2}$	$s_{2,3}$	$s_{3,4}$	$s_{4,5}$	$s_{5,6}$	$s_{6,7}$	$s_{7,8}$	$s_{8,9}$	$s_{9,10}$	$s_{10,11}$	$s_{11,12}$
$ARR_{F,S}$	0.96	0.99	0.96	0.93	0.94	0.97	0.93	0.83	0.63	0.94	0.61
$ARR_{S,F}$	0.90	0.91	0.91	0.87	0.83	0.85	0.36	0.16	0.32	0.20	0.45

Table 4. Summary of $E^P(Y_{10s})$ and $E^P(Y_{10f})$ Using ARR

	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	s_{11}	s_{12}
$E^P(Y_{10s})$	890	107	905	126	730	109	580	21	51	6	7
$E^P(Y_{10f})$	121	819	104	792	111	623	37	91	7	10	3

On the other hand, we calculate the EWRR between each semester using Equation (3) based on the results in Table 2. The calculated EWRRs are shown in Table 6, where $EWRR_{F,S}$ represents the EWRR from fall to spring semesters, and the $EWRR_{S,F}$ represents the EWRR from spring to fall semesters. Assuming that α is set to 0.5 for the purpose of this research, then we can calculate $EWRR_{F,S}$ and $EWRR_{S,F}$ as

$$EWRR_{F,S}(s_{3,4}) = 0.5\times[0.99+(1-0.5)0.96+(1-0.5)^2 0.95+(1-0.5)^3 0.95+ (1-0.5)^4 0.96+(1-0.5)^5 0.96+(1-0.5)^6 0.95]+(1-0.5)^7 0.96 \quad (23)$$

$$EWRR_{S,F}(s_{9,10}) = 0.5\times[0.34+(1-0.5)0.24+(1-0.5)^2 0.33+(1-0.5)^3 0.36+ (1-0.5)^4 0.34 = 0.31. \quad (24)$$

We can now project student enrollment using the EWRR for Y_{10s} and Y_{10f} using Equation (5), as shown in Table 7. When projecting for Y_{10s} , we use the $EWRR_{F,S}$ since it projects for a spring semester. When projecting for Y_{10f} , however, we use the $EWRR_{S,F}$ since it projects for a fall semester, as illustrated in the following examples:

$$E_{10s}(s_3) = E_{09f}(s_2)\times EWRR_{F,S}(s_{2,3}) = 108\times 1.01 = 109 \quad (25)$$

$$E_{10f}(s_8) = E_{10s}(s_7)\times EWRR_{S,F}(s_{7,8}) = 103\times 0.31 = 32\ . \quad (26)$$

We also can calculate the forecasting errors, using university-level analysis with the EWRR, as shown in Table 8. The detailed calculations are illustrated as

$$\varepsilon(Y_{10s}) = \frac{E^A_{total}\ (Y_{10s}) - E^P_{total}\ (Y_{10s})}{E^A_{total}\ (Y_{10s})}\times 100 = \frac{3555-3543}{3555}\times 100 = 0.35\% \quad (27)$$

$$\varepsilon(Y_{10f}) = \frac{E^A_{total}\ (Y_{10f}) - E^P_{total}\ (Y_{10f})}{E^A_{total}\ (Y_{10f})}\times 100 = \frac{2707-2724}{2707}\times 100 = -0.69\%\ . \quad (28)$$

Table 8. Summary of ε for Y_{10s} and Y_{10f} Using EWRR

	E^A_t	E^P_t	ε
Y_{10s}	3555	3543	0.35
Y_{10f}	2707	2724	–0.69

We illustrate the calculation steps through an example using university-level analysis, which forecasts undergraduate student enrollment based on the ARR and EWRR calculations of a university's student enrollment each semester from the previous semester. School- and division-level analyses, on the other hand, forecast undergraduate student enrollment, based on the ARR and EWRR calculations of different schools' or different divisions' student enrollment. The detailed calculation steps are very similar for school- and division-level analyses compared to university-level analysis. The only difference is that after we obtain student enrollment at each school or each division, we calculate the total student enrollment at a university by taking the summation of all schools' or divisions' student enrollment, after which we compare the resulting E^P_{total} with E^A_{total} to obtain ε .

To find the most accurate forecasting model to project undergraduate student enrollment at a university, we compare the forecasting error, ε , for the three forecasting models—university-, school-, and division-level analyses—and the forecasting model with the lowest ε value should be selected. As shown in Table 9, we compare ε value of each forecasting model combined with its corresponding method. It is evident that university- and school-level analyses yield lower ε values than division-level analysis when forecasting for undergraduate student enrollment (excluding first-semester freshman students) for the spring and fall semesters of 2010.

Table 6. Summary of EWRRs for Each Semester

	$s_{1,2}$	$s_{2,3}$	$s_{3,4}$	$s_{4,5}$	$s_{5,6}$	$s_{6,7}$	$s_{7,8}$	$s_{8,9}$	$s_{9,10}$	$s_{10,11}$	$s_{11,12}$
$EWRR_{F,S}$	0.96	1.01	0.96	0.94	0.94	0.96	0.93	0.84	0.64	1.06	0.57
$EWRR_{S,F}$	0.95	0.91	0.95	0.88	0.87	0.85	0.31	0.16	0.31	0.21	0.45

Table 7. Summary of $E^P(Y_{10s})$ and $E^P(Y_{10f})$ Using EWRR

	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	s_{11}	s_{12}
$E^P(Y_{10s})$	894	109	910	126	730	108	579	21	52	7	7
$E^P(Y_{10f})$	128	816	108	793	116	620	32	91	7	10	3

Table 9. Comparison of ε Values of Y_{10s} and Y_{10f}

	University-Level		School-Level		Division-Level	
	ARR	EWRR	ARR	EWRR	ARR	EWRR
Y_{10s}	0.66	0.35	0.46	0.38	–1.94	–0.87
Y_{10f}	–0.38	–0.69	–0.83	–1.13	–3.44	–1.10

Table 10. Comparison of ε Values of Y_{10s} to Y_{10f}

	University-Level		School-Level	
	ARR	EWRR	ARR	EWRR
Y_{07s}	–0.82	–0.76	–1.17	–1.02
Y_{07f}	–1.64	–1.67	–1.97	–1.70
Y_{08s}	0.23	0.55	–0.13	0.33
Y_{08f}	–0.33	0.12	–0.86	–0.22
Y_{09s}	1.02	1.09	0.60	0.84
Y_{09f}	1.25	1.60	0.93	1.48

To determine whether university- and school-level analyses will work in forecasting undergraduate student enrollment during special periods or sudden events (e.g., the economic downturn in 2008), we also analyze the forecasting results from the spring semester of 2007 to the fall semester of 2009 to validate their ε values, as shown in Table 10. It is clear that the forecasting models also work relatively well in forecasting undergraduate student enrollment during such events.

In this research, the forecasting models are mainly developed by the ARR and EWRR based on all available data prior to the projected semester. To further determine whether university- or school-level analysis is better, we will use only the most recent data points for analysis. We compare different ε values based on 1 year (y^1), 2 years (y^2), 3 years (y^3), and 4 years (y^4) of the ARRs, which means that we calculate the ARRs based on using only the most recent years of enrollment data for calculations. Since the calculations are now based on a small numbers of years, the EWRR method will not differ significantly from the ARR; therefore, we will use only the ARR for the analysis. By using the same calculation steps, we show the results for ε values using university-level analysis in Table 11 and Figure 2.

Table 11. ε Values When Using Different Years of ARR by University-Level Analysis for Projecting Y_{10s} and Y_{10f}

	y_1	y_2	y_3	y_4
Y_{10s}	–0.20	0.23	0.68	0.76
Y_{10f}	–1.55	–0.69	–0.12	–0.29

Figure 2. ε Values When Using Different Years of ARR by University-Level Analysis for Projecting Y_{10s} and Y_{10f}

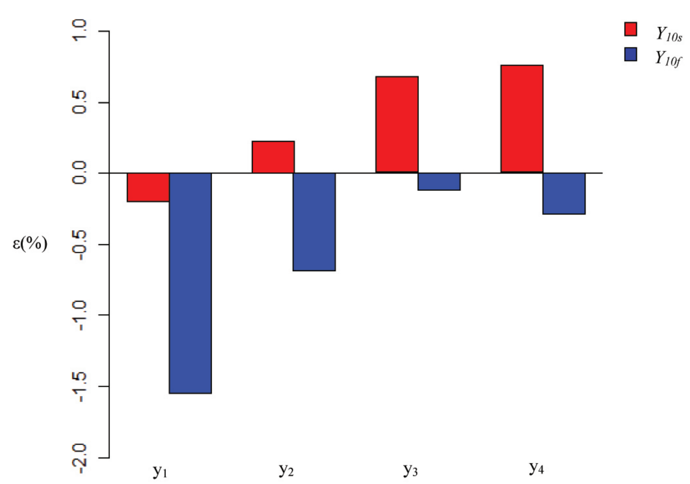


Table 12. ε Values of Using Different Years of ARR by School-Level Analysis for Projecting Y_{10s} and Y_{10f}

	y_1	y_2	y_3	y_4
Y_{10s}	–0.08	0.29	0.76	0.79
Y_{10f}	–1.67	–1.08	–0.49	–0.55

Figure 3. ε Values When Using Different Years of ARR by School-Level Analysis for Projecting Y_{10s} and Y_{10f}

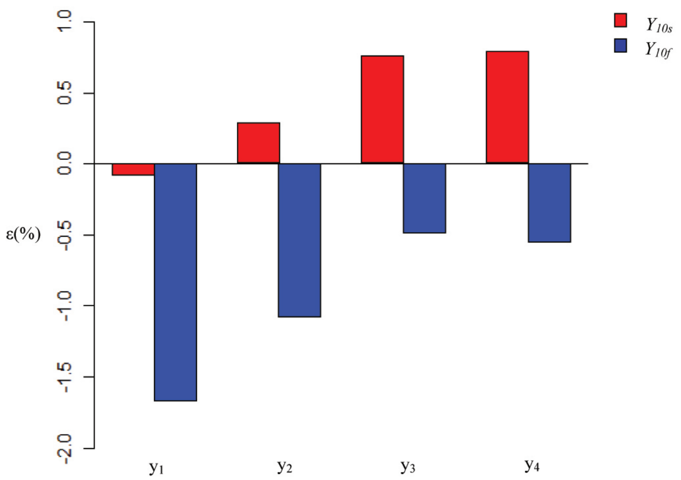


Figure 4. ε Value Comparison for Projecting Y_{10s} and Y_{10f}

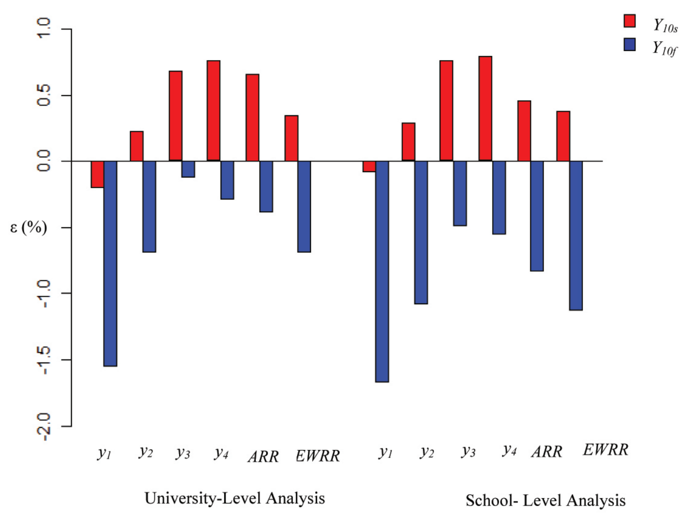


Table 13. ε Value Comparison Using ARR and EWRR for Projecting Y_{09s} and Y_{09f}

	University-Level		School-Level	
	ARR	EWRR	ARR	EWRR
Y_{09s}	1.02	1.09	0.60	0.84
Y_{09f}	1.25	1.60	0.93	1.49

By using the same calculation steps, we show the results for ε values using school-level analysis in Table 12 and Figure 3. The results indicate that when the average number of years used increases, $\varepsilon(Y_{10s})$ generally increases as well. However, $\varepsilon(Y_{10f})$ generally decreases from 1 year to 3 years, with a slight increase from 3 years to 4 years.

To illustrate the lowest ε value, we develop Figure 4 for comparison purposes. It is evident that when projecting for Y_{10s}, school-level analysis using 1-year average provides the lowest ε value. When projecting for Y_{10f}, university-level analysis using 3 years of average yields the lowest ε value.

To examine if the forecasting models can be extended to project for spring and fall semesters, we performed the same calculation steps for forecasting undergraduate student enrollment at a university, projecting for Y_{09s} and Y_{09f}. We summarize $\varepsilon(Y_{09s})$ and $\varepsilon(Y_{09f})$ for both university- and school-level analyses in Table 13.

By using the same calculation steps, we show ε values using university-level analysis for projecting Y_{09s} and Y_{09f} in Table 14 and Figure 5.

Table 14. ε Values When Using Different Years of ARR by University-Level Analysis for Projecting Y_{09s} and Y_{09f}

	y_1	y_2	y_3	y_4
Y_{09s}	0.85	1.29	1.27	1.16
Y_{09f}	1.58	2.00	1.57	1.58

Figure 5. ε Values When Using Different Years of ARR by University-Level Analysis for Projecting Y_{09s} and Y_{09f}

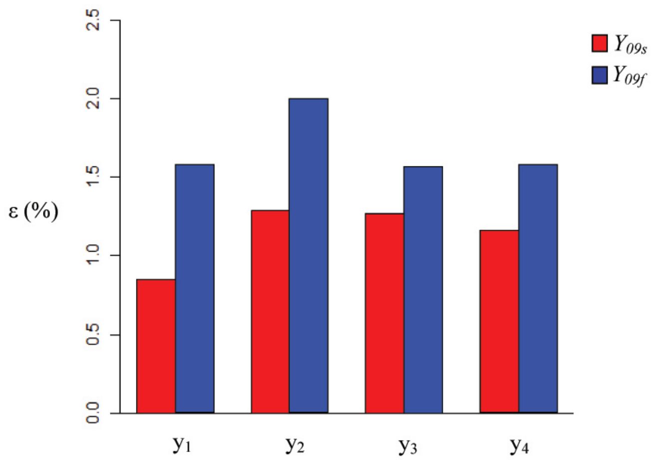
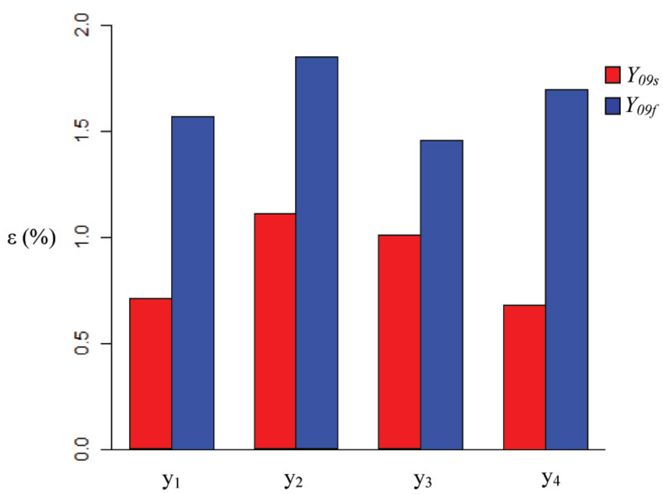


Table 15. ε Values When Using Different Years of ARR by School-Level Analysis for Projecting Y_{09s} and Y_{09f}

	y_1	y_2	y_3	y_4
Y_{09s}	0.71	1.11	1.01	0.68
Y_{09f}	1.57	1.85	1.46	1.70

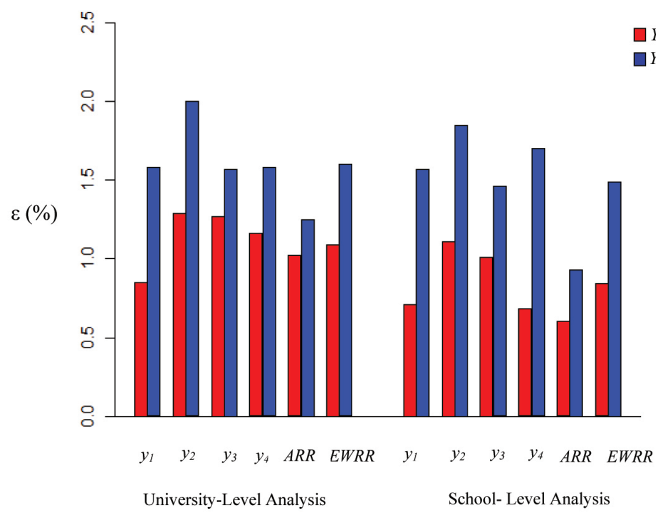
Figure 6. ε Values When Using Different Years of ARR by School-Level Analysis for Projecting Y_{09s} and Y_{09f}



By using the same calculation steps, we show ε values using school-level analysis for projecting Y_{09s} and Y_{09f} in Table 15 and Figure 6.

When finding the lowest ε value for projecting Y_{09s} and Y_{09f} we plotted Figure 7 for comparison purposes. Based on this comparison, it is clear that when projecting for Y_{09s} , school-level analysis based on the ARR yields the lowest ε value. When projecting for Y_{09f} , school-level analysis based on the ARR also provides the lowest ε . Hence, by combining the two cases, we recommend school-level analysis based on a 1-year average (i.e., lowest forecasting error) when projecting for spring semesters. On the other hand, we recommend university-level analysis using the ARR method since this forecasting model yields the lowest forecasting error when projecting for fall semesters. Based on the results, it is reasonable that averaging over more years of data provides a better forecast result when projecting for fall semesters since the data are noisy. This is mainly due to students taking internships or transferring to and from another school starting at the fall semester. On the other hand, forecasting for spring semester using 1 year of data is sufficient since the data are more stable.

Figure 7. ε Value Comparison for Projecting Y_{09s} and Y_{09f}



CONCLUSION AND FUTURE WORK

This research focuses on developing a low-cost and easy-to-use forecasting model for projecting undergraduate student enrollment. Based on a detailed analysis of historical data and different forecasting models, we developed and evaluated two forecasting models using different sets of enrollment data, including university-, school-, and division-level enrollment. University-level analysis is similar to many proposed methods in the literature that have shown high accuracy in projecting student enrollment as discussed in the previous sections of the paper. School- and division-level analyses were not mentioned in the literature we reviewed since these are the units of analysis to test the two new methods proposed in this research. The numerical results indicate that the forecasting errors will not decrease when applying division-level analysis versus school- and university-level analyses. Also, using all available student enrollment data does not necessarily produce a smaller forecasting error than using the most recent enrollment data. Based on the case study, by looking at different years' forecasting errors, school-level analysis using 1-year average should be used when projecting for spring semesters since the model yields the lowest average forecasting error of 0.40%. When projecting for fall semesters, university-level analysis using the ARR method should be used since it yields the lowest average forecasting error of 0.81%. Therefore, to keep the forecasting error rate at the lowest level, it is better to use school-level analysis with 1-year average when projecting for spring semesters and to use university-level analysis with the ARR when projecting for fall semesters.

The research is based on analyzing historical undergraduate student enrollment data from the State University of New York at Binghamton by comparing forecasting errors of different forecasting models. The proposed forecasting models should be updated constantly with current and accurate information regarding student enrollment data, such that new enrollment trends can be analyzed. A user-friendly graphical user interface can also be implemented and applied in the future to make the computations of the forecasting models more efficient and effective.

References

Armstrong, J. S. 2001. *Principles of forecasting: A handbook for researchers and practitioners.* Norwell, MA: Kluwer Academic.

Beck, B. D. 2009. Enrollment forecasting and modeling. Office of Academic Planning and Analysis, University of Wisconsin, Madison.

Brown, R. G. 1959. S*tatistical forecasting for inventory control.* New York: McGraw-Hill.

Desjardins, S. L., Ahlburg, D. A., & McCall, B. P. 2006. An integrated model of application, admission, enrollment, and financial aid. *Journal of Higher Education, 77*(3), 381–429.

Dobbs, S. 2001. Re-engineering the enrollment management system at the Monterey Peninsula Unified School District (MPUSD). MS thesis, Naval Postgraduate School, Monterey, CA.

Gardner, D. E. 1981. Weight factor selection in double exponential smoothing enrollment forecasts. *Research in Higher Education, 14*(1), 49–56.

Glover, R. H. 1986. Designing a decision-support system for enrollment management. *Research in Higher Education, 24*(1), 15–34.

Guo, S. 2002. Three enrollment forecasting models: Issues in enrollment projections for community colleges. Fortieth RP Conference, Pacific Grove, CA, May 1–3.

Holt, C. C. 2004. Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting, 20*(1), 5–10.

Hossler, D., & Bean, J. P. 1990. *The strategic management of college enrollments.* San Francisco: Jossey-Bass.

Magee, J. F. 1958. *Production planning and inventory control.* New York: McGraw-Hill.

Marshall, K. T., & Oliver, R. M. 1970. A constant-work model for student attendance and enrollment. *Operations Research, 18*(2), 193–206.

Muth, J. F. 1960. Optimal properties of exponentially weighted forecasts. *Journal of the American Statistical Association, 55*(290), 299–306.

Norton, D. R. 1998. Universities' enrollment management. State of Arizona Office of the Auditor, Phoenix, AZ.

Porter, S. R. 1999. Including transfer-out behavior in retention models: Using the NSLC enrollment search data. North East Association of Institutional Research Conference, Newport, RI.

Schmid, C. F., & Shanley, F. J. 1952. Techniques of forecasting university enrollment. *Journal of Higher Education, 23*(9), 483–488, 502–503.

Snyder, R. D. 1988. Progressive tuning of simple exponential smoothing forecasts. *Journal of the Operational Research Society, 39*(4), 393–399.

Tersine, R. J. 1994. *Principles of inventory and materials management*. Boston: PTR Prentice-Hall Publishing.

Appendix. List of Nomenclature

t	A specific semester
j	A specific school
k	A specific division
n	Total number of semesters
s_t	Semester t
$s_{\beta,\theta}$	Two consecutive semesters, where $\beta < \theta$
y	Number of years used to calculate ARR
ARR_t	Average return ratio for semester t
$ARR_{t,j}$	Average return ratio for semester t and school j
$ARR_{t,k}$	Average return ratio for semester t and division k
$ARR_{F,S}$	Average return ratio from fall to spring semester
$ARR_{S,F}$	Average return ratio from spring to fall semester
E	Undergraduate student enrollment
E_t	Number of undergraduate student enrollment at semester t
$E_{t,j}$	Number of undergraduate student enrollment at semester t in school j
$E_{t,k}$	Number of undergraduate student enrollment at semester t in division k
E_{t-1}	Number of undergraduate student enrollment at semester $t-1$
$E_{t-1,j}$	Number of undergraduate student enrollment at semester $t-1$ in school j
$E_{t-1,k}$	Number of undergraduate student enrollment at semester $t-1$ in division k
E_{total}	Total undergraduate student enrollment
E_{total}^{A}	Actual total undergraduate students
E_{total}^{P}	Projected total undergraduate students
$EWRR_t$	Exponential weighted return ratio for semester t
$EWRR_{t,j}$	Exponential weighted return ratio for semester t and school j
$EWRR_{t,k}$	Exponential weighted return ratio for semester t and division k
$EWRR_{F,S}$	Exponential weighted ratio from fall to spring semester
$EWRR_{S,F}$	Exponential weighted return ratio from spring to fall semester
RR	Return ratio
RR_t	Return ratio for semester t
RR_{t-1}	Return ratio for semester $t-1$
Y	Academic year
Y_f	Fall academic year
Y_s	Spring academic year
α	Weighting factor for EWRR method, where $0 < \alpha < 1$
ε	Forecasting error (%)

PROFESSIONAL FILE

ARTICLE 130

Association for Institutional Research

ALTERNATIVE ESTIMATES OF THE RELIABILITY OF COLLEGE GRADE POINT AVERAGES

Joe L. Saupe and Mardy T. Eimers

Acknowledgments
The authors acknowledge and appreciate the assistance of Ann Patton, senior programmer; and Nino Kalatozi, doctoral candidate in educational leadership and graduate research assistant, in the preparation of this paper. Both work in the Office of Institutional Research at the University of Missouri.

About the Authors
Joe L. Saupe is professor emeritus at the University of Missouri. Mardy T. Eimers is director, Institutional Research and Quality Improvement, at the University of Missouri.

The authors prepared this paper for the Annual Forum of the Association for Institutional Research, June 2–6, 2012, New Orleans, Louisiana.

Abstract
The purpose of this paper is to explore differences in the reliabilities of cumulative college grade point averages (GPAs), estimated for unweighted and weighted, one-semester, 1-year, 2-year, and 4-year GPAs. Using cumulative GPAs for a freshman class at a major university, we estimate internal consistency (coefficient alpha) reliabilities for the several GPAs. We compare these reliabilities to similar reliabilities found in the literature. Principal findings are that different cumulative GPAs have different degrees of reliability and that GPA reliability increases at a decreasing rate with number of semesters completed. Understanding these differences in reliability has implications for how GPAs are used by institutional researchers in practical as well as theoretical studies. The literature review and methods of the study should be useful to the institutional researcher who undertakes an investigation that involves GPA reliability.

INTRODUCTION

College grade point averages (GPAs) are used as predictors of success in undergraduate education, as predictors of success in graduate or professional education, as criteria for admission to degree programs, as indicators of qualification for employment, and as variables in different types of research (Warren, 1971). For each of these uses it is important that the GPAs possess some minimum degrees of reliability. For this reason, there have been a number of investigations into the reliability of college grades and GPAs (see Barritt, 1966; Clark, 1950; Etaugh, Etaugh, & Hurd, 1972; Ramist, Lewis, & McCamley, 1990). The reliability of the college GPA has also been used as one variable in studies of some other variable (Bacon & Bean, 2006; Millman, Slovacek, Kulik, & Mitchell, 1983; Singleton & Smith, 1978). An early study (Starch & Elliot, 1913) that dealt with grading high school examinations in mathematics and English indicates there has been interest in the reliability of grades for at least 100 years.

The problem that gives rise to the present study is that college GPAs are used as variables in institutional and other research efforts and are drawn upon in decision-making policies, often without consideration given to the reliabilities of the GPAs, to methods of calculating these reliabilities, or to reliability characteristics of alternative GPAs. Thus, the primary focus of this study is to provide greater understanding and clarification concerning these issues that underlie the use of the GPAs.

RELIABILITY AND COLLEGE GPAS

Classical measurement theory describes several basic approaches when estimating reliability (Crocker & Algina, 1986; Feldt & Brennan, 1989). The earliest definition of reliability is the correlation between two parallel forms of the same test (Feldt & Brennan, 1989). Test forms are considered to be parallel when they are constructed to cover the same domain or domains of content. It is not clear that there is a counterpart to parallel forms of tests in the case of college GPAs.

A second approach to estimating reliability is the test-retest procedure. With this approach, one gives the test twice to the same group of subjects and estimates the reliability of the test by the correlation between the two sets of scores. If two semesters or 2 years of college coursework are considered to be measures of the same variable, for example academic achievement, then the correlation between GPAs for the two semesters or 2 years may be viewed as a reliability estimate based on the test-retest situation. Clark (1950) compared correlations between first- and second-term GPAs with an alternative estimate of the reliability of the GPAs for a term. In a second study, Clark (1964) examined both approaches to estimating the reliability of GPAs in conjunction with comparing the reliability of grades on an eight-step grading scale with those on a five-step scale. Elliott and Strenta (1988) used correlations among annual GPAs in a study of differences in departmental grading standards. Humphreys (1968) calculated correlations among eight semesters of GPAs. Rogers (1937) also correlated term GPAs for eight academic terms. Werts, Linn, and Jöreskög (1978) used an eight-by-eight matrix of correlations among semester GPAs in their simplex analysis of that matrix. Finally, Willingham (1985) calculated correlations among yearly GPAs, but did not refer to them as reliabilities.

The third type of reliability is estimated by internal consistency methods (Crocker & Algina, 1986). The internal consistency of a test is the degree to which all of the items in the test are measures of the same characteristic or attribute or combination of characteristics or attributes. This type of reliability is estimated on the basis of a single administration of the test. There are at least three different methods that can be used to estimate internal consistency: (1) the split-half procedure, (2) coefficient alpha, and (3) analysis of variance (ANOVA).

The *split-half procedure* randomly divides the items of a test into two parts and then calculates the correlation between the scores on the two parts. This correlation is an estimate of the reliability of each half of the test. The estimate of the reliability of the whole test is estimated by use of the Spearman-Brown prophecy formula (Brown, 1910; Spearman, 1910), which expresses the reliability of the total test as a function of the correlation between the two halves of the test. Barritt (1966) used the split-half procedure to estimate the reliability of first-semester grades by randomly dividing the grades of students taking 12 or more credits into two sets of courses and correlating the resulting pairs of GPAs. In a similar study involving 38 colleges, Ramist and colleagues (1990) randomly divided freshman grades into two halves, calculated correlations between the GPAs of the two halves, and applied the Spearman-Brown formula. The generalized Spearman-Brown formula can be used to estimate the reliability of a test that is three, four, or some greater number times the length of the test for which there is a reliability estimate (Feldt & Brennan, 1989).

A second procedure for estimating the internal consistency type of reliability is known as the *coefficient alpha* procedure (Cronbach, 1951).[1] The formula for coefficient alpha involves the sum of the variances of the individual item scores and the variance of the total scores on the test. We did not find any studies of the reliability of GPAs using Cronbach's alpha in the literature reviewed.

Analysis of variance (ANOVA) is a third approach to estimating internal consistency. The most straightforward application of this approach involves a subjects-by-items ANOVA (Hoyt, 1941).[2]

The reliability estimate is a function of the mean square for students and the interaction or error mean square. Bendig (1953) estimated the reliability of grades for a single course using the ANOVA approach. Several instructors taught the course in multiple sections and four common tests plus individual instructor-made tests were used.

Other (ANOVA) procedures similar to that of Hoyt are also used. One such procedure is used when some characteristic of a group of subjects is rated, but different raters for different subjects are involved (e.g., Ebel, 1951; Shrout & Fleiss, 1979; Stanley, 1971). For example, Bacon and Bean (2006) used interclass correlation in their study of the reliabilities of GPAs that differed by number of years included, and of GPA in the major versus overall GPA. Etaugh and colleagues (1972) used the interclass correlation procedure to compare the reliabilities of unweighted mean grades with the reliability of mean grades weighted by their credit values for freshman year and senior year GPAs. Millman and colleagues (1983) used the interclass correlation ANOVA procedure to calculate reliabilities of major field GPAs in their study of the effect of grade inflation on the reliability of GPAs.

Other internal consistency procedures for estimating the reliability of GPAs have been suggested. In two previously cited studies, Clark (1950) and Clark (1964) investigated the use of a ratio of two standard deviations as

1 The Kuder-Richardson formula 20 (Kuder & Richardson, 1937), prominent in the literature on reliability, is equivalent to coefficient alpha when all test items are scored as 0 or 1. This situation does not occur when the measure is a college grade or GPA.

2 The reliabilities produced by the coefficient alpha and Hoyt ANOVA formulas are identical and the split-half procedure may be considered to be a special case of the coefficient alpha (Crocker & Algina, 1986). Specifically, the mean of the reliabilities calculated for all possible split halves of a test is very similar to coefficient alpha. The mean is identical to coefficient alpha if the split half is calculated by an alternative formula (Rulon, 1939) that involves differences between the scores on the two half tests rather than the correlation between the half test scores.

the estimate of the reliability of a GPA. Singleton and Smith (1978) calculated the average correlation among the first 20 courses taken by students and reported the results as reliabilities of individual course grades. The procedures for estimating the reliability of GPAs cited above as illustrations of the test-retest model might also be considered to be members of the internal consistency family.

Researchers who have studied the reliability of GPAs have uniformly used internal consistency procedures. In these studies, because GPA is considered to be an indicator of overall academic achievement, the internal consistency method is appropriate and we will employ it in the present study.

The literature on the reliability of college grades includes studies of the reliability of individual course grades (Bendig, 1953; Etaugh et al., 1972; Singleton & Smith, 1978), of single-term GPAs (Barritt, 1966; Clark, 1950, 1964; Rogers,1937; Werts et al., 1978), of 1-year GPAs (Bacon & Bean, 2006; Elliott & Strenta, 1988; Etaugh et al., 1972; Humphreys, 1968; Millman et al., 1983; Ramist et al., 1990; Willingham,1985), and of GPAs for more than 1 year of course work (Bacon & Bean, 2006). There have been relatively few studies of the reliability of the final undergraduate (cumulative) GPA, and that GPA is a focus of the present study.

PURPOSES

The purposes of this study are to focus the attention of researchers and practitioners on the reliability of college GPAs; to provide methods for estimating this reliability, including the method of this study and methods found in the literature; and to provide answers to the following research questions:

1. What are reliability estimates for one-semester, 1-year, 2-year, and 4-year GPAs, and how do they differ?
2. How do results of using the Spearman-Brown formula to estimate the reliabilities of college GPAs compare with the results of using coefficient alpha estimates?
3. What is the effect on reliabilities calculated for multisemester GPAs of weighting semester GPAs by the credits of those GPAs?
4. How do reliabilities found in this study compare with similar reliabilities reported in the literature?

In terms of the first research question, previous research suggests that two factors may affect the reliability of GPAs over time. In a study of the effects of grade inflation on GPA reliability (Millman et al., 1983), there were nonsignificant decreases in GPA reliability over time. However, Bacon and Bean (2006) found that 4-year overall GPAs had a higher reliability (.94) than other limited time frame GPAs, including most recent 1 year (.84) or most recent 2 years (.91). It might be expected that the variance of 4-year GPAs is lower than that of first-year GPAs because of the loss of lower-achieving students between the end of the first year and the end of the fourth year. That lower variance should lead to a lower reliability. On the other hand, adding items to a test can be expected to increase the reliability of the test according to the generalized Spearman-Brown formula (Feldt & Brennan, 1989). In this study, a semester GPA is the counterpart of the test item. Thus, more semesters should lead to higher reliabilities. Consequently, the comparison of reliability estimates of GPAs at different stages of college completion is of interest.

To address research question 2, the reliabilities of two-, four-, and eight-semester GPAs are calculated directly and compared to the reliabilities calculated by the generalized Spearman-Brown formula from a one-semester GPA reliability.

The semester GPAs of different students are based on the differing numbers of credits involved in these GPAs. It might seem that the reliabilities of multiterm GPAs could be improved by giving more weight to those GPAs based on larger numbers of credits. However, Etaugh and colleagues (1972) found that unweighted GPAs had higher reliabilities than did weighted GPAs. The need for additional information on this matter is the basis of the third research question.

The fourth research question has to do with the possibility of some uniformity among colleges and universities in the patterns of the reliability of cumulative GPAs at different stages in the college experience. Information on this possibility is provided by the comparison of GPAs from the literature with those found in this study.

Following are issues about the reliability of college GPAs that are found in the literature but are not dealt with in this study:

1. That different courses taken by different students may be expected to lead to lower GPA reliabilities than those that would occur if all students take the same courses. In a preceding section of this paper, we mention the literature on adjusting GPAs for differences in courses taken by different students (Elliott & Strenta, 1988; Young, 1990, 1993).
2. The reliability of the GPA for the courses of a major might be expected to be higher than the overall GPA. However, Bacon and Bean (2006) found that that the opposite is the case.
3. The fact that some students have the same instructor for two terms and others do not may be expected to affect the comparability, hence reliability, of the resulting grades (Clark, 1964).
4. That some students complete more academic terms than others may affect

the comparability, hence reliability, of their GPAs (Clark, 1964).
5. The number of points on the grade scale may affect the reliability of GPAs (Komorita & Graham, 1965; Masters, 1974).

DATA AND METHODOLOGY

The data for this study come from a large research university in the Midwest. Specifically, the data are for degree-seeking, full-time and part-time, first-time freshmen entering in the fall semester of 2007, including those who had enrolled for the preceding summer session. There were 4,970 students in this entering class. Forty-seven of these students did not remain enrolled long enough for an academic record to be posted for them at the end of that initial semester. End-of-semester credits and semester GPAs are recorded for each student for each of the eight semesters. Summer session and intersession GPAs are not included. We include the numbers of consecutive semesters that the 4,970 students remained enrolled, as well as the students' cumulative GPAs at the end of the first, second, and fourth years as recorded in university records.

From these data, we calculate cumulative GPAs for the end of the first two, first four, and all eight semesters; we also calculate weighted semester GPAs for students completing two, four, eight semesters. We calculate a weighted GPA by multiplying the semester GPA by the ratio of the number of credits in that GPA by the mean number of credits in the GPAs of all students for that semester.

The reliabilities calculated from the semester GPAs are the reliabilities of the sums or the means of the GPAs for the included semesters. These mean GPAs are not identical to the true cumulative GPAs that are recorded in the students' academic records. These GPAs involve the semester-by-semester numbers of credits completed. The reliabilities of the sums or means of the weighted semester GPAs may be better estimates of the reliabilities of the true cumulative GPAs. For this reason, we calculate and include weighted semester GPAs in the study.

We carried out the following data analyses:

- We calculate correlations among the following GPAs for students completing two, four, and eight semesters:
    1. Actual cumulative GPAs
    2. Cumulative GPAs calculated from the semester GPA data
    3. Means of semester GPAs
    4. Means of weighted semester GPAs
- We calculate these correlations in order to determine the degree to which they are interchangeable. Specifically, do the means of semester GPAs accurately reflect the cumulative GPAs? Do the calculated cumulative GPAs that exclude summer and intersession data accurately reflect the actual cumulative GPAs? How are the means of the weighted GPAs related to the other three measures?

We calculate correlations between first-semester and second-semester GPAs and between weighted first-semester and second-semester GPAs in order to estimate the reliability of first-year, one-semester GPAs, and to compare this reliability for unweighted and weighted GPAs.

We calculate internal consistency reliabilities using Cronbach alpha (Cronbach, 1951) for end of two-semester, end of four-semester, and end of eight-semester mean GPAs, unweighted and weighted, in order to compare GPA reliabilities over time and to compare reliabilities of unweighted and weighted GPAs. Using symbols for the GPA, the formula is alpha =

$$\frac{s}{s-1}\left(1-\frac{VARsem}{VARgpa}\right),$$

where *s* is the number of semesters, *VARsem* is the variance of GPAs for a semester, and *VARgpa* is the variance of the sums of GPAs.

Based on the reliability of one-semester GPAs, we use the Spearman-Brown procedure (Brown, 1910; Spearman, 1910) to estimate the reliability of two-semester GPAs, of four-semester GPAs, and of eight-semester GPAs in order to compare the procedures of estimating the reliability for the several comparable GPAs. The basic Spearman-Brown formula for estimating the reliability of a two-semester GPA is SB =

$$\frac{2r}{1 + r},$$

where *r* is the correlation between the two-semester GPAs. The generalized formula for estimating the reliability of a four- or eight-semester reliability is GSB =

$$\frac{sr}{1 + (s-1)\, r},$$

where *s* is the number of semesters for which the reliability is to be estimated.

RESULTS

We carry out the data analyses on groups of students defined on the basis of the number of consecutive semesters they completed. We use this basis for the selection of students to be included in an analysis so that all students included in a calculation of reliability had completed the same number of semesters without gaps in their attendance. Where there are gaps in the sequences of semesters completed, the coefficient alpha procedure would not be applicable. The alpha procedure allows differences among semesters to be ignored in the estimation of the reliability of the sum or mean of semester GPAs.

Table 1 shows the numbers and cumulative numbers of students completing each of the consecutive number of semesters. From the cumulative numbers, 4,606 students are included in the analyses of students completing two consecutive semesters, 3,922 in the analyses for students completing four consecutive semesters, and 2,968 in those for students completing eight consecutive semesters.

Table 2 contains the correlations among the four cumulative or mean GPAs for the groups of students completing two, four, and eight consecutive semesters. Means and standard deviations of the four overall GPAs are included for each group. The correlations among the actual cumulative GPAs, the calculated cumulative GPAs, and the mean GPAs exceed .99 for all three groups of students. The means and standard deviations for these three overall GPAs are comparable within each of the three groups with the mean of the actual cumulative GPA slightly but consistently exceeding the means for the other two measures. Also, the standard deviations for the actual cumulative GPAs are slightly but consistently smaller than those for the other overall GPAs.

The correlations of the means of weighted GPAs with the other three overall GPAs are consistently smaller than the intercorrelations among the first three overall GPAs. While the means of these GPAs are comparable to the means of the first three GPAs, their standard deviations are appreciably higher.

The mean GPAs increase and the standard deviations decrease as the number of semesters included increases. These trends are not surprising. In addition to possibly differing grading

Table 1. Numbers and Percentages of Students Who Completed Given Numbers of Consecutive Semesters

Consecutive Semesters	Number of Students	Cumulative Number	Percent of Students	Cumulative Percent
8	2,968	2,968	59.7%	59.7%
7	290	3,258	5.8%	65.6%
6	228	3,486	4.6%	70.1%
5	183	3,669	3.7%	73.8%
4	253	3,922	5.1%	78.9%
3	206	4,128	4.1%	83.1%
2	478	4,606	9.6%	92.7%
1	317	4,923	6.4%	99.1%
0	47	4,970	0.9%	100.0%
Total	4,970	--	100.0%	--

Table 2. Correlations Among and Means and Standard Deviations of the Four Cumulative or Mean Two-, Four-, and Eight-Semester GPAs

Variable	Calculated Cum GPA[2]	Mean of Sem GPAs[3]	Mean of Whtd Sem GPAs[4]	Mean	S.D.
Two-Semester GPAs (N = 4,606)					
Actual Cum GPA[1]	0.996	0.994	0.941	2.95	0.74
Calculated Cum GPA[2]		0.998	0.945	2.94	0.75
Mean of Sem GPAs[3]			0.944	2.94	0.75
Mean of Whtd Sem GPAs[4]			---	2.97	0.88
Four-Semester GPAs (N = 3,922)					
Actual Cum GPA[1]	0.994	0.993	0.934	3.10	0.55
Calculated Cum GPA[2]		0.998	0.938	3.08	0.57
Mean of Sem GPAs[3]			0.937	3.08	0.57
Mean of Whtd Sem GPAs[4]			---	3.10	0.70
Eight-Semester GPAs (N = 2,968)					
Actual Cum GPA[1]	0.994	0.992	0.916	3.19	0.46
Calculated Cum GPA[2]		0.998	0.921	3.16	0.49
Mean of Sem GPAs[3]			0.920	3.16	0.49
Mean of Whtd Sem GPAs[4]			---	3.17	0.58

[1] Cumulative GPA take from the University's student data base.
[2] Calculated cumulative GPA from semester GPAs and credits.
[3] Mean of semester GPAs.
[4] Mean of weighted semester GPAs.

standards between courses taken by freshmen or sophomores, and courses taken by juniors and seniors, these trends very likely reflect the loss of lower-achieving students over 4 years of the study.

Table 3 provides the several reliability estimates for one-semester, two-semester, four-semester, and eight-semester GPAs. The one-semester reliabilities are correlations between first- and second-semester GPAs for students who completed the first two semesters. The Spearman-Brown estimates are derived from the one-semester reliabilities in the table. The remaining reliabilities are coefficient alphas calculated for each group of students completing two-, four-, or eight-consecutive semesters. The one-semester reliabilities, .72 and .69, are similar, but the value for unweighted GPAs is modestly higher than the value for weighted GPAs. The Spearman-Brown values for two-, four-, and eight-semester unweighted and weighted GPAs are also similar, with differences ranging from .02 to .00. The alpha reliabilities for unweighted GPAs consistently but modestly exceed those for weighted GPAs. The Spearman-Brown estimates for four- and eight-semester GPAs are moderately higher than the corresponding alphas. Finally, in each case the reliability estimate increases from approximately .70 to .91 or higher as the number of semesters increase.

Reliabilities for one-, two-, four-, and eight- semester GPAs from the literature that are comparable to those of this study, including those found in this study, are as follows:

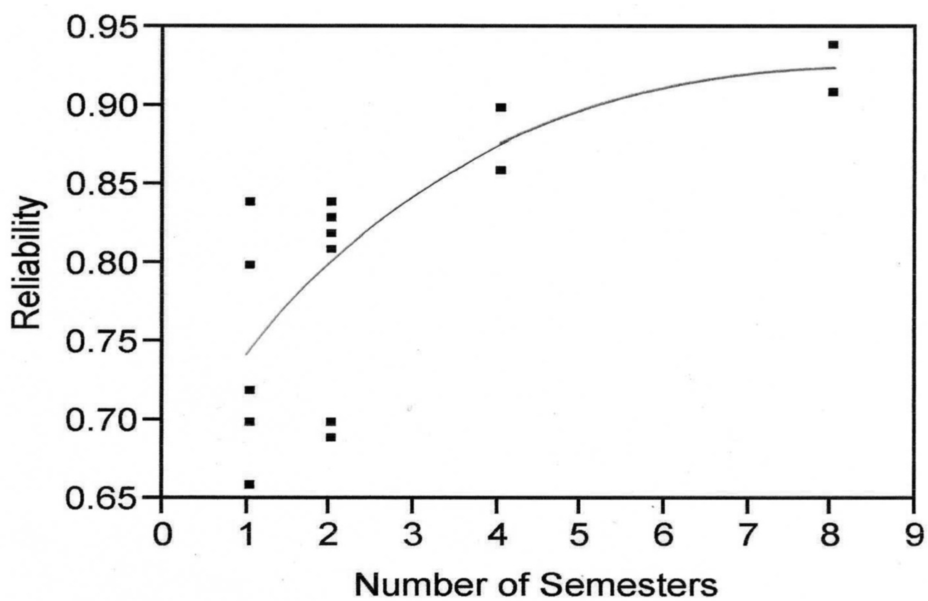
- One-semester GPAs: .72 (this study), .84 (Barritt, 1966), .70 (Clark, 1964), .66 (Humphreys, 1968), and .80 (Rogers, 1937).
- Two-semester GPAs: .84 (this study), .84 (Bacon & Bean, 2006), .69 (Elliott & Strenta, 1988), .81 (Etaugh et al., 1972), .83 (Millman et al., 1983), .82 (Ramist et al., 1990), and .70 (Willingham, 1985).
- Four-semester GPAs: .86 (this study) and .90 (Bacon & Bean, 2006).
- Eight-semester GPAs: .91 (this study) and .94 (Bacon & Bean, 2006).

Other reliabilities of GPAs are reported in the literature, but the above values are the most comparable to the GPAs in this study. We had to make a few decisions to select these comparable reliabilities. For example, in a couple of cases we use the average of two or more reliabilities from a single study. Also, the one-semester reliabilities used here are first-semester (or second-semester) reliabilities; we do not select values for subsequent semesters.
To facilitate comparisons of these reliabilities, we provide Chart 1. The chart shows the relationship between the number of semesters, one through

Table 3. Internal Consistency Reliability Estimates by Number of Semesters and Method of Estimating Reliability

Method of Reliability Estimate	One Semester	Two Semesters	Four Semesters	Eight Semesters
N	4,606	4,606	3,922	2,968
Correlation - Unweighted GPAs	0.72	--	--	--
- Spearman-Brown	--	0.84	0.91	0.95
Correlation - Weighted GPAs	0.69			
- Spearman-Brown		0.82	0.90	0.95
Alpha - Unweighted GPAs	--	0.84	0.86	0.91
Alpha - Weighted GPAs	--	0.81	0.85	0.85

Chart 1. Reliability of GPA by Number of Semesters, Data from This Study, and Data from the Literature



eight, of coursework on which a GPA is based, and the reliability of that GPA. The values in the chart are given above.

These reliabilities were derived using a variety of procedures. This study is the only one that made use of coefficient alpha. The split-half procedure and the Spearman-Brown formula are used in this study and others. Other studies employed various ANOVA approaches to estimating GPA reliability. It might be expected that values of reliabilities estimated by different procedures would to some degree be dependent on the procedure used. Also, the various studies were carried out with data from a variety of colleges and universities. The reliability of a GPA might be expected to vary from one type of institution to another. For example, the university from which the data of this study come is comprehensive, offering a great variety of undergraduate majors. To the degree that grading standards vary to some degree among majors, this variety of majors might be expected to depress the reliability of overall GPAs. Thus, Chart 1 should be considered to be suggestive and not definitive. It does suggest there is a generally positive relationship between the two variables.

DISCUSSION

As previously noted, the alpha reliabilities of this study are the reliabilities of sums of semester GPAs. They correspond to the total scores on a test for which an alpha is calculated. To make these sums of GPAs comparable to other GPAs, we divided them by the appropriate number of semesters and expressed them as means. Also, these means of semester GPAs exclude grades earned in summer sessions or intersessions. The GPAs that should be of interest are the cumulative GPAs that appear in the students' official records. These GPAs are, of course, influenced by the numbers of credits on which each semester GPA is based and include grades earned in summer sessions and intersessions. The correlations, over .99, between the means of semester GPAs and the actual cumulative GPAs and the similarity of the means and standard deviations of these two variables indicate that the alpha reliabilities of this study are very good estimates of the reliabilities of the cumulative GPAs in the students' records. The third indicator of overall achievement, the cumulative GPA calculated from semester GPAs and credits, also excludes grades earned in summer sessions and intersessions and is included in the study in order to discern if the exclusion of these grades impacts the accuracy of the alpha reliabilities. The near 1.00 correlations among these three overall GPAs and the similarity of their means and standard deviations suggest they are essentially interchangeable and provide confidence that the alpha reliabilities are very good estimates of the reliabilities of the actual cumulative GPAs. The lower correlations involving the means of weighted GPAs and the higher standard deviations for these variables indicate that the weighting procedure does not improve the comparability of these overall GPAs to the actual cumulative GPAs. As a matter of fact, the weighting procedure distorts the validity of the resulting GPAs. This finding is reinforced by the fact that the reliabilities of the GPAs resulting from the weighting procedure are lower than the reliabilities of the corresponding unweighted GPAs. Etaugh and colleagues (1972) also found that weighting GPAs results in lower reliabilities for composite GPAs than does not weighting GPAs.

The one-semester reliabilities of .72 (unweighted) and .69 (weighted) are correlations between semester-one and semester-two GPAs. The Spearman-Brown values for two-semester GPAs are the results of applying the basic Spearman-Brown formula to the respective correlations and the Spearman-Brown values for two-, four-, and eight-semester GPAs are products of the generalized Spearman-Brown formula. The similarity of the two reliabilities for two-semester GPAs and of the six reliabilities for two-, four-, and eight-semester GPAs indicates that the Spearman-Brown technique, as applied here, produces quite reasonable estimates of the reliabilities of GPAs for more than one semester of coursework. That the reliabilities of weighted GPAs are consistently lower than the reliabilities of unweighted GPAs is another indication that the weighting procedure is undesirable. The conclusion must be that the weighting procedure contributes error variance to the resulting average GPAs. In other words, it decreases the validity of the overall GPAs as indicators of a student's academic achievement.

The Spearman-Brown estimates of reliabilities for four- and eight-semester GPAs exceed their corresponding alpha reliabilities. Although the differences are not large, this result suggests that the alpha reliabilities are affected by the decrease in the variances of overall GPAs as the number of semesters increase. The Spearman-Brown estimates are not affected by this decrease in variance.

Reliabilities of GPAs found in this study are not unlike those taken from the literature. For the five one-semester GPAs, the range is from .66 to .84 (.72 in this study). Seven two-semester GPA reliabilities range from .69 to .84 (.84 in this study). There are only two four-semester reliabilities, .90 and, from this study, .86, and two eight-semester reliabilities, .90 and, from this study, .91. There are clearly too few values for a meta-analysis of these values, but these data suggest a trend in the relationship between the reliability of the GPA and the number of semesters on which it is based. As portrayed by the line fitted in Chart 1, the GPA reliability increases at a decreasing rate as the number of semesters increases. Additional research is needed to confirm this relationship.

The reliability of a GPA determines an upper bound to the correlation of that GPA with another variable. If the GPA were perfectly reliable, the correlation would be higher than that observed with the GPA that has a reliability of less than 1.00. For example, Saupe and Eimers (2011), in a study of how restriction of range in high school GPA depresses correlations in the prediction of success in college, note that unreliability in the college success variable is another factor that depresses such correlations. They find a correlation of .56 between high school core course GPA (CCGPA) and freshman year GPA (FYGPA). If the reliability of the FYGPA is .84, as found in the present study, then using the relationship provided by Walker and Lev (1953), the correlation between CCGPA and a perfectly reliable FYGPA would be .61. [3]

CONCLUSIONS

The following conclusions seem warranted:

1. Means of semester GPAs are almost identical to actual cumulative GPAs. Consequently, the reliabilities of sums (or means) of semester GPAs are good estimates of the reliabilities of actual cumulative GPAs.
2. Reliabilities of cumulative GPAs increase from the first semester to the end of the undergraduate program at a decreasing rate. In the present study, the increase is from .72 for the first-semester GPA, .84 for the two-semester GPA, and .86 for the four-semester GPA, to .91 for the eight-semester or near-final undergraduate GPA. Similar values and trends are likely to be found at other colleges and universities.
3. The use of the Spearman-Brown generalized formula to estimate reliabilities of longer-term GPAs from the reliability of first-semester GPA provide generally accurate, but moderately overstated, values.
4. Reliabilities calculated from weighted semester GPAs understate the reliabilities calculated from unweighted GPAs, and weighted GPAs do not provide good estimates of actual cumulative GPAs.

LIMITATIONS AND FURTHER RESEARCH

One limitation of this research is that the data came from a single institution and from a single entering class of that institution. This limitation is not uncommon. It is mitigated to some degree by the comparisons of the GPA reliabilities estimated from these data with reliabilities found in the literature. A second limitation is that students who completed one, two, or four semesters and then were not enrolled for one or more semesters before reenrolling are excluded from some of the reliability estimates. This limitation may also be mitigated because the reliabilities estimated using the Spearman-Brown procedure are similar to those estimated directly by coefficient alpha. Additional research on the reliability of college GPAs could be directed toward the question of whether the relationship between reliability values and number of semesters completed is similar across institutions. The suggestion of this study that this relationship may be similar for different colleges and universities needs further study. Also, further research could attempt to discern whether the reliabilities of college GPAs differ among different types of institutions. For example, are GPA reliabilities lower for selective institutions than for those not selective due to the smaller variance in levels of ability in the former?

IMPLICATIONS

The true standard of academic success is represented by a student's GPA. Whether the GPA is cumulative, by semester, or calculated in some other manner, it is critically important. The GPA can impact a college student's ability to pursue that coveted major, maintain or qualify for a financial aid award or scholarship, get into the graduate school of choice, or land the job that propels the graduate to greater opportunities. As easily as it can open doors, GPA thresholds can also keep doors closed. Consequently, it is important to know as much about the GPA as possible—including its reliability.

The purpose of this study was to examine the reliability of college GPAs, to provide different methods for estimating these reliabilities, and to add to the knowledge base in terms of the research literature and practical application in colleges and universities. Thus, we propose the following implications. First, the user of college GPAs should be aware that the reliabilities of GPAs vary according to the stage of the college career at which the GPAs are determined. It appears that the reliability increases as the student completes additional coursework. Also, it can be expected that even as early as the end of the first year, the reliability of the GPA may well be at an acceptable level of .80 or higher.

Second, there are a number of methods that can be used to estimate the reliability of a college GPA. This study introduced coefficient alpha as a method for determining the reliability of a GPA.

[3] $R^c_{hc} = R_{hc} / \sqrt{R_{cc}}$, where R_{hc} is the original correlation between HSGPA and FYGPA, R_{cc} is the reliability of FYGPA, and R^c_{hc} is the estimated correlation between HSGPA and FYGPA assuming the reliability of FYGPA is 1.00.

This method may prove to be beneficial to institutional researchers and faculty researchers who examine the reliability of college GPAs.

Third, frequently researchers and practitioners alike do not think about the reliability of college GPA. They may be interested in understanding how well admission tests (e.g., ACT, SAT, etc.), high school rank in class, high school GPA, and similar variables predict success in college. Success in college is almost always tied to the student's GPA in some manner. However, how often is the reliability of the dependent variable, the GPA, considered? How often is the reliability of the GPA at different periods over a student's career questioned? If this study has highlighted the importance of GPA reliability in both practical and scholarly pursuits, it will have accomplished a principal goal.

REFERENCES

Bacon, D. R., & Bean, B. (2006). GPA in research studies: An invaluable but overlooked opportunity. *Journal of Marketing Education, 28*(1), 35–42.

Barritt, L. S. (1966). The consistency of first-semester grade point average. *Journal of Educational Measurement, 3*(3), 261–262.

Bendig, A. W. (1953). The reliability of letter grades. *Educational and Psychological Measurement, 13*(2), 311–321.

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology,* 3(3), 296–322.

Clark, E. L. (1950). Reliability of college grades. *American Psychologist, 5*(7), 344.

Clark, E. L. (1964). Reliability of grade point averages. *Journal of Educational Research, 57*(8), 428–430.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* New York: Wadsworth.

Cronbach, L. J. (1951). Coefficient alpha and the internal consistency of tests. *Psychometrika, 16*(3), 297–334.

Ebel, R. L. (1951). Estimation of the reliability of ratings. *Psychometrika, 16*(4), 407–424.

Elliott, R., & Strenta, A. C. (1988). Effects of improving the reliability of the GPA on prediction generally and on comparative predictions for gender and race. *Journal of Educational Measurement, 25*(4), 333–347.

Etaugh, A. F., Etaugh, C. F., & Hurd, D. E. (1972). Reliability of college grades and grade point averages: Some implications for the prediction of academic performance. *Educational and Psychological Measurement, 32*(4), 1045–1050.

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), Educational measurement (pp. 105–146). Washington, DC: American Council on Education.

Hoyt, C. (1941). Test reliability estimated by analysis of variance. *Psychometrika, 6*, 153–160.

Humphreys, L. G. (1968). The fleeting nature of the prediction of college academic success. *Journal of Educational Psychology, 59*(5), 375–380.

Komorita, S. S., & Graham, W. K. (1965). Number of scale points and the reliability of the scales. *Educational and Psychological Measurement, 4*, 987–995.

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika, 2,* 151–160.

Masters, J. R. (1974). The relationship between number of response categories and reliability of Likert-type questionnaires. *Journal of Educational Measurement, 11*, 49–53.

Millman, J., Slovacek, S. P., Kulick, E., & Mitchell, K. J. (1983). Does grade inflation affect the reliability of grades? *Research in Higher Education, 19*(4), 423–429.

Ramist, L., Lewis, C., & McCamley, L. (1990). Implications of using freshman GPA as the criterion for the predictive validity of the SAT. In W. W. Willingham, C. Lewis, R. Morgan, & L. Ramist (Eds.), *Predicting college grades: An analysis of institutional trends over two decades* (pp. 253–288). Princeton, NJ: Educational Testing Service.

Rogers, H. H. (1937). The reliability of college grades. *School and Society, 45*, 758–760.

Rulon, P. J. (1939). A simplified procedure for determining the reliability of a test by split-halves. *Harvard Educational Review, 9*(1), 99–103.

Saupe, J. L., & Eimers, M. T. (2011). Correcting correlations when predicting success in college. *IR Applications, 31*, 1–11.

Shrout, P. E. & Fleiss, J. L. (1979). Intraclass correlation: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428.

Singleton, R., & Smith, E. R. (1978). Does grade inflation decrease the reliability of grades? *Journal of Educational Measurement, 15*(1), 37–41.

Spearman, C. (1910). Correlation calculated from faulty data. British Journal of Psychology, 3(3), 271–295.

Stanley, J. C. (1971). Reliability. In R. L Thorndike (Ed.), Educational measurement (pp. 356–442). Washington, DC: American Council on Education.

Starch, D., & Elliot, E. C. (1913). Reliability of grading work in mathematics. *School Review, 21*, 294–295.

Walker, H. M., & Lev, J. (1953). *Statistical inference.* New York: Henry Holt.

Warren, J. R. (1971). *College grading practices: An overview.* Washington, DC: ERIC Clearinghouse on Higher Education.

Werts, C., Linn, R. L., & Jöreskög, K. G. (1978). Reliability of college grades from longitudinal data. *Educational and Psychological Measurement, 38*(1), 89–95.

Willingham, W. W. (1985). *Success in college.* New York: College Entrance Examination Board.

Young, J. E. (1990). Are validity coefficients understated due to correctable defects in the GPA? *Research in Higher Education, 31*(4), 319–325.

Young, J. E. (1993). Grade adjustment methods. *Review of Educational Research, 63*(2), 151–165.

PROFESSIONAL FILE

ARTICLE 131

Association for Institutional Research

PROGRAMS FOR ENGAGEMENT AND ENHANCEMENT

Gloria Crisp, Lisa Palacios, and John Kaulfus

About the Authors
Gloria Crisp is associate professor of higher education at The University of Texas at San Antonio. Lisa Palacios is director of graduate recruitment at The University of Texas at San Antonio. John Kaulfus is assistant vice president/dean of students at Texas A&M University-Commerce.

Abstract
The following article describes programs used by universities and colleges to engage students; these programs include mentoring, learning communities, and first-year success courses and programs. We begin with a brief overview of student development theory, program descriptions and citations, and article summaries for key references. Next, we introduce prominent national surveys available to institutions that are interested in measuring student engagement (inside or outside formal programs). We conclude with additional references and recommendations for institutional researchers involved in program review and/or student outcomes assessment of student engagement programs.

INTRODUCTION

Higher education is not a passive experience that leaves students untouched. Rather, college life involves a variety of experiences, both inside and outside the classroom, designed to engage students and enhance their lives by introducing new ideas, challenging past behaviors or events, and creating intellectual discord and tension (Keeling, Wall, Underhile, & Dungy, 2008). Institutional effectiveness is dependent, in part, on institutions providing students with opportunities to purposefully engage (Harper & Quaye, 2009). According to Pascarella, "an excellent undergraduate education is most likely to occur at those colleges and universities that maximize good practices and enhance student engagement" (2001, p. 22). As such, institutions that value student success will take every opportunity to engage students both academically and socially (Culp, 2007).

Simply defined, student engagement is how universities organize their human capital and resources to encourage students to involve themselves in academic, interpersonal, and cocurricular activities (Astin, 1993). Student engagement is typically not viewed as a direct measure of student learning, but rather is used as a measurement of participation in meaningful educational experiences and activities that facilitate both social and academic integration (Tinto, 2000) and lead to student development (LaNasa, Olson, & Alleman, 2007). More specifically, opportunities for students to engage are provided through formalized programs designed to directly support student integration and/or development outcomes (i.e., study strategies, time/stress management skills, motivation, academic self-confidence, connections with peers, and out-of-class interactions with faculty), that in turn directly impact traditional measures of student success (i.e., grades, persistence).

According to a survey involving 185 colleges and universities across the country, the most prevalent services and programs provided to students to promote student engagement in the first year include tutoring, academic coaching and counseling, writing support services, academic advising, and testing services (National Resource Center for the First-Year Experience and Students in Transition, 2008). In large part due to work by Kuh and colleagues, engagement programs and activities have become increasingly viewed as an important component of student success (Kuh, Kinzie, Buckley, Bridges, & Hayek, 2007). As such, increasing attention has been given to the implementation, administration, and assessment of educational experiences designed to engage students.

Although engagement programs are typically created and managed by student affairs professionals, institutional researchers should be familiar with programmatic efforts on their campus, and should understand how program outcomes can be used to address accreditation standards and institutional planning and assessment goals (as demonstrated in volume 141 of *New Directions for Institutional Research*, 2009). The present article describes several programs currently used by postsecondary institutions to engage students with the intent of providing institutional researchers with knowledge to support assessment efforts. The article begins with an overview of relevant student development theory that serves as a conceptual grounding for engagement programs. Next, we provide program descriptions for programs that have been linked to engagement (e.g., social engagement, academic skills, time management, and career selection) and academic outcomes. Citations and article summaries for key references are provided in table form following each section for institutional research professionals who are interested in learning more about student engagement and/or enrichment programs. Third, we highlight prominent national surveys available to institutions that are interested in measuring student engagement (inside or outside of formal programs). The article concludes with additional references and recommendations for institutional researchers involved in program review and/or student outcomes assessment of student engagement programs.

OVERVIEW OF STUDENT DEVELOPMENT THEORY

Research and theory by Erikson and Chickering provide a foundation for our current understanding of student development. Erikson's theory of psychosocial development (1968) explains that individuals must work through eight stages in order to successfully form an identity and discover purpose and meaning in life. According to Erikson, adolescents move through a developmental stage during college termed the "identity versus role confusion" stage before moving into adulthood. This stage involves students successfully, or in some cases unsuccessfully, developing a personal identity; it is defined by a "crisis" that must be resolved in order for students to avoid an "identity crisis" that leads to stagnation or regression. Similarly, Chickering's seven vectors of student development (1969) explain that college students move through seven vectors or stages as they become more self-aware and as they have more complex thoughts, which is spurred

Table 1. Student Engagement Theory References

References for Student Engagement Theory	Impact for Institutional Researchers
Astin, A. (1993). *What matters in college: Four critical years revisited*. San Francisco: Jossey-Bass.	Presents the results from a study of almost 25,000 students at 200 colleges and universities. Findings demonstrate how colleges and universities can enhance student development during college through a variety of in-class and out-of-class experiences.
Hamrick, F. A., Evans, N. J., & Schuh, J. H. (2002). *Foundations of student affairs practice: How philosophy, theory, and research strengthen educational outcomes*. San Francisco: Jossey-Bass.	Provides an overview of theories that relate to student growth and development during college and explores ways that institutions can enhance students' educational experiences.
Harper, S. R., & Quaye, S. J. (2009). *Student engagement in higher education: Theoretical perspectives and practical approaches for diverse populations*. New York: Routledge.	Based on theory, explores ways that diverse populations of students (e.g., racial and ethnic minorities, LGBT students) as compared to nondiverse populations might struggle to engage during college.
Kuh, G. D., Kinzie, J., Schuh, J. H., & Whitt, E. J. (2005). *Assessing conditions to enhance educational effectiveness. The inventory for student engagement and success*. San Francisco: Jossey-Bass.	Provides a theoretical framework, the Inventory for Student Engagement and Success (ISES), to examine student engagement within a program, division, college, or entire institution. Explains how information can be used for program reviews, planning, and accreditation.
Pascarella, E., & Terenzini, P. T. (2005). *How college affects students: A third decade of research*. San Francisco: Jossey-Bass.	Presents findings from 15 years of research on college's impact on student learning, growth, and development. Also presents implications for research, policy, and practice.
Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition* (2nd ed.). Chicago: University of Chicago Press.	Synthesizes research on student retention demonstrating the importance of institutions providing students with opportunities to engage with the campus community.

by interactions with peers and faculty and the introduction of new concepts and ideas. Chickering's work has since been updated to be inclusive of non-traditional students (i.e., Chickering & Reisser, 1993).

More recently, Erikson and Chickering's work has been expanded by Astin and Tinto in an effort to understand the factors related to student success and persistence. Astin's theory of involvement (1984, 1999) postulates that student involvement in college has a direct impact on psychosocial development and assists in identity formation as students work toward graduation. Astin's work also demonstrates that student learning and development are dependent on active involvement in academic and social aspects of a college experience. Moreover, his theory argues that development is influenced by both the quality and the quantity of involvement.

Similarly, Tinto's theory of student departure (1993) demonstrates that students are more likely to persist toward graduation if they become socially and academically integrated into the college environment. He postulates that integration is achieved when a student and the institution share similar values and the student is engaged in positive social and academic interactions. Tinto's work demonstrates the importance of support from faculty and university staff. Table 1 contains key references related to theory underpinning student engagement programs to guide the development of programmatic activities and goals.

ENGAGEMENT PROGRAMS

The following section provides a program overview for mentoring, learning communities, and first-year success courses and programs. This is not meant to provide a comprehensive overview of engagement programs, but rather to provide institutional researchers with examples of programs currently employed on college and university campuses across the country that have been empirically shown to enhance students' experiences and to promote students engagement.

Mentoring Programs

Mentoring programs that involve a variety of engagement activities such as academic advising, academic skills development, personal development, and career selection are becoming increasingly prevalent. Mentoring programs and experiences have been empirically shown to be associated with numerous academic and developmental outcomes, including improving critical thinking skills, self-confidence, persistence, and academic performance. Mentoring has also been found to help students develop their latent abilities, and to raise students' expectations and future aspirations (e.g., Astin, 1999; Bank, Slavings, & Biddle, 1990; Campbell & Campbell, 1997; Freeman, 1999; Girves, Zepeda, & Gwathmey, 2005; Johnson, 1989; Mangold, 2003; Pagan & Edwards-Wilson, 2003; Roberts, 2000; Ross-Thomas & Bryant, 1994).

Institutional researchers should consider and draw from published program overviews and evaluations when assisting in the development and/or assessment of programmatic efforts. Unfortunately, there is little agreement regarding how college students experience mentoring, or on the components that should be included in a mentoring program. Moreover, it has been noted that the majority of empirical work on mentoring has been limited due to methodological weaknesses including limitations in how mentoring is defined and measured, a lack of sophisticated data analysis and theoretical grounding, failure to control for selection bias, and an overreliance of self-reported benefits of mentoring as the assessment measure (Crisp & Cruz, 2009). However, according to a comprehensive review of the psychological, business, and education literature by Nora and Crisp (2007), students perceive a holistic mentoring experience to include four separate yet interrelated types of support: (1) psychological and emotional support, (2) support for setting goals and choosing a career path, (3) academic subject knowledge support aimed at advancing a student's knowledge relevant to his or her chosen field, and (4) support in the form of a role model.

An assortment of mentoring programs designed to serve a variety of student populations including first-generation, minority, at-risk, and/or low-income students have been described in the literature (e.g., Bordes & Arredondo, 2005; Pagan & Edwards-Wilson, 2003; Wallace, Abel, & Ropers-Huilman, 2000). For instance, the Puente Project, evaluated by Laden (1999), is a nationally recognized program designed to raise Latino/a students' educational and career aspirations. Other examples of programs that involve a mentoring component include TRIO Programs (Wallace et al., 2000), the Adventor Program (Shultz, Colton, & Colton, 2001), and the Search for Education, Elevation and Knowledge (SEEK) Program (Sorrentino, 2007). Table 2 provides a list of select published work including a mentoring theory and scale (i.e., Crisp, 2009) to guide assessment efforts.

Learning Communities

Recently, there has been increased interest from both academic and student affairs practitioners to enhance and/or expand innovative programs such as learning communities and first-year experiences (Dale & Drake, 2007). Learning communities provide college students with the opportunity to get to know other students as well as faculty; these communities integrate students into the university commu-

Association for Institutional Research

Table 2. Mentoring Program References

References for Designing/ Assessing Outcomes for Mentoring Programs	Impact for Institutional Researchers
Crisp, G. (2009). Conceptualization and initial validation of the College Student Mentoring Scale (CSMS). *Journal of College Student Development, 50*(2), 177–194.	Offers a theoretically grounded survey to be used by institutions that are interested in measuring the mentoring experiences of undergraduate college students. Includes the 25-item survey as an appendix.
Laden, B. V. (1999). Socializing and mentoring college students of color: The Puente Project as an exemplary celebratory socialization model. *Peabody Journal of Education, 74*(2), 55–74.	Presents highlights from the Puente Project, a program designed to support first-generation Latino/a college students at California community colleges.
Shultz, E. L., Colton, G. M., & Colton, C. (2001). The Adventor Program: Advisement and mentoring for students of color in higher education. *Journal of Humanistic Counseling, Education, and Development, 40*, 208–218.	Describes the Adventor Program, designed by the College of Education at Kutztown University to serve racial/ minority students. Presents information about the program design and implementation, and follows with pilot findings.
Sorrentino, D. M. (2007). The SEEK mentoring program: An application of the goal-setting theory. *Journal of College Student Retention*, *8*(2), 241–250.	Presents a description and overview of the SEEK Program at the College of Staten Island, City University of New York (CUNY), that provides academic mentoring to students at risk for academic dismissal.
Wallace, D., Abel, R., & Ropers-Huilman, B. R. (2000). Clearing a path for success: Deconstructing borders through undergraduate mentoring. *The Review of Higher Education, 24*(1), 87–102.	Utilizes qualitative interview data to explore first-generation, low-income students' experiences with and perceptions of formalized mentoring programs.

nity in a meaningful way (Price, 2005; Taylor, Moore, MacGregor, & Lindblad, 2003; Tinto, 1998; Weber, 2000). More specifically, learning communities provide students with an educational environment that supports student engagement through an integrated and interdisciplinary curriculum that may cross departments or divisions and focuses on high levels of participation and support from faculty (Brower & Dettinger, 1998; Oertel as cited in Taylor et al., 2003; Price, 2005). Learning communities exist in a variety of formats to facilitate students' connection to faculty, other students, and the institution (Tinto, 1998); these formats include team-taught programs, paired or clustered courses, cohorts of students enrolled together in large courses, and residence-based programs (Price, 2005; Taylor et al., 2003; Tinto, 1998; Weber, 2000).

There is a wealth of literature on learning communities to suggest that programmatic efforts can be used to influence retention and learning outcomes. Namely, ongoing evaluations of the Opening Doors Learning Communities (ODLC) program by MDRC are utilizing experiments that test a cause-and-effect relationship between participation in learning communities and outcomes for various groups of students through the use of random assignments (e.g., Bloom & Sommo, 2005; Richburg-Hayes, Visher, & Bloom, 2008; Scrivener, Bloom, LeBlanc, Paxson, & Sommo, 2008). Additionally, nonexperimental research by Zhao and Kuh (2004) has revealed that students who participate in the learning community have higher levels of academic effort, active learning, interactions with faculty, and participation in diversity activities. Participants also reported more positive associations with advisers, campus support services, and overall experiences, as well as self-reported gains in personal and social development and basic skills advancement. Furthermore, qualitative work by Tinto and Goodsell (1993) involving a linked writing course and seminar found that learning communities supported the development of students' time management, writing, and study skills. Table 3 provides a list of key references to studies on learning communities.

First-Year/Orientation/Success Programs

According to a survey by the National Resource Center for the First-Year Experience and Students in Transition (2008), nearly 85% of colleges and universities currently offer a first-year program. First-year programs, student success courses, and orientation courses all focus on assisting college students' transition and/or enhancing engagement and success in college (Cook, 1996). These programs are designed to teach students strategies for

Table 3. Learning Communities References

References for Designing/ Assessing Outcomes for Learning Communities	Impact for Institutional Researchers
Richburg-Hayes, L., Visher, M. G., & Bloom, D. (2008). Do learning communities affect academic outcomes? Evidence from an experiment in a community college. *Journal of Research on Educational Effectiveness, 1*(1), 33–65.	Details the design and results of a study that utilized an experimental design to explore a cause-and-effect relationship between participation in a learning community and students' engagement and attachment to the institution's community.
Scrivener, S., Bloom, D., LeBlanc, A., Paxson, C. E., & Sommo, C. (2008). A good start: Two-year effects of a freshmen learning community program at Kingsborough Community College. MDRC, New York.	Explains how Kingsborough's Opening Doors Learning Communities (ODLC) program utilizes an experimental design with random assignment to study the effect of learning communities on student success.
Taylor, K., Moore, W. S., MacGregor, J., Lindblad, J. (2003). *What we know now. National Learning Communities Project Monograph Series*. The Washington Center for Improving the Quality of Undergraduate Education at The Evergreen State College in cooperation with the American Association for Higher Education, Washington, DC.	Presents findings from a systematic literature review of research and assessment specific to learning communities conducted by the National Learning Communities Project.
Tinto, V. (1997). Classrooms as communities: Exploring the educational character of student persistence. *Journal of Higher Education, 68*(6), 599–623.	Presents a mixed methods study of the use of learning communities at Seattle Central Community College. The study investigates the impact on collaborative learning strategies on both student persistence and learning.
Tinto, V. (1998). Learning communities and the reconstruction of remedial education in higher education. Prepared for presentation at the "Conference on Replacing Remediation in Higher Education" at Stanford University, January 26–67, 1998, sponsored by the Ford Foundation at the U.S. Department of Education.	Details a case study specific to the effectiveness of a developmental education learning community. Provides a description of the program, research procedures, and findings.

success in college by introducing them to campus facilities, resources, and services; and/or by enhancing students' health or well-being, study skills, time management, or learning styles (e.g., Derby, 2007; Derby & Smith, 2004; Derby & Watson, 2006; Glass & Garrett, 1995; Grunder & Hellmich, 1996; National Resource Center for the First-Year Experience and Students in Transition, 2008; O'Gara, Karp, & Hughes, 2008; Raymond & Napoli, 1998). Programmatic efforts may be offered as both credit and noncredit courses (Donnangelo & Santa Rita, 1982), and may be required or optional (Zimmerman, 2000). Possible program offerings range from a half-day orientation (Hollins, 2009) to semester- or year-long programs (Donnangelo & Santa Rita, 1982; Glass & Garrett, 1995).

Although the majority of research to date has focused on examining the impact of programs on retention or learning outcomes (e.g., Derby & Smith, 2004; Glass & Garrett, 1995; Pascarella, Terenzini, & Wolfle, 1986; Raymond & Napoli, 1998; Stovall, 1999), findings from the National Resource Center for the First-Year Experience and Students in Transition (2008) survey indicate that success courses may also be associated with engagement outcomes such as increasing peer connections, use of campus services, participation in campus services, and out-of-class interaction with faculty. Additionally, research conducted by the Community College Research Center (CCRC) at Teachers College, Columbia University, found that programmatic efforts may integrate students both socially and academically by helping to facilitate the development of students' relationships with faculty and other students (O'Gara et al., 2008). Moreover, evaluation efforts at the Virginia Community College System examined the impact of a comprehensive approach to student orientation that included a half- to full-day program (Seeking Opportunities through Academic Recruitment [SOAR]), group advising, and an orientation course. Findings indicated that the program increased students' personal adjustment during the transition process and academic gains among first-semester students. The orientation course was also found to assist students in developing effective study habits, career and academic planning, and knowledge regarding college resources (Hollins, 2009).

Resources and support regarding first-year programs are available to researchers through the First-Year Experience (http://www.sc.edu/fye/). Additionally, Table 4 provides a list of references specific to designing and assessing outcomes for first-year, orientation, and student success courses and programs.

TOOLS FOR ASSESSING STUDENT ENGAGEMENT

Several national surveys are available to institutions interested in assessing student engagement and/or students' experiences during college, including the National Survey of Student Engagement (NSSE; http://nsse.iub.edu/). This survey contains items assumed to measure different components of student engagement, including academic challenge (e.g., preparing for class, using higher-order thinking skills), active and collaborative learning (e.g., contributing to class discussions, working with students outside of class), and student interactions with faculty members (e.g., talking about career plans, working on activities other than coursework) (Kuh, 2004). Additionally, seniors report whether they participated in various programs and on-campus activities, including learning communities. The NSSE is typically administered in the spring using a paper or online version of the survey to a random sample of first-year and senior students (Kuh, Cruce, Shoup, Kinzie, & Gonyea, 2008).

Variations of the NSSE that measure engagement of different student populations are also available, including the Community College Survey of Student Engagement (CCSSE) (http://www.ccsse.org/) and Beginning College Survey of Student Engagement (BCSSE) (http://bcsse.iub.edu/). Moreover, many institutions compare student responses from the NSSE with faculty perceptions measured by the Faculty Survey of Student Engagement (FSSE) (http://fsse.iub.edu/). Tips and recommendations for analyzing and interpreting the NSSE survey data are available in a 2009 issue of New Directions for Institutional Research by Chen and colleagues. Another survey available to institutions interested in assessing students' development during the first year of college is the Your First College Year (YFCY) survey, developed through collaboration between the Higher Education Research Institute (HERI) and the Policy Center on the First Year of College at Brevard College (http://www.heri.ucla.edu/yfcyoverview.php). This survey allows colleges and universities to

Table 4. First-Year/Orientation/Success Program References

References for Designing/ Assessing Outcomes for First Year/Orientation Success Programs	**Impact for Institutional Researchers**
Engberg, M. E., & Mayhew, M. J. (2007). The influence of first year "success" courses on student learning and democratic outcomes. *Journal of College Student Development, 49*(2), 95–109.	Examines the impact of a first-year program on a variety of outcomes, including multicultural awareness, commitment to social justice, and the complexity of attributes. Also includes a discussion of connecting theory with practice.
Keup, J. R., & Barefoot, B. O. (2005). Learning how to be a successful student: Exploring the impact of first year seminars on student outcomes. *Journal of the First year Experience, 17*(1), 11–47.	Utilizes data from the Cooperative Institutional Research Program's (CIRP) 2000 freshman survey and the Your First College Year (YFCY) 2001 survey to study the relationship between taking a first-year course and academic and social experiences of students.
O'Gara, L. Karp, M. M., & Hughes, K. L. (2009). Student success courses in the community college: An exploratory study of student perspectives. *Community College Review, 36*(3), 195–218.	Investigates the perceived impact of taking a success course on individual benefits such as building relationships with faculty and developing study skills at two urban community colleges in the Northeast.
Schwitzer, A. M., McGovern, T. V., & Robbins, S. B. (1991). Adjustment outcomes of a freshman seminar: A utilization-focused approach. *Journal of College Student Development, 321*, 484–489.	Presents findings of an evaluation investigating the relationship between participation in a college orientation seminar and students' social and academic adjustment in college.
Zeidenberg, M., Jenkins, D., & Calcagno, J. C. (2007). *Do student success courses actually help community college students succeed?* Community College Research Center (CCRC) Brief No. 36, June.	Examines the impact of enrolling in a student success course over the course of 17 semesters on various student outcomes, controlling for possible extraneous variables.

identify students' experiences during the first year that encourage and support student involvement, satisfaction, and learning, as well as other success indicators that enhance first-year programs. Similar to the NSSE, the YFCY allows for comparisons to national and institutional peer groups among participating institutions as well as trend and longitudinal analyses. The YFCY is offered in both paper and web format and is conducted at the end of the students' first academic year (somewhere between the months of March to June).

Third, the Degrees of Preparation survey may also be of interest to institutions in measuring ways that college experiences are related to various developmental and civic outcomes, including critical thinking skills, career-related experiences, and civic engagement. This survey's major components and question descriptions are available in an issue of New Directions for Institutional Research (Ouimet & Pike, 2008). A copy of the piloted version of the survey is available at http://www.aascu.org/accountability/survey/?u=1. Additional information regarding the above-mentioned instruments as well as an inventory of other potentially relevant surveys and tools used to assess student engagement outcomes is posted on the Association for Institutional Research (AIR) website at http://applications.airweb.org/surveys/Default.aspx.

CONCLUSIONS AND RECOMMENDATIONS

We hope that the information presented in this article is useful to institutional researchers involved with program planning, assessment, and/or accreditation efforts; we offer the following conclusions and recommendations. First, we recommend that student engagement programs be clearly connected to the institution's core mission (Culp, 2007) and that they be grounded in student development theory (Dale & Drake, 2007). Institutional researchers should work with faculty as well as academic and student affairs personnel to utilize previously validated assessment tools and survey items that are grounded in theory, rather than developing home-grown surveys that may or may not be accurate measures of students' experiences. Additional recommendations specific to using engagement data in assessment and planning efforts are provided by Banta, Pike, and Hansen (2009).

When possible, we also strongly encourage the use of experimental designs that utilize random assignment to groups and an experimental (i.e., students in the program) and control group (i.e., group of students who do not participate) to assess cause-and-effect relationships between program activities and engagement outcomes. Examples of evaluation work utilizing experimental designs are provided in the learning community section of this article. Because experimental designs are rarely possible, we also recommend the use of quasi-experimental designs that adequately control for possible confounding variables (e.g., matching groups). Furthermore, in cases where the program is already in place or the independent variable (i.e., program) cannot be manipulated, we suggest the use of nonexperimental designs that adequately control for students' background characteristics and precollege characteristics that have been previously found to impact student outcomes (see discussion by Cole, Kennedy, & Ben-Avie, 2009). Finally, we suggest that institutional researchers consider using qualitative methods to answer "how" and "why" questions specific to program assessment.

Next, we encourage institutional researchers to actively seek out collaborations with faculty and student and academic affairs programs/offices. Student affairs personnel have knowledge of long-standing and professionally accepted student development theory (King & Howard-Hamilton, 2000) that is likely to strengthen assessment efforts and properly take into account factors such as ethnicity, culture, and career choice (Pickering & Sharpe, 2000). Student and academic affairs practitioners and faculty also have ready access and can encourage student participation in surveys needed to properly assess student outcomes (Smith & Mather, 2000). Moreover, institutional research offices may be able to provide precollege data to student affairs divisions to guide and inform the development of programmatic activities (Cole et al., 2009). Kinzie and Pennipede (2009) provide further discussion and recommendations for collaborating with student affairs in using data. Additionally, a New Directions article by Nelson Laird, Smallwood, Niskode-Dossett, and Garver (2009) offers ideas for involving faculty in assessment efforts specific to student engagement.

REFERENCES

Astin, A. W. (1984). Student involvement: A developmental theory for higher education. *Journal of College Student Development*, *25*(4), 297–308.

Astin, A. W. (1993). *What matters in college: Four critical years revisited.* San Francisco: Jossey-Bass.

Astin, A. W. (1999). Student involvement: A developmental theory for higher education. *Journal of College Student Development, 40*(5), 518–529.

Bank, B. J., Slavings, R. L., & Biddle, B. (1990). Effects of peer, faculty, and parental influence on students' persistence. *Sociology of Education, 63*, 208–225.

Banta, T. W., Pike, G. R., & Hansen, M. J. (2009). The use of engagement data in accreditation, planning and assessment. *New Directions for Institutional Research, 141*, 21–34.

Bloom, D., & Sommo, C. (2005). *Building learning communities: Early results from the*

Opening Doors demonstration at Kingsborough Community College. MDRC, New York.

Bordes, V., & Arredondo, P. (2005). Mentoring and 1st year Latino/a college students. *Journal of Hispanic Higher Education, 4*(2), 114–133.

Brower, A. M., & Dettinger, K. M. (1998). What is a learning community? Toward a comprehensive model. *About Campus*, Nov–Dec, 15–21.

Campbell, T. A., & Campbell, D. E. (1997). Faculty/student mentor programs: Effects on academic performance and retention. *Research in Higher Education, 38*(6), 727–742.

Chen, P. D., Gonyea, R. M., Sarraf, S. A., BrckaLorenz, A., Korkmaz, A., Lambert, A. D., . . . & Williams, J. M. (2009). Analyzing and interpreting NSSE data. *New Directions for Institutional Research, 141,* 35–54.

Chickering, A. (1969). *Education and identity*. San Francisco: Jossey-Bass.

Chickering, A., & Reisser, L. (1993). *Education and identity* (2nd ed.). San Francisco: Jossey-Bass.

Cole, J. S., Kennedy, M., & Ben-Avie, M. (2009). The role of precollege data in assessing and understanding student engagement in college. *New Directions for Institutional Research, 141,* 55–69.

Cook, L. P. (1996). A description of new student orientation programs at two-year colleges in the United States. (Dissertation Abstracts International UMI No. 9619361).

Crisp, G. (2009). Conceptualization and initial validation of the College Student Mentoring Scale (CSMS). *Journal of College Student Development, 50*(2), 177–194.

Crisp, G. & Cruz, I. (2009). Mentoring college students: A critical review of the literature between 1990 and 2007. *Research in Higher Education, 50*(6), 525–545.

Culp, M. M. (2007). Doing more of what matters: The key to student success. *New Directions for Community Colleges, 131*, 77–87.

Dale, P. A., & Drake, T. M. (2007). Connecting academic and student affairs to enhance student learning and success. N*ew Directions for Community Colleges, 131*, 51–64.

Derby, D. C. (2007). Predicting degree completion: Examining the interaction between orientation course participation and ethnic background. *Community College Journal of Research and Practice, 31*(11), 883–894.

Derby, D. C., & Smith, T. (2004). An orientation course and community college retention. *Community College of Research and Practice, 28*, 763–773.

Derby, D. C., & Watson, L. W. (2006). African-American retention within a community college: Differences in orientation course enrollment. *Journal of College Student Retention, 8*(3), 377–390.

Donnangelo, F. P., & Santa Rita, E. D. (1982). The effects of two collage orientation courses upon the academic performance and retention of entering freshmen. Bronx Community College, New York. (ERIC Document Reproduction Service No. ED232747).

Erikson, E. (1968). Identity: Youth and crisis. New York: W. W. Norton.

Freeman, K. (1999). No services needed?: The case for mentoring high-achieving African American students. *Peabody Journal of Education, 74*(2), 15–26.

Girves, J. E., Zepeda, Y., & Gwathmey, J. K. (2005). Mentoring in a post-affirmative action world. *Journal of Social Issues, 61*(3), 449–479.

Glass, J. C. Jr., & Garett, M. S. (1995). Student participation in a college orientation course, retention, and grade point average. *Community College Journal of Research and Practice, 19*(2), 117–132.

Grunder, P. G., & Hellmich, D. M. (1996). Academic persistence and achievement of remedial students in a community college's college success program. *Community College Review, 24*(2), 21–33.

Harper, S. R., & Quaye, S. J. (2009). Student engagement in higher education: Theoretical perspectives and practical approaches for diverse populations. New York: Routledge.

Hollins, T. N. (2009). Examining the impact of a comprehensive approach to student orientation. *Journal of the Virginia Community Colleges, 15–27*.

Johnson, C. S. (1989). Mentoring programs. In M. L. Upcraft & J. Gardner (Eds.), *The freshman year experience: Helping students survive and succeed in college (pp. 118–128). San Francisco: Jossey-Bass.*

Keeling, R. P., Wall, A. F., Underhile, R., & Dungy, G. J. (2008). Assessment reconsidered: Institutional effectiveness for student success. International Center for Student Success and Institutional Accountability, USA.

King, P., & Howard-Hamilton, M. (2000). Using student development theory to inform institutional research. *New Directions for Institutional Research, 108*, 19–36.

Kinzie, J. & Pennipede, B. S. (2009). Converting engagement results into action. *New Directions for Institutional Research, 141*, 83–96.

Kuh, G. D. (2004). Forging a new direction: How UTEP created its own brand of excellence. *About campus,* 9(5), 9–15.

Kuh, G. D., Cruce, T. M., Shoup, R., Kinzie, J., & Gonyea, R. M. (2008). Unmasking the effects of student engagement on first-year college grades and persistence. *Journal of Higher Education, 79*(5), 540–563.

Kuh, G. D., Kinzie, J., Buckley, J., Bridges, B., & Hayek, J. C. (2007). Piecing together the student success puzzle: Research, propositions, and recommendations. *ASHE Higher Education Report, 32*(5).

Laden, B. V. (1999). Socializing and mentoring college students of color: The Puente Project as an exemplary celebratory socialization model. *Peabody Journal of Education, 74*(2), 55–74.

LaNasa, S. M., Olson, E., & Alleman, N. (2007). The impact of on-campus student growth on first-year student engagement and success. *Research in Higher Education, 48*(8), 941–966.

Mangold, W. D. (2003). Who goes, who stays: An assessment of the effect of a freshman mentoring and unit registration program on college persistence. *Journal of College Student Retention, 4*(2), 95–122.

National Resource Center for the First-Year Experience and Students in Transition. (2008). National Resource Center Survey. Retrieved from http://www.sc.edu/fye/research/surveyfindings/

Nelson Laird, T. F., Smallwood, R., Niskode-Dossett, A. S., & Garver, A. K. (2009). Effectively involving faculty in the assessment of student engagement. *New Directions for Institutional Research, 141*, 71–81.

Nora, A., & Crisp, G. (2007). Mentoring students: Conceptualizing and validating the multi-dimensions of a support system. *Journal of College Student Retention: Research, Theory and Practice, 9*(3), 337–356.

O'Gara, L. Karp, M. M., & Hughes, K. L. (2009). Student success courses in the community college: An exploratory study of student perspectives. *Community College Review,* *36*(3), 195–218.

Ouimet, J. A., & Pike, G. R. (2008). Rising to the challenge: Developing a survey of workplace skills, civic engagement, and global awareness. New Directions for Institutional Research, Assessment Suppl. 2007, 71–82.

Pagan, R., & Edwards-Wilson, R. (2003). A mentoring program for remedial students. *Journal of College Student Retention, 4*(3), 207–225.

Pascarella, E. T. (2001). Identifying excellence in undergraduate education: Are we even close? *Change, 33*(3, May/June), 18–23.

Pascarella, E. T., Terenzini, P. T., & Wolfle, L. M. (1986). Orientation to college and freshman year persistence/withdraw decisions. *Journal of Higher Education, 57*(2), 155–175.

Pickering, J., & Sharpe, M. (2000). Slicing the pie: Institutional research, assessment, and student affairs research. *New Directions for Institutional Research, 108*, 79–89.

Price, D. V., with research support from Lee, M. (2005). *Learning communities and student success in postsecondary education*. A background paper. December. MDRC, New York.

Raymond, L., & Napoli, A. R. (1998). An examination of the impact of a freshman seminar course on student academic outcomes. *Journal of Applied Research in the Community College, 6*(1), 27–34.

Richburg-Hayes, L., Visher, M. G., & Bloom, D. (2008). Do learning communities affect academic outcomes? Evidence from an experiment in a community college. *Journal of Research on Educational Effectiveness, 1*(1), 33–65.

Roberts, A. (2000). Mentoring revisited: A phenomenological reading of the literature. *Mentoring and Tutoring, 8*(2), 145–170.

Ross-Thomas, E., & Bryant, C. E. (1994). Mentoring in higher education: A descriptive case study. *Education, 115*(1), 70–76.

Scrivener, S., Bloom, D., LeBlanc, A., Paxson, C. E., & Sommo, C. (2008). A good start: Two-year effects of a freshmen learning community program at Kingsborough Community College. MDRC, New York.

Shultz, E. L., Colton, G. M., & Colton, C. (2001). The Adventor Program: Advisement and mentoring for students of color in higher education. *Journal of Humanistic Counseling, Education, and Development, 40*, 208–218.

Smith, K., & Mather, P. (2000). Best practices in student affairs research. *New Directions for Institutional Research, 108*, 63–78.

Sorrentino, D. M. (2007). The SEEK mentoring program: An application of the goal-setting theory. *Journal of College Student Retention, 8*(2), 241–250.

Stovall, M. L. (1999). Relationships between participation in a community college student success course and academic performance and persistence. Unpublished doctoral dissertation, University of Illinois, Urbana-Champaign.

Taylor, K., Moore, W. S., MacGregor, J., Lindblad, J. (2003). What we know now. National Learning Communities Project Monograph Series. The Washington Center for Improving the Quality of Undergraduate Education at The Evergreen State College in cooperation with the American Association for Higher Education, Washington, DC.

Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition (2nd ed.)*. Chicago: University of Chicago Press.

Tinto, V. (1998). Learning communities and the reconstruction of remedial education in higher education. Prepared for presentation at the "Conference on Replacing Remediation in Higher Education" at Stanford University, January 26–67, 1998, sponsored by the Ford Foundation at the U.S. Department of Education.

Tinto, V. (2000). What have we learned about the impact of learning communities on students? *Assessment Update, 12*(2), 1–12.

Tinto, V., & Goodsell, A. (1993, April). Freshmen interest groups and the first-year experience: Constructing student communities in a large university. Paper presented at the Annual Meeting of the College Reading and Learning Association, Kansas City, MO.

Wallace, D., Abel, R., & Ropers-Huilman, B. R. (2000). Clearing a path for success: Deconstructing borders through undergraduate mentoring. R*eview of Higher Education, 24*(1), 87–102.

Weber, J. (2000). Learning communities in higher education: A field observation case study. Doctoral dissertation, Widener University, Chester, PA. (ERIC Document Reproduction Service No. ED459882).

Zhao, C. M., & Kuh, G. D. (2004). Adding value: Learning communities and student engagement. *Research in Higher Education, 45*(2), 115–138.

Zimmerman, A. (2000). A journal-based orientation course as a predictor of student success at a public two-year technical college. *Journal of the First-Year Experience, 12*(1), 29–43.

Thank You!

AIR expresses sincere appreciation for the members who serve as publications volunteers. Authors who submit materials to AIR Publications receive feedback from AIR members. The following individuals reviewed manuscripts submitted to this volume of Professional Files.

David Blair
Patrick Button
Kelly Duarte
Yao Hill
John Leonard
Terry McCamish
Tim Merrill
John Milam
Brian Moss
Kristi Nelms
Sharron Ronco
Kenneth Scott
Charles Secolsky
Suzanne Simpson
Cynthia Tweedell
C. Nevada Winrow

About AIR *Professional Files*

Biannual *Professional Files* volumes are journal-length publications grounded in relevant literature that synthesize current issues, present new processes or models, or share practical applications related to institutional research. All submissions are peer-reviewed. For more information about AIR Publications, including *Professional Files* and instructions for authors, visit www.airweb.org/publications.

Association for Institutional Research
Randy L. Swing, Executive Director
1435 E. Piedmont Drive, Suite 211
Tallahassee, FL 32308
www.airweb.org

ISSN 2155-7535