ASSOCIATION FOR INSTITUTIONAL RESEARCH
Data and Decisions for Higher Education

Back to All Proposals

PROPOSAL DETAILS

Personal Information

Name	Dr. Xueli Wang
Informal Name	Xueli
Affiliation	University of Wisconsin-Madison
Unit/Department	
Title	Assistant Professor, Educational Leadership and Policy Analysis
Year began this position	2010
Email	xwang273@wisc.edu
Preferred Mailing Address	270-H Education Building 1000 Bascom Mall Madison, Wisconsin 53706-1326 United States Phone: 608-263-5451 Fax: 608-265-3135
Secondary Address	

Financial Representative

Name	Kim Moreland
Informal Name	
Institution / Affiliation	University of Wisconsin-Madison
Unit / Department	Graduate School/Administration
Title	Director, Research & Sponsored Programs
Preferred Mailing Address	21 N. Park. St. Suite 6401
City	Madison
Country	United States
State	Wisconsin
Zip/Postal Code	53715-1218
Email	preaward@rsp.wisc.edu
Phone	(608) 262-3822
Fax	

Secondary Investigator(s)

Project Description I

Title:

Toward Viable STEM Pathways from Community Colleges to 4-year Institutions: Course-taking Patterns of Community College STEM Transfers

Overview of the Study

Course-taking patterns are a prime indicator of the academic experience and progression of college students, including STEM baccalaureate aspirants beginning at community colleges. Yet remarkably little empirical knowledge exists to illuminate viable STEM pathways for these students. The proposed study examines course-taking patterns of beginning community college students who plan to transfer into 4-year STEM majors and how these patterns are mapped against critical junctures of the students' STEM pathways, such as upward transfer and persistence and attainment at 4-year institutions. Drawing upon rich and recent postsecondary transcript data, this project will use data mining techniques that, although underutilized in higher education research, are powerful and appropriate analytical tools to investigate complex transcript data. Thus focusing on a pivotal yet extremely understudied topic dealing with postsecondary STEM education, this study will offer new insight into course and program features that help contribute to efficient and effective baccalaureate STEM pathways for interested community college students.

National Context and Importance of the Study

In recent years, researchers and policymakers have grappled with how to address the shortage of college students pursuing science, technology, engineering, and mathematics (STEM) degrees (Fox, 2003; Hagedorn & DuBray, 2010; Hagedorn & Purnamasari, 2012; Lowell & Salzman, 2007). While the demand continues to rise as the United States looks to maintain a global competitive edge both academically and economically, there exists a serious deficit in the number of students entering STEM areas of study and successfully completing college degrees in these fields (Dowd, 2011; Espinosa, 2011). In order to tackle this problem, numerous policies and initiatives have focused on improving STEM participation, persistence, and completion at American postsecondary institutions.

The role of community colleges in this important endeavor cannot be overemphasized. Nationally, nearly 1,200 community colleges enroll over eight million students annually, including 43% of all undergraduates (American Association of Community Colleges, 2011; Knapp, Kelly-Reid, & Ginder, 2009). Community colleges also serve a disproportionately large number of minority students (Cohen & Brawer, 2008) who are underrepresented in STEM fields but represent the country's STEM future. Traditionally underappreciated, these pubic 2-year institutions have the potential to help increase the number and diversity of students pursuing STEM degrees (Hagedorn & Purnamasari, 2012). With their transfer function, community colleges also have the capacity to assist students to continue their education at 4-year colleges and universities and succeed in a STEM baccalaureate.

Recently, the pivotal role of community colleges in STEM education has gained national attention. In 2011, the National Academy of Sciences, the National Science Foundation, and the Carnegie Institution for Science co-sponsored the summit on Community Colleges in the Evolving STEM Education Landscape, which highlighted the vital role these public 2-year institutions can play in expanding the educational opportunities for students pursuing STEM degrees and occupations. With this policy priority in mind, it is critical to identify viable STEM educational pathways facilitated by community colleges.

Although a fair amount of empirical attention has been devoted to STEM education focusing on high schools and 4-year institutions (e.g., ACT, 2006; Crisp, Nora, & Taggart, 2009; Porter & Umbach, 2006; Wao, Lee, & Borman, 2010), less research has been conducted on the role of community colleges and how they may assist students to transfer into and succeed in STEM areas of study at baccalaureate-granting institutions. In particular, we know virtually nothing about what course-taking trajectories are followed by transfer-aspiring community college students who are interested in studying STEM. Nor do we have much information on which course-taking pathways align with successful transfer and eventual attainment at 4-year schools. As Dowd (2011) rightfully pointed out, "the body of literature focusing specifically on transfer in STEM is not robust enough to substantiate conclusions about the unique programmatic features that are necessary to design effective STEM transfer pathways" (p. 122).

Therefore, I propose to study the course-taking trajectories of beginning community college students and the resulting STEM outcomes. By exploring and identifying patterns of course-taking that are common to those students who successfully transfer and obtain a baccalaureate in STEM, this study can illuminate course pathways STEM-aspiring students navigate through community colleges and 4-year institutions (for those who successfully transfer upward) and which particular trajectories are most viable for STEM transfer and eventual success at baccalaureate institutions. This knowledge will assist educators and policy makers to improve curriculum and program offerings and strengthen intercollegiate course and program articulations, thus facilitating effective and efficient STEM educational pathways for interested community college students.

Review the literature and establish a theoretical grounding for the research:

STEM Transfer from Community Colleges to 4-Year Institutions

A small but growing body of research has touched upon community college transfer into STEM fields at 4-year institutions. A handful of studies have focused on the required academic preparation. Hagedorn and DuBray (2010) highlighted the appropriate amount of preparation from community colleges, which is a critical factor in successful STEM transfer given that many students attending community colleges already underperform in essential STEM courses such as math and science (Bragg, 2011). The courses needed to perform at the level required for transfer into STEM fields have a tendency to become gatekeeper courses, discouraging many students from pursing this route (Hagedorn & DuBray, 2010; Packard, 2011). Therefore, scholars have called for community colleges to improve math and science courses and student progress in them to promote STEM transfer and baccalaureate achievement (Hagedorn & Purnamasari, 2012; Hoffman, Starobin, Laanan, & Rivera, 2010). Yet, there is little agreement on what constitutes "necessary" and "appropriate" course preparation to achieve this goal. Furthermore, we have little empirical knowledge as to which course-taking patterns help students progress in an efficient manner and adequately meet the rigor of STEM course requirements.

Looking at the STEM transfer issue more holistically involving both 2-year and 4-year institutions, Bensimon and Dowd (2012) argued that course and program alignment and articulation through the collaboration between community colleges and 4-year institutions can increase STEM transfer rates. However, there remains poor collaboration between community colleges and universities and significant issues of curricular distrust (Dowd, 2011; Gabbard et al., 2006). Clearly, additional research exploring factors associated with curricular alignment and articulation is critical, especially in STEM fields; otherwise, this problem will continue to prevent students who begin at community colleges from obtaining baccalaureate degrees in STEM. One promising approach would be to explore STEM-aspiring community college students' course-taking patterns and their connection to successful transfer and baccalaureate attainment. This knowledge will inform both 2-year and 4-year institutions in their efforts to set up curricular and articulation agreements to facilitate STEM transfer and completion.

Several other studies focused on measures, such as outreach programs, to improve STEM transfer (e.g., Bensimon & Dowd, 2012; Packard, 2011). For example, by informing students of options to pursue a STEM degree, outreach has been suggested as a means to improve STEM transfer from community colleges to 4-year institutions (Packard, 2011). Measures like this, however, have not pointed to how actual course-based pathways influence students' transfer and progress toward a STEM baccalaureate, and no concrete or consistent results on a larger scale exist to inform curricular improvement and change (Dowd, 2011; Packard, 2011).

In summary, a limited number of studies have explored various characteristics that help or hinder community college students in their pursuit of a 4-year degree in STEM. Yet, these studies invariantly neglect a key dynamic—students' actual course-taking pathways along the STEM pipeline and what course patterns conducive to STEM transfer, persistence, and attainment look like. Should this knowledge gap continue, the empirical research base on STEM transfer would offer only limited insight into potential opportunities for designing and structuring viable academic offerings and pathways. By increasing our knowledge on course-taking patterns, we can effectively inform a number of current curricular concerns surrounding STEM transfer, such as lack of articulation of coursework, lengthy remedial course sequences, and the separation of special programs from the core curriculum (Dowd, 2011).

Conceptual Grounding of the Research

The proposed study is grounded in the tradition of Astin's (1991, 1993) input-environment-outcome (I-E-O) model, one of the most enduring and widely cited college impact models (Pascarella & Terenzini, 2005). Although the I-E-O model assumes that students' outcomes also depend on the characteristics that they bring to college, Astin (1993) uses the model to focus more on the influences of the educational environment on student outcomes net of other non-college factors. The educational environment is composed of curricular and instructional features, as well as other college activities students are exposed to. Like most studies anchored in the I-E-O conceptual framework, the proposed study focuses on the environmental influences—namely academic coursework—over which institutions have programmatic and policy control.

The study also draws upon Bahr's (2013) deconstructive approach to understanding community college students' pathways and success that calls for an in-depth understanding of how students progress or fail to progress through community colleges. Bahr breaks down the actual process in which students progress in their coursework toward any given end, whether it is an actual qualification, transfer, subject competency, or other intended educational goals. As Bahr forcefully argued, without understanding how students actually progress through their college programs, "institutional adjustments and interventions will be more a product of guesswork than of sound and empirically-based reasoning" (p. 13).

Taking into consideration both Astin's (1993) and Bahr's (2013) conceptual approaches, this study focuses on course-taking as the educational environment transfer-aspiring community college students are exposed to in their pursuit of a STEM baccalaureate. Although co-curricular experiences are also highlighted in the I-E-O model, I argue that for beginning community college students, the college environment intersects substantially with the classroom. Given that community colleges are largely commuter schools and many community college students primarily engage with the college through coursework, exposure to courses represents the primary college environment, and course-taking behaviors and patterns are the primary indicators of student engagement with this academic environment. Indeed, as Hagedorn and Kress (2008) argued, for many community college students the only trace of their presence is found in the transcripts. As a whole, a student's transcript serves as a map of the curriculum—the principal college environment traveled by students. When analyzed appropriately, course-taking patterns may offer valuable insight into a student's academic history and momentum through college and illuminate patterns that effectively and wisely engage academic resources.

Describe the research method that will be used:

Research Questions

Three questions drive the study: What distinctive or frequent course-taking patterns exist among STEM baccalaureate aspiring students beginning at community colleges? How do these identified course-taking patterns contribute to upward STEM transfer? Among all community college students who transfer into 4-year STEM majors, what are viable course-based pathways to a STEM baccalaureate?

Data, Sample, and Measures

This study will use data from the Beginning Postsecondary Students Longitudinal Study (BPS:04/09) and Postsecondary Education Transcript Study (PETS:09). Following a nationally representative first-time beginning cohort (N=18,610) in 2003–2004, BPS:04/09 contains survey data at three points in time: in respondents' first year of college and then 3 and 6 years after they first started in postsecondary education. Of critical importance to the proposed study, transcripts were collected under PETS:09 from all 3,030 eligible postsecondary institutions attended by the BPS respondents over a 6-year period. Of the eligible institutions, 2,620 (87%) provided transcripts for a total of 16,960 students. The transcript data are invaluable in studying course-taking patterns, credit transfer, and the link among course-taking, institutional transfer, and baccalaureate persistence and attainment in STEM.

To answer the first two research questions, the sample is restricted to beginning postsecondary students at community colleges who were in a STEM field of study when first enrolled in 2003–2004 and who expected to earn a bachelor's degree or above. Of the BPS panel respondents, 5,550 began at a public 2-year institution and among these community college entrants, nearly 500 were in STEM fields and aspired to earn at least a bachelor's degree. To address the third question, all 2-year to 4-year STEM transfer students (despite their reported educational goals and fields of study during the first year) are retained, resulting in an analytical sample of 480.

PETS:09 will provide most of the measures for this inquiry. These include detailed transcript records at the student-, course-, term-, degree-, institution-, and transfer-level. Three "outcome" measures indicating the success along the STEM transfer pathway are of interest:

(a) TRANSFER: whether STEM baccalaureate aspiring students beginning at community colleges transferred into a 4-year STEM major;

(b) PERSISTENCE: among all community college students who transfer into 4-year STEM majors, whether students were enrolled as a STEM major at a 4-year institution as of 2009; and,

(c) ATTAINMENT: whether students attained a baccalaureate in a STEM field as of 2009.

Finally, the study includes demographic variables for cross-group comparison of course patterns. See Appendix A for a complete variable list.

Data Analysis

Use of data mining techniques. To best answer the research questions, I will primarily rely on data mining techniques (specifically, frequent pattern/association rule mining). Data mining is a form of exploratory analysis that extracts implicit and useful patterns and relationships from massive quantities of data rather than testing pre-formulated hypotheses (Han, Kamber, & Pei, 2011; Luan & Zhao, 2006). Data mining and traditional statistical procedures both can perform association and prediction analysis, but, for my study, data mining is more appropriate to analyzing course-taking patterns. This is because of the complex, seemingly unstructured nature of transcript data in PETS:09, which holds tens of course records for each student over numerous academic terms. It would be extremely challenging to apply traditional parametric analysis to make sense of the wealth of such data. By employing data mining, however, meaningful and frequent patterns and sequences among the course records will be teased out and information such as the optimal combination of remedial work, for-credit STEM courses, and general education courses becomes evident and can be used for subsequent analysis to offer a foundation for educational change.

Data formulation. Because the transcript data were collected from a bundle of 3,030 postsecondary institutions, each with a different course numbering system, finding course patterns without standardizing the course numbering system would be impossible. Therefore, to reduce these course records into a manageable dataset, each student's course-taking records will first be classified and aggregated according to their corresponding top-level Classification of Instructional Programs taxonomy (2-digit CIP code). Aggregated credit hours of each CIP category will be further "discretized," where data of continuous value type are put into equal interval or percentile sets. Table 1 in Appendix B is an example of the coding and discretization schema.

As this study will mine both sequential and non-sequential course patterns, data will be prepared in both transactional and tabular formats. The original PETS:09 course dataset is in a transactional format for sequential mining, where each course represents one unique record (see example in Table 2 of Appendix B). To construct tabular data for non-sequential mining, the original vertical and repeating (per student) course data format will be converted into a horizontal and unique record format for each student (see example in Table 3 of Appendix B).

Analytical procedures. After data preparation, the study will proceed as follows: First, for STEM baccalaureate aspiring students beginning at community colleges, the study will analyze their course-taking patterns (both sequential and non-sequential) by applying the "frequent pattern/association rule data mining" (Han et al., 2011) for discovery of salient associations and correlations among student course-taking records. (A technical description of this technique is in Appendix C.) At this stage, the mining of course-taking patterns is single dimensional, i.e., only course data are involved. In the second stage, the identified course-taking patterns will be linked to other data dimensions such as educational outcomes. Thus, after mining the frequent course-taking patterns, I will perform multi-dimensional association rule mining to examine whether certain patterns (or sequences) of course-taking are more conducive STEM transfer. These procedures address the first two research questions.

Using similar strategies, the third phase of the study will include community college students who transfer into 4-year STEM majors and map their course-taking trajectories in both 2-year and 4-year institutions. In this step, pre-transfer and post-transfer course data are separated, and frequent course-taking patterns will be identified and linked to STEM baccalaureate persistence and attainment.

I will use Microsoft Access for relational database manipulation and transformation and SPSS Modeler (formerly Clementine) for transcript data mining.

Uploaded Appendix Document(s):

- Appendix A B C

Project Description II

Will you use NCES target dataset? Yes

Please check all NCES datasets that apply
- Beginning Postsecondary Student (BPS) Longitudinal Study and Transcript Data

Explain why each dataset best serves this research. Include a variable list for each dataset used.

As noted, this study will use data from the BPS:04/09 and PETS:09. BPS:04/09 contains survey data on demographic characteristics, college experiences, major fields of study, persistence, transfer, and degree attainment. PETS:09 includes comprehensive transcript data spanning 6 years. Together, BPS:04/09 and PETS:09 represent an invaluable and ideal resource to study course-taking patterns, credit transfer, and the link among course-taking, institutional transfer, and baccalaureate persistence and attainment in STEM among recent baccalaureate-aspiring students who begin at community colleges and who are interested in STEM fields of study. The list of variables to be used in the study is in Appendix A.

Will you use NSF target dataset? No

Explain why each dataset best serves this research. Include a variable list for each dataset used.

Will you address the NPEC focus topic? Yes

If yes, please briefly describe:
This study addresses the 2013 NPEC focus topic—The Impact of Data on the College Search and Selection Process—by focusing on the importance of transcript data in identifying the most efficient and effective course-taking patterns in the STEM pipeline across community colleges and 4-year institutions. The emphasis on transcript data, available to both institutions and individual students and their families, allows a readily available yet underutilized, rarely-analyzed data source to help facilitate the identification of effective educational pathways in STEM or other domains. Institutions can use results from this research to improve their curriculum and program offerings and strengthen intercollegiate course and program articulations to help develop viable STEM pathways. This information will then help students and families choose postsecondary institutions, 2-year and/or 4-year, that best serve their academic, educational, and financial needs and interest. This information may be particularly helpful for beginning community college students looking to transfer into 4-year STEM majors.

Project Description III

Provide a timeline of key project activities:

2013 Activities	
May–June	· Receive funding and prepare data
July–September	· Complete data preparation in ACCESS · Perform descriptive analysis
October–December	· Conduct and complete data mining · Prepare and submit mid-year progress report to AIR
2014 Activities	
January–February	· Write the results and discussion sections · Refine literature review and research methods sections as needed
March	· Complete a research manuscript based on the study and present the study at AERA, contingent upon proposal acceptance · Publish the manuscript as a working paper on the website of the Wisconsin Center for the Advancement of Postsecondary Education (WISCAPE)
April	· Finalize the manuscript based on feedback from University of Wisconsin–Madison and WISCAPE colleagues and AERA · Submit the manuscript to *Research in Higher Education*
May–June	· Present the refined manuscript at AIR annual forum · Prepare and submit final report to AIR

List deliverables such as research reports, books, and presentations that will be developed from this research initiative:

The following deliverables will result from this research initiative:

- Conference presentations at the 2014 AERA and AIR annual meetings;

-Research reports to AIR, including mid-year progress report and final report;

-Working paper(s) to be published on the website of the Wisconsin Center for the Advancement of Postsecondary Education (WISCAPE);

-Research article(s) to be submitted for publication in *Research in Higher Education* and/or *The Journal of Higher Education*.

Describe how you will disseminate the results of this research:

Manuscripts based on this study will be submitted for publication in *Research in Higher Education* and/or *The Journal of Higher Education*. In addition, the research papers will be disseminated as working papers through the Wisconsin Center for the Advancement of Postsecondary Education (WISCAPE) where I serve as a scholar. All WISCAPE papers will be distributed via the center's website, e-newsletter (3,000+ subscribers), targeted e-mail announcements, and limited print production. As noted, results will also be presented at national conferences, including AIR and AERA.

Provide a reference list of sources cited:

ACT. (2006). *Developing the STEM education pipeline*. Iowa City, IA: ACT.

American Association of Community Colleges. (2011). Community college trends and statistics. Retrieved from http://www.aacc.nche.edu/AboutCC/Trends/Pages/default.aspx

Ammon, B. V., Bowman, J., & Mourad, R. (2008). Who are our students?: Cluster analysis as a tool for understanding community college student populations. *Journal of Applied Research in the Community College*, 16, 32–44.

Astin, A. W. (1991). *Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education*. Phoenix, AZ: The Oryx Press.

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

Bahr, P. R. (2013). The deconstructive approach to understanding community college students' pathways and outcomes. *Community College Review, 41,* in press.

Bahr, P. R. (2010). The bird's eye view of community colleges: A behavioral typology of first-time students based on cluster analytic classification. *Research in Higher Education*, 51, 724–749.

Bensimon, E. M., & Dowd, A. C. (2012). *Developing the capacity of faculty to become institutional agents for Latinos in STEM*. Los Angeles, CA: University of Southern California.

Bensimon, E. M., Dowd, A. C., Chase, M. M., Sawatzky, M., Shieh, L. T., Rall, R. M. & Jones, T. (2012). *Community college change agents at HSIs: Stewarding HSI-STEM funds for Latino student success in STEM*. Los Angeles, CA: University of Southern California.

Bragg, D. D. (2011). Two-year college mathematics and student progression in STEM programs of study. Paper presented at Community Colleges in the Evolving STEM Education Landscape: A summit, Washington, DC.

Cohen, A., & Brawer, F. (2008). *The American community college* (5th ed.). San Francisco, CA: Jossey-Bass.

Crisp, G., Nora, A., & Taggart, A. (2009). Student characteristics, pre-college, college, and environmental factors as predictors of majoring in and earning a STEM degree: An analysis of students attending a Hispanic Serving Institution. *American Educational Research Journal*, *46*(4), 924–942.

Dowd, A. C. (2010). Improving transfer access for low-income community college students. In A. Kezar (Ed.), *Recognizing and serving low-income students in postsecondary education: An examination of institutional policies, practices, and culture* (pp. 217–231). New York, NY: Routledge.

Dowd, A. C. (2011). *Developing supportive STEM community college to four-year college and university transfer ecosystems*. Los Angeles, CA: University of Southern California.

Espinosa, L. L. (2011). Pipelines to pathways: Women of color in undergraduate STEM majors and the college experiences that contribute to persistence. *Harvard Educational Review*, *81*(2), 209–240.

Fox, M. A. (2003). *Pan-organizational summit on the U.S. science and engineering workforce: Meeting summary*. Washington, DC: National Academies Press.

Gabbard, G., Singleton, S., Macias, E. E., Dee, J., Bensimon, E. M., Dowd, A. C., et al. (2006). *Practices supporting transfer of low-income community college transfer students to selective institutions: Case study findings*. Boston, MA and Los Angeles, CA: University of

Massachusetts Boston and University of Southern California.

Gonzalez, J., & DesJardins, S. (2002). Artificial neural networks: A new approach to predicting application behavior. *Research in Higher Education 43*(2), 235–258.

Hagedorn, L. S., & DuBray, D. (2010). Math and science success and nonsuccess: Journeys within the community college. *Journal of Women and Minorities in Science and Engineering*, *16*(1), 31–50.

Hagedorn, L. S., & Kress, A. M. (2008). Using transcripts in analyses: Directions and opportunities. *New Directions for Community Colleges, 143*, 7-17.

Hagedorn, L. S., & Prather, G. (2005). The community college solar system: If university students are from Venus community college students must be from Mars. Paper presented at the 2005 Annual Forum of the Association for Institutional Research, San Diego, California.

Hagedorn, L. S. & Purnamasari, A. V. (2012). A realistic look at STEM and the role of community colleges. *Community College Review*, *40*(2), 145–164.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *SIGKDD Explorations*, *11*(1), 10–18.

Han, J, Kamber, M., & Pei, J. (2011). *Data mining: Concepts and techniques* (3rd ed.). Waltham, MA: Morgan Kaufmann.

Hoffman, E., Starobin, S. S., Laanan, F. S., & Rivera, M. (2010). Role of community colleges in STEM education: Thoughts on implications for policy, practice, and future research. *Journal of Women and Minorities in Science and Engineering*, *16*(1), 85–96.

Knapp, L. G., Kelly-Reid, J. E., & Ginder, S. A. (2009). Enrollment in postsecondary institutions, Fall 2007; graduation rates, 2001 & 2004 cohorts; and financial statistics, fiscal year 2007 (NCES 2009-155). Washington, DC: National Center for Education Statistics.

Lowell, B. L., & Salzman, H. (2007). Into the eye of the storm: Assessing the evidence on science and engineering education, quality, and workforce demand. Retrieved from http://www.urban.org/publications/411562.html

Luan, J., & Zhao, C. (2006). Practicing data mining for enrollment management and beyond. In J. Luan & C. Zhao (Eds.), *New Direction for Institutional Research, 131*. San Francisco, CA: Jossey-Bass.

Packard, B. W. L. (2011). Effective outreach, recruitment, and mentoring into STEM pathways: Strengthening partnerships with community colleges. Paper presented at National Academy of Science Meeting, "Realizing the Potential of Community Colleges for STEM Attainment," Washington, DC.

Pascarella, E .T., & Terenzini, P. T. (2005). *How college affects students: A third decade of research*. San Francisco, CA: Jossey-Bass.

Porter, S. R., & Umbach, P. D. (2006). College major choice: An analysis of person-environment fit. *Research in Higher Education*, *47*(4), 429–449.

VanDerLinden, K. (2002). *Credit student analysis: 1999 and 2000*. Annapolis Junction, MD: Community College Press, American Association of Community Colleges.

Wao, H. O., Lee, R. S., & Borman, K. (2010). Climate for retention to graduation: A mixed methods investigation of student perceptions of engineering departments and programs. *Journal of Women and Minorities in Science and Engineering*, *16*(4), 293–318.

IRB Statement

Statement of Institutional Review Board approval or exemption:

The IRB at the University of Wisconsin–Madison specifies that research projects involving analysis of secondary data from NCES do not require prior IRB approval.

Statement of Use of Restricted Datasets

This study will employ restricted-use data from the 2004/09 Beginning Postsecondary Students Longitudinal Study (BPS:04/09) and its supplementary Postsecondary Education Transcript Study (PETS:09). I am currently authorized to access the restricted-use BPS and PETS data through a site license at the School of Education of UW–Madison.

Biographical Sketch

Xueli Wang's Biography Sketch

Dr. Xueli Wang is an assistant professor in the Department of Educational Leadership and Policy Analysis at the University of Wisconsin–Madison and a scholar at the Wisconsin Center for the Advancement of Postsecondary Education (WISCAPE). She holds a Ph.D. in Higher Education and a graduate minor in quantitative research methods from The Ohio State University. Wang's current research deals with several issues pertaining to community colleges and their students, including baccalaureate aspirations, transfer access to 4-year institutions, participation in STEM fields of study, educational pathways, as well as the role of community colleges in STEM education. Wang's work has been published or is forthcoming in highly regarded journals such as *Teachers College Record*, *The Journal of Higher Education*, *Research in Higher Education*, *Journal of College Student Development*, and *Community College Review*. As Principal Investigator, Wang received a 2007–2008 Dissertation Fellowship from AIR/NCES/NSF/NPEC, a 2011–2012 AIR/NSF research grant, and a 2012–2014 AERA research grant. She is also co-Principal Investigator and Research Director of a large-scale NSF funded research project that examines educational outcomes in manufacturing engineering technologist and technician education programs in Wisconsin's 2-year technical college system. In summer 2011, Wang was selected as a Young Academic Fellow by the Institute of Higher Education Policy and the Lumina Foundation.

Budget Requirements

Xueli Wang' Budget

Personnel-Time on Project
%(FTE) Academic Year: 25.00
%(FTE) Summer: 37.45

Personnel-Salary & Benefits
Academic Year: $ 99107.00
Summer: $ 29402.00

Graduate Research Assistant's Budget

Personnel-Time on Project
%(FTE) Academic Year: 0.00
%(FTE) Summer: 0.00

Personnel-Salary & Benefits
Academic Year: $ 0.00
Summer: $ 0.00

Total Salary and Wages: $35787.80

Travel: $1386.00
Other travel related expenses: $0.00
Other research expenses: $2826.00
Total Request: $39999.80

Funding History

I have not received prior or current funding for the proposed research. I have been awarded funding from AIR for projects unrelated to this study. Specifically, I received a 2006 AIR/NSF Summer Data Policy Institute Fellowship, a 2007–2008 AIR/NPEC dissertation grant, and a 2010–2011 AIR/NSF research grant. I fulfilled all obligations required of the funding in a timely fashion. Three research articles based on my AIR-funded dissertation have been published in prestigious peer-reviewed journals such as *Research in Higher Education* and *The Journal of Higher Education*. Supported by my 2011–2012 AIR grant, I completed two research manuscripts, both under revision at top-tiered journals. Over the past 6 years, support from AIR has been highly instrumental in my development as a junior scholar. As I begin my third year on the tenure track at UW–Madison, I have established a coherent research agenda focusing on the role of community colleges in educational and workforce development. Funding for this research will therefore greatly assist me to further advance my research on the STEM pipeline from community colleges to 4-year institutions—a policy and research issue of national importance—and what is inside the course-taking "black box" that makes for efficient and viable pathways.

Appendix A. List of Variables

Variable name	Description	BPS/PETS label
Sample selection criteria		
RQ1&2: Sample restricted to respondents aspiring to earn at least a bachelor's degree and majoring in STEM fields upon entering a public 2-year institution		
	First institution type 2003–04	FSECTOR9
	Major when first enrolled in 2003–04 is in STEM fields (comparable to 2006, 2009)	MAJ04A
	Highest degree ever expected in 2003–04 is a baccalaureate or above	HIGHLVEX
RQ3: Sample restricted to beginning community college students who transferred into a STEM major at a 4-year institution		
	Transcript: Level and control transfer type	QGTRTYPE
	Transcript: Major at destination school	QGMAJRS
Outcome variables (against which course patterns are to be mapped)		
STEM transfer	Transcript: Level and control transfer type Transcript: Major at destination school	QGTRTYPE QGMAJRS
Associate's degree attainment in STEM	Highest degree attained anywhere through 2009 Transcript: Field of Study: Associate's degree or certificate program	ATHTY6Y MT11ASCT
Baccalaureate attainment in STEM by 2009	Highest degree attained anywhere through 2009	ATHTY6Y
	PETS reported field of bachelor's degree	MT11BACH
Being enrolled as a STEM major at a 4-year institution by 2009	Attainment or level of last institution enrolled through 2009 PETS reported field of study in 2009	PRLVL5Y MT11BACH
Transcript data used in data mining		
Course attributes:		
Course taker ID	Transcript : Student ID	ID
Course name	Course name	MTCRSNAM
Course number	Course number	MTCRSNUM
Term ID	Course taken term ID	MTTMID
Term start date	Course taken term start date	MTTMBEG
Term end date	Course taken term end date	MTTMEND
Institution where course was taken	Coded from institution's IPEDS ID where course was taken	MTINSTID
Transcript institution	Transcript institution's IPEDS ID (or receiving institution of transfer course)	MTTRIPDS
Course 6-digit CIP code	PETS code (i.e., CIP) for course	MTPETC
Course top-level CIP category	Course top-level (2-digit) CIP code	MTPETGEN
Credits counting toward GPA	Course credits count toward GPA indicator (1=yes, 0=no)	MTCRDCT

Variable name	Description	BPS/PETS label
Credit or clock hours indicator	Credit or clock hours (1=credit, 2=clock)	MTCRDCLK
Earned credits	Earned credits	MTCRDERN
Earned clock hours	Earned clock hours	MTHRSERN
Transfer course indicator	Course is a transfer course (1=yes, 0=no)	MTTRNSFR
STEM indicator	Course is in STEM category (1=yes, 0=no)	MTSTMFLG
Remedial course indicator	Course is remedial (1=yes, 0=no)	MTCRSREM
Vocational course attribute	Course is vocational (1=yes, 0=no)	MTCRSVOC
Course grade	Normalized grade received for course	MTNGRAD
Course quality points	Normalized quality points received for course	MTNORMQP
Demographic variables		
Respondent’s gender	Dummy variable (1=female, 0=male)	GENDER
Respondent’s race/ethnicity	Race category	RACE
Respondent’s parental education	Whether respondent is a first-generation student, recoded from parents’ highest level of education (1=yes, 0=no)	PAREDUC

Appendix B. Example Tables for Data Formulation

Table 1. *Example Coding Schema of Course-Taking Based on Top-Level CIP and Credit-Hours Category*

	Classification of Instructional Programs (CIP 2010) Top-Level Category				
Category of Credits Earned	11 (Computer sciences)	14/15 (Engineering)	26 (Biological sciences)	27 (Mathematics)	…
0	Bin CS-0: Yes/No	Bin EG-0: Yes/No	Bin BI-0: Yes/No	Bin MA-0: Yes/No	…
1~6	Bin CS-1: Yes/No	Bin EG-1: Yes/No	Bin BI-1: Yes/No	Bin MA-1: Yes/No	…
7~12	Bin CS-2: Yes/No	Bin EG-2: Yes/No	Bin BI-2: Yes/No	Bin MA-2: Yes/No	…
13~18	Bin CS-3: Yes/No	Bin EG-3: Yes/No	Bin BI-3: Yes/No	Bin MA-3: Yes/No	…
19 and above	Bin CS-4: Yes/No	Bin EG-3: Yes/No	Bin BI-4: Yes/No	Bin MA-4: Yes/No	…

Note. The CIP category represents the course subject, and the credit category represents the size of the course subject.

Table 2. *Sample Data in Transactional Format*

Student ID	Year	Course	Credits
1	2003	Math 102	3
1	2004	Biology 105	3
1	2004	Chemistry 103	3
2	2003	Math 102	3
2	2004	English 204	3
3	2003	Math 102	3
3	2003	English 204	3
3	2004	Math 202	3
3	2004	Math 302	3
3	2004	Math 333	2

Table 3. *Sample Data in Tabular Format*

Student ID	Bin MA-1	Bin BI-1	Bin CH-1	Bin EN-1	…	Bin MA-2	…
1	Yes	Yes	Yes		…		…
2	Yes			Yes	…		…
3				Yes	…	Yes	…

Note.
MA-1: Math (credit 1~6)
MA-2: Math (credit 7~12)
BI-1: Biology (credit 1~6)
CH-1: Chemistry (credit 1~6)
EN-1: English (credit 1~6)

Appendix C. Technical Details

Frequent Pattern/Association Rule Mining

In data mining, an association rule is a pattern indicating that an itemset *A* occurrence implies that another itemset *B* also occurred, i.e., $A \Longrightarrow B$, or *A* implies *B*. The occurrence count (or percentage) of itemset *A* among the universal set *U* is called support of *A*, and the occurrence count (or percentage) of itemsets *A* and *B* occurring together is referred to as *support* (*A*&*B*). A frequent pattern is an association rule whose occurrence count is greater than or equal to a minimum threshold (the so-called minimum support) set by the researcher. Therefore, an association rule is usually reported as:

$A \Longrightarrow B$ (support = *x* %, confidence = *y* %)

Where *support* ($A \Longrightarrow B$) = *Pr* (*A*&*B* occurred together), and

Confidence ($A \Longrightarrow B$) = *Pr* (*A*&*B* occurred together) / *Pr* (*A*) = $Pr(B|A)$

If an association rule satisfies both minimum support and minimum confidence, then this rule is called "strong." The thresholds of minimum support and minimum confidence are discretionarily set by the researcher, depending on how many association rules are discovered. If the number of such rules is large, the threshold can be set to a higher value. If none or too few association rules are found, the threshold can be set to a lower value. In this study, finding frequent patterns or association rules is essentially a process of searching for and counting frequency of all existing itemsets (including their subsets) from all of the available course baskets. For example, if a student's course itemset has 10 courses in the set, then this course itemset has 1,023 subsets. That is,

$$\binom{10}{1} + \binom{10}{2} + \ldots + \binom{10}{10} = 2^{10} - 1 = 1023$$

Clearly, as the amount of course data increases, this searching process becomes time-consuming. Thus, when performing association rule data mining, the basic and popular Apriori algorithm (Agrawal & Srikant, 1994) is utilized to identify the frequent patterns. In short, the Apriori algorithm is a method to improve efficiency in mining large datasets by applying the Apriori property, which states that if a given itemset does not satisfy minimum support, then none of its subsets will satisfy minimum support. Therefore, based on the Apriori property, if the course itemset does not satisfy the minimum support criterion, there is no need to check all of its subsets' frequency of occurrence. Through this property, the search process for frequent course patterns is substantially shortened. In this study, the initial minimum support will be set to 10%. As the analysis evolves, depending on the amount of identified patterns, the minimum support will be adjusted to a reasonable value for this research. An initial minimum confidence value will be set to 50%, which means that the probability of observing such an association is better than the likelihood of getting heads (or tails) in a coin toss.

Other Analytical Details of Secondary Importance to the Study

Association rule mining vs. cluster analysis

The data mining approach used in this study—frequent pattern/association rule mining—is not to be confused with clustering, a different data mining technique that has been used by a limited number of higher education scholars to study college student typologies (e.g., Ammon, Bowman, & Mourad, 2008; Bahr, 2010; Hagedorn & Prather, 2005; VanDerLinden, 2002). Cluster analysis, typically involving the *k*-means algorithm, is a process of grouping similar objects together so that the inter-cluster similarity is low and the intra-cluster similarity is high. That is, members of a cluster are more like each other than they are like members of a different cluster. Obviously, cluster analysis would be appropriate if the interest is in identifying groupings of students who share similar characteristics on multiple dimensions, which is *not* the focus of the study. Instead, this research is interested in capturing the co-occurrence of course records in large volumes of transcript data in an attempt to discover recurring relationships among course records, which makes it appropriate to employ the frequent pattern/association rule mining technique, where frequent course patterns can be identified and represented in the form of association rules.

Multi-dimensional association rule mining

After performing the primary analytical steps described in the proposal narrative, student demographic data fields will be joined to the course dataset, making the previous one-dimensional course dataset multidimensional. Accordingly, multi-dimensional association rules mining will be executed to examine whether course-taking behaviors and patterns differ among racial and gender groups. Given the disproportionately high STEM attrition rates among female students and members of underrepresented minority groups, this set of nuanced analyses will help pinpoint potential areas of intervention for these students in regard to their course trajectories.

Defining STEM Fields of Study

Finally, when defining destination fields of study for transfers and STEM persistence and attainment, both the traditional umbrella approach (i.e., lumping all STEM fields together) and a more nuanced approach accounting for subject diversity within STEM will be adopted. Although STEM disciplines share similarities, there also exist notable differences across these fields and, consequently, distinctive course-taking patterns may exist depending on the specific pathways into and through a sub area within STEM defined in a more detailed fashion.