

An Assessment Database for Supporting Educational Research

Mark Urban-Lurain, Diane Ebert-May, Jennifer Momsen, Ryan McFall, Matthew B. Jones,
Ben Leinfelder, Jon Sticklen
urban@msu.edu, ebertmay@msu.edu, momsenj@msu.edu, mcfall@hope.edu, jones@nceas.ucsb.edu,
leinfelder@nceas.ucsb.edu, sticklen@msu.edu

Abstract - One of the challenges of research in science education is storing, managing and querying the large amounts of diverse student assessment data that are typically collected in many Science, Technology, Engineering and Mathematics (STEM) courses. Furthermore, longitudinal studies across courses and ABET accreditation necessitate tracking students throughout their academic programs in which each course will have different types of data. Researchers need to manage, assign metadata to, merge, sort, and query all of these data to support instructional decisions, research and accreditation. To address these needs we have constructed a database to support both data-driven instructional decision making and research in STEM education. We have built upon existing metadata standards to define an extensible Educational Metadata Language (EdML) that enables assessments to be tagged based on taxonomies, standard psychometrics such as difficulty and discrimination, and other data to facilitate cross-study analyses. Once a collection of assessment data are available, faculty can examine their assessment data to evaluate historical trends, analyze the effectiveness of pedagogical techniques and strategies, or compare the performance of different teaching and assessment techniques within their course or across institutions.

Index Terms - Assessment Database, Infrastructure, Interdisciplinary, STEM Education Research

INTRODUCTION

Universities throughout the nation are under increasing pressure to provide evidence of learning outcomes using assessment metrics that move beyond course grades. Addressing calls for improving education with evidence-based outcomes requires systematically collecting and analyzing amounts of data on a scale not previously required. Faculty who are engaged in the scholarship of teaching and learning and data-driven instructional design need tools to support the collection, management and analysis of assessment data. Institutional researchers need to manage a wide variety of data about student learning. However, these data are often poorly structured and stored in a wide variety of ad hoc formats that are distributed

across unconnected systems. These factors impede collecting, managing and analyzing data from educational research.

One of the challenges of data-driven instructional decision making and research in STEM education is storing, managing and querying the large amounts of diverse student assessment data that are typically collected in many STEM courses. Assessments may include handwritten assignments, quizzes in a learning management system, multiple choice “bubble sheet” exams, classroom “clickers” data, lab or programming assignments, and many other assessments. Furthermore, longitudinal studies across courses and ABET accreditation necessitate tracking students throughout their academic programs in which each course will have different types of data. Researchers need to manage, assign metadata to, merge, sort, and query all of these data to support instructional decisions, research and accreditation.

There are a large number of assessment databases (e.g., NSDL, MERLOT, FLAG) but these are “item” databases, sources of assessments that faculty may use, many of which have metadata about their subjects, grade levels, difficulty, etc. However, these databases do not support storing student performance data on these assessments nor do they support the sharing of data about student performance for research purposes. To address these needs we have constructed a database to support both data-driven instructional decision making and research in science education. We modeled our work after the Long Term Ecological Research (LTER) network which facilitates the collection, management and sharing of ecological data. We believe that sharing data on student learning is an important part of advancing educational research and improving STEM education.

PRIOR WORK

Faculty Institutes for Reforming Science Teaching (FIRST) has engaged faculty from over 50 institutions in professional development focused on active, inquiry-based teaching designed to improve student learning. We built on the student assessment data and the national FIRST network to construct a prototype, proof-of-concept database to support research on undergraduate education. The database supports storing, searching and analyzing assessment data and

DATABASE DESIGN

facilitates both data-driven instructional decision making and research in STEM education (<http://first.ecoinformatics.org>).

We partnered with the National Center for Ecological Analysis and Synthesis (NCEAS) and Dr. Ryan McFall [1] at Hope College to build upon existing metadata standards [e.g., Ecological Metadata Language [2], IMS (<http://www.imsglobal.org>), Dublin Core Metadata Initiative (<http://dublincore.org>)] to define an extensible Educational Metadata Language (EdML). EdML provides a language to describe a wide variety of assessment data types and metadata about those assessments. This enables us to tag assessments based on taxonomies, standard psychometrics such as difficulty and discrimination, and other data to facilitate cross-study analyses [3]. We also extended an existing metadata and data storage server, Metacat [4], as well as an existing metadata creation application, Morpho [5], in order to accelerate development of our prototype system.

MOTIVATION

An example from a large introductory Bioscience course at Michigan State University (MSU) is illustrative of need for such an assessment database. A team of faculty and researchers began studying student learning in BS111 in 2005. Data collected include:

- Student demographic data from the registrar (GPA, ACT scores, major, class standing, gender, ethnicity, etc.)
- Multiple choice exam data (word processing files with the exam questions, student response data from scantron sheets read by the scoring office.)
- Homework assignments and data administered using an online course management system.
- In-class “clicker” questions and student response data.
- Classroom observation data about instruction
- Data from interviews with students
- Classification of assessment items by content (i.e., cellular respiration, photosynthesis) and by intellectual level using Bloom’s taxonomy [6].

Research questions focus on the instructional changes predicting improved student understanding; e.g., what assessments effectively reveal higher-level student understanding in large enrollment courses? To answer questions about student learning and the impact of instructional changes, we constructed datasets that merged these data in various ways to permit analyses such as comparing data across different sections of the course, comparing data across different semesters, comparing different types of assessment questions on the same topic, and comparing different instructional interventions. As in any research, most analyses prompted new questions, which in turn required different datasets. Each dataset was constructed manually, requiring hours of effort and data management expertise absent among most faculty.

A large variety of assessment data, collected across the nation, informed our database design specifications for developing both the Educational Metadata Language to describe the assessment data, and to test and evaluate the database back-end and front-end tools. We generated a large number of use cases and research questions from which the functional requirements were derived.

The data system consists of several linked software systems that together provide an intuitive and efficient means of managing educational assessment data from faculty across the nation. Features of the database system satisfy a series of use cases developed by educators and researchers to inform the system design. The components of the system (shown in Figure 1) include:

- A data system that stores assessment data and relevant metadata about the context in which the assessment occurred: courses, institutions, assessment questions, and responses.
- Desktop software to collect and document assessment data for faculty. The system includes wizards for collecting course information, and an efficient Assessment import tool that reads documents in PDF format and extracts the relevant question items from the text for direct insertion into the data system, without retyping the questions or laborious cut and paste operations. The import tool also reads student response data from common data sources like spreadsheets and data extracted from common course management systems like Moodle, Blackboard, and WebCT.
- A searchable, web-based application for locating assessment data and seeing useful data summaries from the system. Users can query the system based on a variety of metadata to generate downloadable datasets for analysis.
- EdML, an exchange schema for assessment data and metadata for passing information among components of the database system and to external applications.

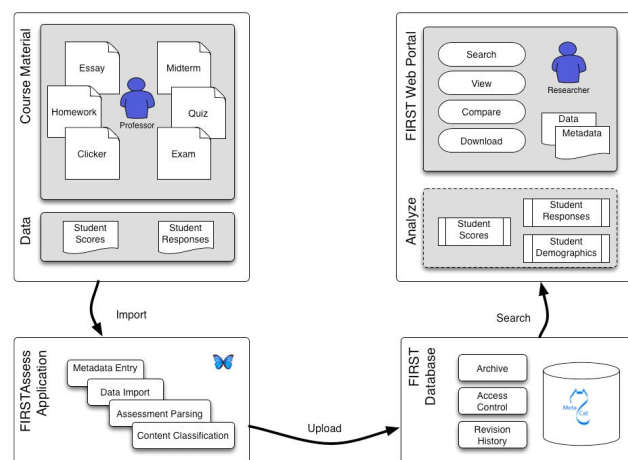


FIGURE 1
DATABASE STRUCTURAL OVERVIEW

Once a collection of assessment data is available, faculty can examine their assessment data to evaluate historical trends, analyze the effectiveness of pedagogical techniques and strategies, or compare the performance of different teaching and assessment techniques.

In the next sections, we elaborate on the major components of the database system.

STORING ASSESSMENTS

The first step in uploading student assessment data is to upload the assessments from which the data was obtained. We designed the database to comply with the IMS Question Test Interoperability (QTI) standard which specifies the exchange of item, assessment and results data between authoring tools, item banks, learning systems and assessment delivery systems. Our goal is to allow faculty members to create assessments in software that is QTI compliant and have the database import the QTI conforming assessments.

However, most faculty exams are still prepared using word processing software, or, if they are prepared using course management or other assessment software that is QTI compliant, the resulting assessment is a printed exam. To support the primary / lowest common denominator format, we created a tool to convert assessments from PDF to QTI. Using PDF as the input to this tool allows faculty to create assessments using virtually any editing application. This software reads a PDF of an assessment and parses the file to extract individual questions, which is what are stored in the database. Since getting the the conversion completely correct is a difficult problem, the user is provided a GUI to correct parsing errors, edit items, add metadata to the items, and save the resulting data package on the database. The user interface is shown in Figure 2.

While PDF is a convenient format, there are some aspects of PDF that make the conversion process difficult. As the name Page Description Format implies, the document contains only the information needed to render the document on a page. For our application, notably absent is structure information such as how lines of text are grouped into paragraphs. Style information such as bold, underline and super/subscript is not readily available. Furthermore, there is no information which questions tables and figures are associated with. Tables are actually composed of unrelated text and graphics – on the printed page the human eye naturally groups these together, but the conversion tool is given no hints that the objects making up the table should be semantically grouped together. These difficulties are instances of the general machine recognition problem of *segmentation*. In order to strike a balance between ease of use and overly incorrect conversion, our tool asks users for input on some of the more complex tasks, such as grouping primitives like text and lines together into tables, and determining pages of the assessment which should not be included in the database (such as pages of equations).

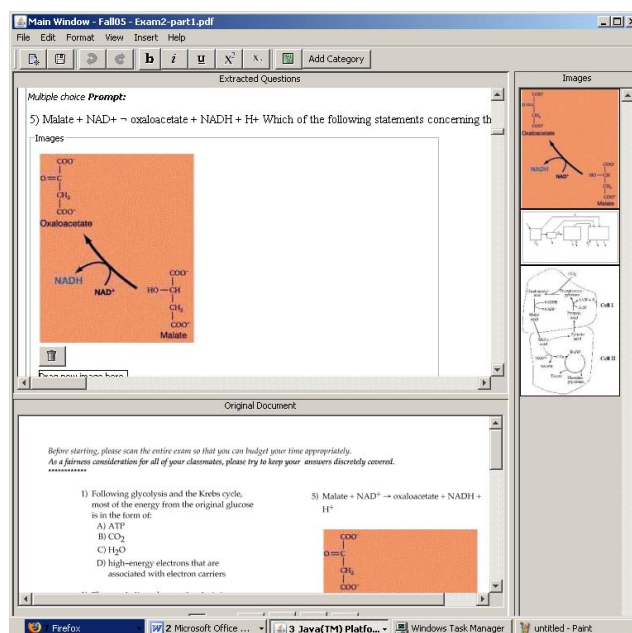


FIGURE 2

ASSESSMENT PARSER INTERFACE. THE BOTTOM PANE DISPLAYS THE ORIGINAL PDF OF THE ASSESSMENT. THE TOP PANE DISPLAYS THE ITEMS THE PARSER HAS EXTRACTED. THE RIGHT PANE CONTAINS IMAGES THAT THE PARSER HAS EXTRACTED.

Even with a perfect segmentation algorithm, there are still areas of ambiguity that must be dealt with. A particularly prominent example is that in many cases, it is difficult to correctly distinguish between a **multiple-choice** question and a question that consists of **multiple parts**. Dependencies between questions on a assessment, such as when a common set of choices are used for a block of multiple choice questions, comprise a second example of ambiguity. These dependencies are frequently, but not always, explicitly mentioned in the assessment.

While there is always room for improvement, the conversion process works reasonably well. Simple multiple choice assessments are usually convert with few to no errors, while more diverse assessments have taken on average 30 or more minutes to convert.

METADATA

To promote the classification and searching of assessments to support STEM education research, users tag assessment items with metadata. We built upon existing metadata standards (e.g., Ecological Metadata Language, IMS, Dublin Core) to define an extensible Educational Metadata Language (EdML). We broadly categorize the metadata as two types: *disciplinary metadata* that addresses particular domain knowledge and general *educational metadata* that includes all other non-disciplinary metadata. Since every discipline has unique metadata and these metadata may have many formats, we define a general metadata structure to support plugin metadata modules. In biology, for example, we use the National Biological Information Infrastructure (NBII) Biocomplexity Thesaurus (<http://thesaurus.nbio.gov>)

which is a tree-structured taxonomy of biological terms consolidating several other thesauri. NBII provides a web-services Simple Object Access Protocol (SOAP) interface with which the assessment parser communicates to provide the user with biological terms. The Assessment Parser uses this Application Programming Interface (API) to provide the user with the ability to broaden/narrow a metadata keyword search in the same fashion as the native NBII interface. In this way, users familiar with the NBII vocabulary can quickly categorize their assessments.

Other disciplinary metadata frameworks can be accommodated, either via their APIs, or by defining the necessary content in an XML file. For example we are creating disciplinary metadata for our first year engineering sequence. These courses use both a custom published text and a standard commercial textbook. The tables of contents and indices of these books provide the foundation for the metadata for these courses. We are also incorporating ABET criteria into the engineering metadata. By tagging assessments with the relevant ABET metadata, it will be possible to provide detailed analysis of student performance as outcomes evidence for accreditation.

The general *educational metadata* framework is extensible and can include standard psychometrics (e.g., difficulty and discrimination), Bloom's Taxonomy of Educational Objectives, Professional society tags, whether an assessment item is copyrighted (as in a textbook question or a published concept inventory) and other user-defined metadata to support any research questions.

The *educational metadata* also includes information about courses and institutions in which the assessments were administered. These metadata facilitate longitudinal studies in addition to comparisons among courses across institutions.

DATA SOURCES

We wanted the database to be independent of course management software or other restrictions that would inhibit use. Because of the diverse data streams that are idiosyncratic to institutions, disciplines and instructors, we designed the database to accommodate a wide range of assessments types and sources (see Figure 3).

Assessments can be broadly classified in two broad categories: closed- and open-ended. Closed-ended assessments have fixed choices (i.e., multiple choice, true false, matching, etc.) from which the student selects responses. The parser will recognize most types of closed-ended assessment questions.

Open-ended assessments require that students construct, rather than select, their answers. Essays, design problems, mathematics problems, programming or lab assignments are examples. For open-ended assessments, the instructor must provide a rubric by which the assessments are scored. The next version of the database will store the rubric to allow faculty to upload files containing the students' responses (essays, scans of handwritten responses, etc.) and allow

faculty to upload scores for each student that were assigned based on the rubric.

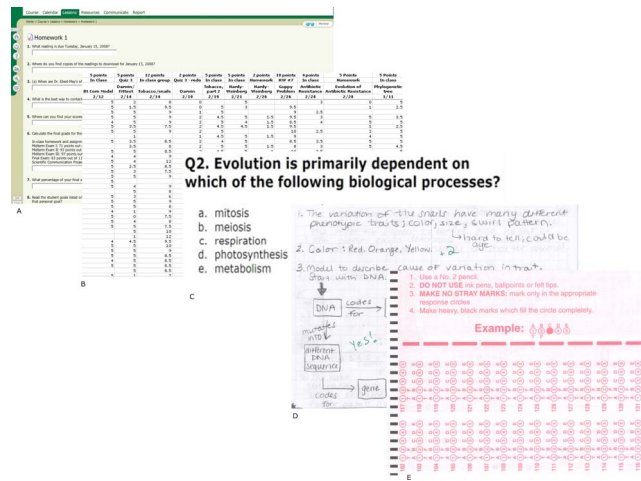


FIGURE 3
SAMPLE DATA SOURCES FROM A FACULTY MEMBER, INCLUDING: (A) OUTPUT FROM COURSE MANAGEMENT SOFTWARE, (B) EXCEL SPREADSHEETS OF STUDENT EXAM, HOMEWORK AND IN-CLASS ASSIGNMENTS, (C) CLICKER DATA, (D) STUDENT OPEN-ENDED RESPONSES AND (E) SCANTRON EXAM SHEETS.

Most courses have data for student responses on multiple assessments. When uploading data, each student is identified by a unique identifier, usually assigned by the institution. The software uses hashing functions to encrypt the unique identifier so that the student ID is not stored in the database. The encrypted identifier allows the database to merge multiple datasets so that it is possible to query data across assessments, or even across courses within an institution, and receive datasets with the data for individual students merged in the returned dataset. To maintain security, the encrypted identifier is never returned to any user in a query.

QUERYING THE DATABASE

Users query the database using a web interface (Figure 4). The interface uses a “shopping cart” model, allowing the user to browse for assessment items based on metadata, institution, type of assessment and other parameters. Once the user identifies possible assessments, it is possible to examine the metadata in the web interface to determine if the assessment items are of interest. The user adds items to, or removes items from, the shopping cart and when finished, the data query is run and the database creates a CSV file containing all of the data selected, along with metadata about the items.

By focusing your analysis at the item level, you can concentrate on assessments that address the particular topics (evolution and modeling) to understand student learning in detail. By comparing the performance of students in your course with students in courses at other universities, it is possible to understand the impact of different instructional environments or interventions.

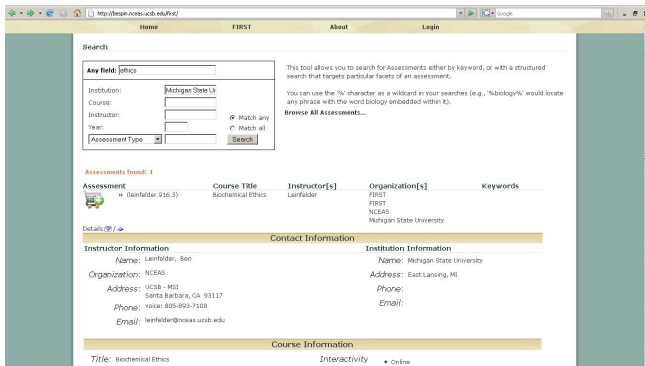


FIGURE 4
DATA QUERY INTERFACE

The dataset includes data for all students to the item level. The data includes the student responses (in the case of closed ended items) and the scores the students received on the items (Figure 5). These data may then be imported into whatever analytic software the user wishes to use.

The screenshot shows a spreadsheet titled 'sample_student_question.xls'. The data is organized in columns: institution, course, year, term, assessment, question, concept, bloomKater_A, bloomKater_B, studentID, response, and score. The rows represent individual student assessments across various institutions (MSU, UCSB) and courses (PLB203, BIO222) from 2007 to 2008. The 'response' column contains alphanumeric strings representing student answers, and the 'score' column shows the corresponding scores.

FIGURE 5
EXAMPLE CSV FILE OUTPUT

SAMPLE RESEARCH ANALYSIS

Imagine you are teaching a unit on evolution and modeling and want to know whether teaching students to make models improved their learning of evolution. You query your current course and two past courses for data on evolution and modeling. The database returns a single spreadsheet file with the data from each course (Figure 5). You import the data into your favorite statistical analysis software for analysis.

As a correlated question, you are interested in comparing your students' modeling abilities with those of students from other institutions. You again search the database for courses at other colleges or universities using model-based education to teach evolution. You limit your query to include only large introductory biology courses at institutions similar in size to yours. You then limit the returned data to include only assessments related to modeling in evolution. The database again returns a spreadsheet that you can import into statistical software for analysis.

IRB, SECURITY AND INTELLECTUAL PROPERTY

Institutional review boards (IRB) are unique to each college and university. This reality demands a flexible database that is both secure and protective of student data. The Assessment Database supports two levels of data access - restricted (available only to the faculty member who uploaded the data) and public (available to all registered database users). As with any human subject research, the IRB at the each researcher's institution will need to determine what level of data sharing is appropriate on a project-by-project basis.

To ensure confidentiality of data, unique student identifiers (usually the school student ID number) are de-identified using an encryption function that ensures student anonymity. Only the encrypted identifier is stored in the database. To support longitudinal studies the database can link students across courses at an institution based on the encrypted identifier. Neither the unique student identifiers nor the encrypted identifiers are ever returned to any user in a query. This prevents reverse engineering the identifiers. However, this also means that researchers must query the database for complete datasets. It is not possible to "rejoin" the multiple queries from the database as there is no unique identifier that can be used to join different queries.

For publicly released data (data that are available to researchers other than those who uploaded it) a faculty member may choose or be directed by their institution's IRB to remove identifying information about themselves and the institution from the resulting dataset.

Many assessment items are copyrighted - as in textbook questions or published concept inventories. Faculty members using the database must be aware of - and take responsibility for - obtaining appropriate permissions for copyrighted or otherwise protected material. Without such permissions, an assessment item cannot be part of the public database. For more information on copyright issues and intellectual property rights, please see our website (<http://first.ecoinformatics.org>).

CURRENT AND FUTURE STATUS

At the time of writing, we have just released Alpha release 3 of the database software and have begun pilot testing at Michigan State University with data from several biology and engineering courses and with institutional research data such as tests of scientific and quantitative literacy that are administered to incoming students at orientation. These tests are intended to verify the end-to-end data flow: parsing data, tagging with metadata, upload student data, and

querying the results. We will use feedback from this testing to refine the user interfaces and metadata frameworks.

We are seeking additional funding to continue development of the database. We envision a distributed system with individual nodes installed at institutions to support local faculty and institutional research at each institution. Each institution and researcher, with approval from their IRB, would decide what data to publish for public access. Each node would then make its public data available to be part of a federated database that would be open to interested researchers.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under awards 0618501, 0353566 and 0088847. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation (NSF). We would also like to thank our REU participants who did much of the work on the exam parser: Lidiya Ilcheva, Daniel Savoie, Garrison Benson, Bridger Hamilton, Eleanor Poley and Gabriel Larriuz.

REFERENCES

- [1] R. McFall, M. Urban-Lurain, and D. Weinshank, "A Web-to-database system for collecting student data," in *Frontiers in Education*, Boston, MA, 2002, pp. F1E14-F1E18.
- [2] E. Fegraus, S. J. Andelman, M. B. Jones, and M. Schildhauer, "Maximizing the value of ecological data with structured Metadata: An introduction to Ecological Metadata Language (EML) and principles for metadata creation," *Bulletin of the Ecological Society of America*, vol. 86, pp. 158-168, 2005.
- [3] W. K. Michener, J. W. Brunt, J. J. Helly, T. B. Kirchner, and S. G. Stafford, "Nongeospatial metadata for the ecological sciences," *Ecological Applications*, vol. 7, pp. 330-342, 1997.
- [4] M. B. Jones, C. Berkley, J. Bojilova, and M. Schildhauer, "Managing scientific metadata," *IEEE Internet Computing*, vol. 5, pp. 59-68, 2001.
- [5] D. Higgins, C. Berkley, and M. B. Jones, "Managing heterogeneous ecological data using Morpho," in *14th International Conference on Scientific and Statistical Database Management*, 2002.
- [6] B. S. Bloom, J. T. Hastings, and G. F. Madaus, *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill Book Company, 1971.

AUTHOR INFORMATION

Mark Urban-Lurain, Director of Instructional Technology Research & Development, Division of Science and Mathematics Education, Michigan State University, urban@msu.edu

Diane Ebert-May, Professor, Department of Plant Biology, Michigan State University, ebertmay@msu.edu

Jennifer Momsen, Research Associate, Department of Plant Biology, Michigan State University, momsenj@msu.edu

Ryan McFall, Associate Professor, Computer Science Department, Hope College, mcfall@hope.edu

Matthew B. Jones, Director of Informatics Research and Development, National Center for Ecological Analysis and Synthesis, jones@nceas.ucsb.edu

Ben Leinfelder, Software Engineer, National Center for Ecological Analysis and Synthesis, leinfelder@nceas.ucsb.edu

Jon Sticklen, Director Applied Engineering Sciences, College of Engineering, Michigan State University, sticklen@msu.edu