

Design and Implementation of Web-Enabled Labs for Data Warehousing

Jiangping Wang and Janet L. Kourik

Abstract—In the distance learning environment students are often working from remote locations. All or most of their interaction will be through the online course management system. Since students cannot attend a physical lab classroom there is a need for accessible labs to serve remote students and improve their understanding of data warehouse (DW) and the textbook materials. For the DW course specifically, a platform that supports student exploration of several features such as SQL, DW, Online Analytical Processing (OLAP), as well as design of the DW is needed. A data warehousing course that meets many recommendations from industry and academia has been successfully offered for several years at our school. The DW course is offered solely in the online format using Oracle products and datasets for labs. The class incorporates a data warehouse design process that includes planning the star schema as well as defining the metadata, physical design, and application views. The paper describes the lab for the data warehousing course, its design and implementation, the exercises students will complete in both design and data analysis, and the data warehousing tool itself.

Index Terms—Web-enabled lab, data warehousing, online learning, distance education.

I. INTRODUCTION

In the distance learning environment students are often working from remote locations. All or most of their interaction will be through the online course. Since students cannot attend a physical lab classroom, accessible labs for remote students are needed to improve their understanding of data warehouse (DW) and the textbook materials. In the distance learning environment, data warehousing courses face specific challenges. In particular students need free software. The software needs to work with a variety of operating systems to provide platform independence. Further, the installation or use of DW should require no or minimal installation. Another barrier is the difficulty students must face to configure the DW software correctly for their systems [1]. For the DW course, a platform that supports student exploration of several features such as SQL, DW, Online Analytical Processing (OLAP), as well as design of the DW is needed.

Data warehousing is one of many approaches in the broader field of Business Intelligence (BI). Research reveals

rapid growth in the demand for professionals with skills in BI and analytics. Findings also indicate that there is a significant gap between demand and individuals with needed BI skills. Hence, there is a move to encourage more academic offerings to build capacity in all BI areas including data warehousing. Research also reveals that BI is truly an interdisciplinary field. Practitioners highlight that the need for analytical skills extends beyond technicians. Analytic skills are needed throughout the organization. Therefore educators have been encouraged to increase offerings that help develop BI and analytical skills for students across disciplines [2]-[4].

Prior literature often examined ways to include data warehousing in on-ground courses or independent study [5]-[7]. We have successfully offered a data warehousing course in the online format at our school that meets many recommendations from industry and academia. The DW course was created specifically for students in programs other than technology such as the MBA and a variety of MA graduate programs. Students enrolled in the course are completing degree programs in many disciplines such as marketing, finance, regulatory compliance, leadership, and healthcare quality management. The DW course is only offered in the online format using Oracle products and datasets for labs. The class incorporates a data warehouse design process that includes defining the star schema, metadata, physical design, and application views. The remainder of the paper describes the lab for the data warehousing course, its design and implementation, the exercises students will complete in both data analysis, design, and the data warehousing tool itself.

II. ONLINE CLASS AND ITS AUDIENCE

The data warehouse course primarily serves working adults who typically have prior work experience and need to take the course online. Whenever labs are needed to support the course learning objectives it is beneficial for students to have hands-on experience with a data warehousing system [8]. Students in the DW course are mostly students in less technical degree programs and may have little or no experience in installing and configuring software. In the past, we found that often the required software was not compatible with students' computers. Students with less technical experience may find it difficult to identify and communicate the problems they are experiencing. The result is that both instructors and students spend more time on technical support rather than course content. Further, we find that a number of students rely heavily on the computers belonging to others or employers. Students may not be permitted to install any software on computers in the workplace. In the nine-week

Manuscript received March 10, 2013; revised June 24, 2013.

The authors are the George Herbert Walker School of Business and Technology, Webster University, St. Louis, MO 63119 USA (e-mail: wang@webster.edu, kourikjl@webster.edu).

accelerated format it can be very difficult for students to download, install, and configure the software in a time to complete the first assignments in the course.

One solution to these problems is to deliver the DW capabilities via the Web. If a student’s computer system can run a web browser to reach and use the DW, many of the downloading, installation and configuration problems can be avoided. Compatibility issues may still occur depending on which web browser is used and its version. However, the compatible browser versions can be clearly identified. Further, most distance education programs require access to a computer system that can support a list of selected browser versions. An instructor need only be familiar with a few browsers instead of many operating systems and a variety of hardware. As a result configuration problems can be minimized and in turn, students can dedicate more time to learning.

III. CLASS COMPONENTS AND ONLINE LABS

The data warehousing class is designed to provide a foundation in data warehouse technology by examining the way DW are designed, managed, and used. The course explores the fundamental concepts related to data warehouses [9]. Major topics covered in this class are the business justification for a data warehouse, management of a data warehouse project, the design, construction, and operation of a data warehouse, issues of data quality and performance, and using the data warehouse.

Just as in other computing oriented courses, implementing the curricula depends heavily on technology and hands-on labs. Labs are design to clarify reading materials and help students understand fundamental principles. An online lab for a data warehousing class should support a basic SQL environment, access to different schema, and a graphical user interface for easier use by non-majors.

The components in Table I address the learning objectives. Course work is guided by weekly learning objectives and each week’s reading will be enhanced by related lab modules. Student learning is assessed by weekly assignments, a mid-term exam and a final exam that include both reading-related questions and lab-related questions.

TABLE I: CLASS COMPONENTS AND LAB ENGAGEMENT

| Components | Class Components | Lab Engagement |
|------------|--|------------------------------|
| Objective | ----- Weekly learning objectives ----- | |
| Activity | Textbook reading | Hands-on lab module |
| Discussion | Reading related questions | Lab related discussion forum |
| Assignment | Reading related | Lab related |
| Test | Reading related exam questions | Lab related exam questions |

Two sample data warehouses are used throughout the class, one for learning in the labs and another for tests, so students can apply what they’ve learned to new data warehouse designs. When students learn the star schema as a design model for a data warehouse, seeing a live data warehouse schema populated with data will help them easily understand the relationship between fact and dimension tables and how the “star” is constructed by tables, primary key and foreign

keys. Hands-on labs also provide a way to assess students understanding on concepts. Discussion forum questions can ask students to identify dimensional hierarchies in a live data warehouse.

IV. LAB STRUCTURE

Oracle Application Express (APEX) has enabled us to successfully deliver online labs for the online data warehousing classes while the literature more often reports using Microsoft Access, Microsoft SQL Server, or Teradata [1], [3]-[8]. Oracle APEX is a web-based environment that provides capabilities for both SQL manipulation and web application development. It contains a comprehensive browser-based development environment that supports both single SQL statements and SQL scripts. The development environment enables students to use only a web browser yet build data warehouse projects. They do not need to install any software other than a compatible web browser on their computers.

The hands-on portion of the class consists of a series of lab modules that cover topics from data warehousing, OLAP, and web-driven web analytics. The structure is shown in the following Fig. 1.

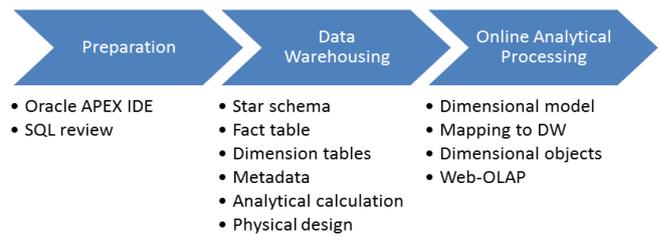


Fig. 1. Lab structure.

A. Preparing Students for Business: Orientation on Tools

In this section of the lab, students are guided to familiarize themselves with the web lab environment. Students are also asked to review their SQL skills and focus on how to use SQL to answer basic business questions. Samples of the questions used during the lab are:

- What is the present customer base?
- What are top product groups as measured by item gross revenue?
- What are the sales measured by sales dollar per customer per product?
- How much revenue did new products generate by customer demographics?
- What are sales statistics by products and by product categories?

Operational database systems are not optimized to store and retrieve data directly for answering questions such as these nor are they optimized for decision making. Operational databases are designed for running day-to-day transactions. If the operational database were used, the resources needed for analytic queries would vie for the resources used to process transactions.

Normalization is implemented during the design process on operational databases to reduce data anomalies such as errors during insertion, updating, and deletion. Problems

with operational data are discussed including complex SQL, complex joins, too many joins, non-integrated data, difficulties in getting hierarchical aggregation, and difficulties in pivoting data. These problems set the stage for introducing the Star schema warehouse design because existing normalized operational databases do not yield a structure that serves advance data analysis requirements well.

The ultimate purpose for this section is twofold. First, students gain the understanding of potential issues involved in answering business questions in operational database with standard SQL. Second, students are introduced to the need for data warehouse to answer complex business questions and make business decisions.

B. Preparing Students for Business: Data Warehouse Design

The primary topic students explore in this section of the course is the star schema as a dimensional modeling technique for data warehouses. Critical business measures, such as sales and revenue, are captured and summarized from multiple business perspectives known as dimensions. The star schema focuses on dimensions and facts that map to business metrics. Advantages of star schema include simple implementation, easier querying for complex business problems, and, for many tools, a user interface that is more intuitive to users.

The sample data warehouse star schema is populated from a transactional system containing data for a fictional electronics store. The operational data have already been cleaned by the extracting, transforming, and loading (ETL) process. The sample data warehouse is designed with a fact table and 4 dimensional tables, as depicted in Fig. 2. The fact table, SALES_FACT, contains 2.8 million rows of data and the biggest dimension table, CUSTOMERS, hosts over 253 thousand records.

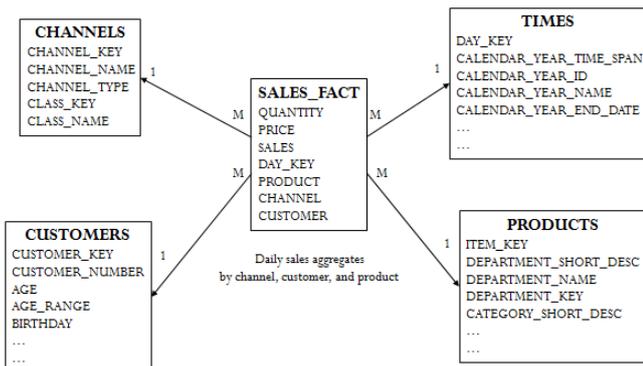


Fig. 2. Sample star schema.

Students often have difficulty envisioning the contents of the dimension table and their use in queries. In the lab, students can switch to the sample data warehouse schema and browse star schema tables, so that they can actually experience data and how they are organized. For example, students can query and see that dimension tables are wide (with large number of attributes) and shallow (with fewer of records), while the fact table is narrow (with only a few attributes) and deep (with large number of records).

The reference to a “wide” dimension table is based on displaying the table with columns and rows, the table has

many columns, and therefore it is wide horizontally. TIMES dimension table shown in Fig. 2 can be used to demonstrate a wide dimension with its list of 50 attributes, in addition to the key for days, DAY_KEY. Fig. 3 lists some of these attributes in the sample data warehouse.

Wide dimension tables in the star schema design represent hierarchies for subtotals. Multiple hierarchies provide for analysis at different levels throughout the dimension. For example the TIMES dimension includes hierarchies such as calendar year, fiscal year, and retail year. Analysis can be performed by drilling down along the calendar year hierarchy from all years, to one calendar year, calendar quarter, or month. Another hierarchy supports drill-down by fiscal year hierarchy from all years, one fiscal year, a fiscal quarter, or month. Similarly, alternative drilldown analysis can be achieved along retail year hierarchy from all years, to retail year, retail season, retail quarter, retail month, and retail week. This hierarchy effectively replaces the traditional five-year sales book maintained by most retail stores.

| TIMES | |
|-------------------------------|---------|
| PK | DAY_KEY |
| CALENDAR_YEAR_TIME_SPAN | |
| CALENDAR_YEAR_ID | |
| CALENDAR_YEAR_NAME | |
| CALENDAR_YEAR_END_DATE | |
| CALENDAR_QUARTER_NAME | |
| CALENDAR_QUARTER_TIME_SPAN | |
| CALENDAR_QUARTER_END_DATE | |
| CALENDAR_QUART_NUMBER_IN_YEAR | |
| CALENDAR_QUARTER_ID | |
| MONTH_TIME_SPAN | |
| MONTH_NAME | |
| MONTH_END_DATE | |
| MONTH_ID | |

Fig. 3. Partial attribute list in TIMES dimension.

TABLE II: SAMPLE RESULTS FROM DRILL DOWN EXERCISES

| | | | | | |
|--------------------------|-----------------------|---------------|--|--|--|
| Query 1 result | SUM(QUANTITY) | 20052 | | | |
| Query 2 result | CALENDAR_QUARTER_NAME | SUM(QUANTITY) | 01-CY2008 5259 02-CY2008 4217 03-CY2008 4642 04-CY2008 5934 | | |
| Query 3 result (partial) | CALENDAR_QUARTER_NAME | TYPE_NAME | SUM(QUANTITY) | 01-CY2008 Camcorder Accessories 5240 01-CY2008 Camcorders 19 02-CY2008 Camcorder Accessories 4199 02-CY2008 Camcorders 18 03-CY2008 Camcorder Accessories 4636 03-CY2008 Camcorders 6 | |
| Query 4 result (partial) | CALENDAR_QUARTER_NAME | TYPE_NAME | COUNTRY_NAME | SUM(QUANTITY) | 01-CY2008 Camcorder Accessories Canada 183 01-CY2008 Camcorder Accessories Cuba 31 01-CY2008 Camcorder Accessories Mexico 714 01-CY2008 Camcorder Accessories United States 4312 01-CY2008 Camcorders Mexico 1 |

Students learn how to answer business questions by analyzing sales price along attributes in the various dimension tables. They also learn how to implement attributes in the dimension tables as constraints and use filters in the queries. The following example queries are used to help students learn how to perform drilldown in summary numbers in order to explore the details at lower levels. Table II show some sample results from the following queries.

- Query 1: How many units in the Camcorder category were sold in North America for the calendar year 2008?
- Query 2: Drill down to the level of quarter (3-months) in 2008 for the same region and product category.
- Query 3: Go down to the level of individual product subgroups in the Camcorder category.
- Query 4: Go to the level of details by individual countries

in the North America region.

Other built-in interactive features that can be manipulated to make the report suit the user’s analytical needs, including layout and format options, filtering, selecting columns, control breaks, computing and aggregating, searching designated fields, report pagination, and conditional display to name a few. Working with interactive reports, students users can customize data presentation and perform OLAP operations on their own without the involvement of application designers.

C. Preparing Students for Business: Analytical Reporting

Although standard SQL for relational databases can perform basic aggregation operations, the performance for complex analytic calculations is poor. OLAP needs to support complex calculations, complex analytical processing, and “what-if” analysis. To improve performance, analytical operations are performed by an OLAP engine that prepares the derived data for storage as values in a logical cube.

Enhanced aggregate and analytic SQL functions have been introduced, making advanced analyzing and reporting significantly easier. Some examples of advanced aggregation functions include ROLLUP, CUBE, GROUPING, GROUPING SETS, RANKING, and relative contribution functions, as listed in Table III.

TABLE III: EXAMPLES OF ADVANCED AGGREGATION FUNCTIONS

| Function | Description |
|-----------------|--|
| ROLLUP | Group the selected rows and return a single row summary for each group. |
| CUBE | Group the selected rows based on the values of all possible combinations and return a single row summary for each group. |
| GROUPING SETS | Specify multiple groupings of data and prune the unneeded aggregates. |
| RANK | Calculate the rank of a value in a group of values. |
| RATIO_TO_REPORT | Compute the ratio of a value to the sum of a set of values. |

Students complete a series of activities in the lab to practice each advanced aggregation functions on the sample data. Business reports are created using aggregate functions to meet user requirements such as the following examples.

- Business Requirement: Compute totals for different quarters and a grand total in the same report. Action: Write and run a query with aggregate function ROLLUP.
- Business Requirement: Combine selected subtotals in a data cube, but not all possible subtotals. Action: Write and run a query with aggregate function GROUPING SETS.
- Business Requirement: Add a column to show the rank of any revenue number within the quarter. Action: Write and run a query with aggregate function RANK.

| QUARTER | RANKING | CHANNEL | CATEGORY | REVENUE | PERCENT |
|-----------|---------|----------|----------------------------|-----------------|---------|
| Q1-CY2007 | 1 | --all-- | All Computer Furniture | \$17,153.20 | 0.1 |
| Q1-CY2007 | 2 | --all-- | PDAs | \$23,425.86 | 0.1 |
| Q1-CY2007 | 3 | --all-- | Total Server Computers | \$459,045.60 | 1.5 |
| Q1-CY2007 | 4 | --all-- | Cameras and Accessories | \$1,014,613.66 | 3.3 |
| Q1-CY2007 | 5 | --all-- | Camcorders and Accessories | \$1,160,413.50 | 3.7 |
| Q1-CY2007 | TOTAL | Direct | TOTAL | \$49,480,488.76 | 20.6 |
| TOTAL | - | Indirect | TOTAL | \$70,855,269.66 | 29.4 |
| TOTAL | - | --all-- | All Computer Furniture | \$62,522.00 | 0.1 |
| TOTAL | - | --all-- | PDAs | \$77,040.25 | 0.1 |

Fig. 4. Partial results of analytical reporting.

The above series of activities guide students’ learning in this section and provides practice with each specialized

aggregate function. The partial results of a final analytical report are presented in Fig. 4, with all the specified business requirements met.

D. Preparing Students for Business: On-Line Analytical Processing

OLAP plays a unique role by serving as the bridge between end users and their data. OLAP presents data to the user in a format that facilitates multidimensional-hierarchical analysis. OLAP enables users to access, manage, analyze, and present decision support data from the DW to support their business intelligence needs. Using the DW, in turn, reduces the demand on the operational transaction system.

In the OLAP related lab modules, students examine the OLAP dimensional model and dimensional objects, such as cubes, hierarchies, levels, and attributes, as shown in Fig. 5. Oracle OLAP provides three types of views: cube view, dimension view, and hierarchy view. The cube view serves the same function as fact tables and hierarchy views while dimension views serve the same function as star schema dimension tables. The relationship among these views can be discovered through OLAP metadata. Students create and practice basic SQL queries to answer business questions. For example, students learn how to apply a filter to every dimension since the cube contains both detail level and aggregated data to prevent returning more data that users need.

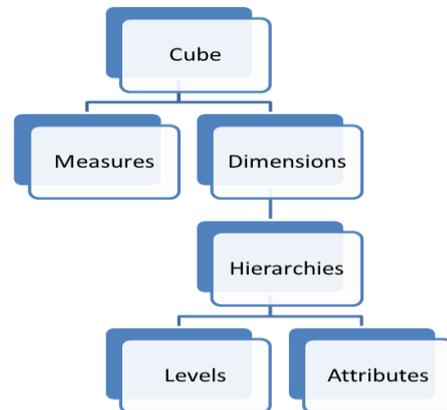


Fig. 5. OLAP dimensional model objects.

Web-enabled OLAP allows analysis to be performed via a common Web browser. It enhances accessibility with considerably lower investment and training costs. A Web browser will be the only client software needed for users to perform analytical processing. Therefore, Web-enabled OLAP can directly impact the delivery of decision support data and the way that users retrieve, analyze and share business intelligence information. OLAP seamlessly increases the access to information with lower burden on users, lower communication costs, and no specialized software being required on the local computer. During this section of the lab students are assigned to use a web-enabled OLAP application implemented via Oracle APEX’s interactive report application, which has features such as drill-down reporting, dynamic report customization, and chart generation [10]. Students can create customized OLAP queries and reports independently, without the involvement of application designers.

V. OUTCOME ASSESSMENT

Student assessments of the online DW labs indicate the online labs are perceived as helpful in learning textbook material. A student commented on the lab that “viewing real business examples was an extreme benefit.” Another student commented that “I enjoyed working with the Oracle APEX tools and was interested with the OLAP tool.” The student was pleased to learn about the logical design and to write queries in the Oracle APEX accessing the dimension and hierarchical views. “It was important to take the conceptual view and enhance the understanding with the practical lab.” Students feel they are able to use the SQL skills to run queries when needed and “will also be able to better understand the background of the work” and “the architecture of the data warehouse”.

Student assessment and course evaluation data also support the narrative evidence. Students have been more successful in the course using web-based labs for more than a half dozen offerings.

VI. FUTURE DIRECTIONS

Potential areas for further exploration include exploring the relationship between student experience with hands-on labs and the ability of students to foster the organizational buy-in that is so important to effective use of a data warehouse in industry [11]. The labs may be reframed and modeled as a cloud service to increase exposure essential cloud computing concepts. While this course is designed for graduates that are not technology major, it would be interesting to serve students from a broader range of disciplines.

VII. CONCLUSIONS

In response to a growing need for the shortage of BI and analytics, industry and academia are encouraging more course offerings in the subject area. The BI and analytics community has encouraged dissemination of such curricular and educational practices. Thus we were motivated to share the design and implementation of an existing data warehouse course that has been successful. The course serves online students from many disciplines and provides extensive hands-on experience with industrial scale tools and datasets. Because the course is offered online, a significant challenge has been a practical way to provide hands-on experience with a needed level of realism in the course labs.

Labs should be made more accessible to student without introducing the burden of installing and configuring the DW environment. Better hands-on lab design remains a challenging curricular component for data warehousing courses. Online web-based labs provided a way for students to experience applied learning. Such technical details detract from the primary course content. More time for online labs in data warehousing and OLAP greatly improves course quality

and enhances students’ learning. This paper offers a contribution to educators interested in teaching data warehousing by detailing the design and implementation of a hands-on course. A particular area of emphasis has been the way in which we successfully provided realistic hands-on experience for online students.

REFERENCES

- [1] B. Wixom, T. Ariyachandra, M. Goul, P. Gray, U. Kulkarni, and G. Phillips-Wren, “The Current State of Business Intelligence in Academia,” *Communications of the Association for Information Systems (CAIS)*, vol. 29, pp. 299-312, 2011.
- [2] L. Farmer, “Data Mining Technology Across Academic Disciplines,” in *Proc. the iConference 2011*, Seattle, WA, USA, 2011, pp. 503-507.
- [3] H. J. Watson and B. H. Wixom, “The Current State of Business Intelligence,” *Computer*, vol. 40, no. 9, pp. 96-99, 2007.
- [4] B. Wixom, H. J. Watson, O. Marjanovic, and T. Ariyachandra, “Educating the Next-Generation BI Workforce,” *The Business Intelligence Journal*, vol. 15, no. 3, pp. 26-31, 2010.
- [5] R. Fang and S. Tuladhar, “Teaching Data Warehousing and Data Mining in a Graduate Program of Information Technology,” *Journal of Computing Sciences in Colleges*, vol. 21, no. 5, pp. 137-144, 2006.
- [6] R. McFadyen, “Data Warehousing in an Undergraduate Curriculum,” *Journal of Computing Sciences in Colleges*, vol. 22, no. 4, pp. 221-227, 2007.
- [7] E. Udoh, “Teaching Database in an Integrated Oracle Environment,” in *ITiCSE-WGR '06 Working Group Reports on ITiCSE on Innovation and Technology in Computer Science Education*, New York, NY, USA, pp. 71-74, 2006.
- [8] N. Jukic and P. Gray, “Using Real Data to Invigorate Student Learning,” *ACM SIGCSE Bulletin*, vol. 40, no. 2, pp. 6-10, 2008.
- [9] M. Murray and M. Guimaraes, “Expanding the Database Curriculum,” *Journal of Computing Sciences in Colleges*, vol. 23, no. 3, pp. 69-75, 2008.
- [10] J. Wang, J. L. Kourik, and P. Maher, “Web-Enabled Online Analytical Processing for Intelligent Business Decision-Making,” in *Proc. the 10th Wuhan International Conference on E-Business*, 2010, pp. 592-597.
- [11] R. Bose, “Advanced Analytics: Opportunities and Challenges,” *Industrial Management & Data Systems*, vol. 109, no. 2, pp. 155-172, 2009.



Jiangping Wang is an associate professor of computer science at Webster University. He has a B.A. from Chongqing University, Chongqing, China, an M.S. from the University of Leeds, Leeds, United Kingdom and a Ph.D. from the Missouri University of Science and Technology, Rolla, Missouri, USA. Dr. Wang's areas of teaching include database design, database applications, data warehousing, web databases, database in web services, and distributed application development. His areas of research include database management systems, decision support systems, business intelligence, e-commerce data processing, and software project management.



Janet L. Kourik is a professor in the Mathematics and Computer Science Department of Webster University in St. Louis, Missouri, US. She has a B.S.C.S. from Webster University, an M.A. from Webster University and a Ph.D. from Nova Southeastern University. Dr. Kourik's areas of teaching include database concepts and applications, information systems, operating systems, and distributed systems. Her areas of research include databases and analytics, agile methods, informatics, and computer science education. She has been published in the *Journal of Business and Educational Leadership*, *Journal of Research in Higher Education*, as well as national and international conference proceedings.