JOURNAL OF CHEMICAL EDUCATION

Article

pubs.acs.org/jchemeduc

Blending Education and Polymer Science: Semiautomated Creation of a Thermodynamic Property Database

Roselyne B. Tchoua,[†] Jian Qin,[∇] Debra J. Audus,[§] Kyle Chard,[⊥] Ian T. Foster,[†,⊥,#] and Juan de Pablo*[,‡,⊥,¶]

[†] The Department of Computer Science, The University of Chicago, Chicago, Illinois 60637, United States

[∇] The Department of Chemical Engineering, Stanford University, Stanford, California 94305, United States

[§] Materials Science and Engineering Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States

[‡] The Institute for Molecular Engineering, The University of Chicago, Chicago, Illinois 60637, United States

[⊥] The Computation Institute, The University of Chicago, Chicago, Illinois 60637, United States

[#] Mathematics and Computer Science Division, Argonne National Laboratory, Lemont, Illinois 60439, United States

[¶] Materials Science Division, Argonne National Laboratory, Lemont, Illinois 60439, United States

S *Supporting Information*

ABSTRACT: Structured databases of chemical and physical properties play a central role in the everyday research activities of scientists and engineers. In materials science, researchers and engineers turn to these databases to quickly query, compare, and aggregate various properties, thereby allowing for the development or application of new materials. The vast majority of these databases have been generated manually, through decades of labor-intensive harvesting of information from the literature, yet while there are many examples of commonly used databases, a significant number of important properties remain locked within the tables, figures, and text of publications. The question addressed in our work is whether and to what extent the process of data collection can be automated. Students of the physical sciences and engineering are often confronted with the challenge of finding and applying property data from the literature, and a central aspect of their education is to develop the critical skills needed to identify such data and discern their meaning or validity. To address shortcomings associated with automated information extraction while simultaneously preparing the next generation of scientists for their future endeavors, we developed a novel course-based approach in which students develop skills in polymer chemistry and physics and apply their knowledge by assisting with the semiautomated creation of a thermodynamic property database.

KEYWORDS: *Polymer Chemistry, Physical Properties, Materials Science, Computer-Based Learning, Collaborative/Cooperative Learning, Curriculum, First-Year Undergraduate, General Public*

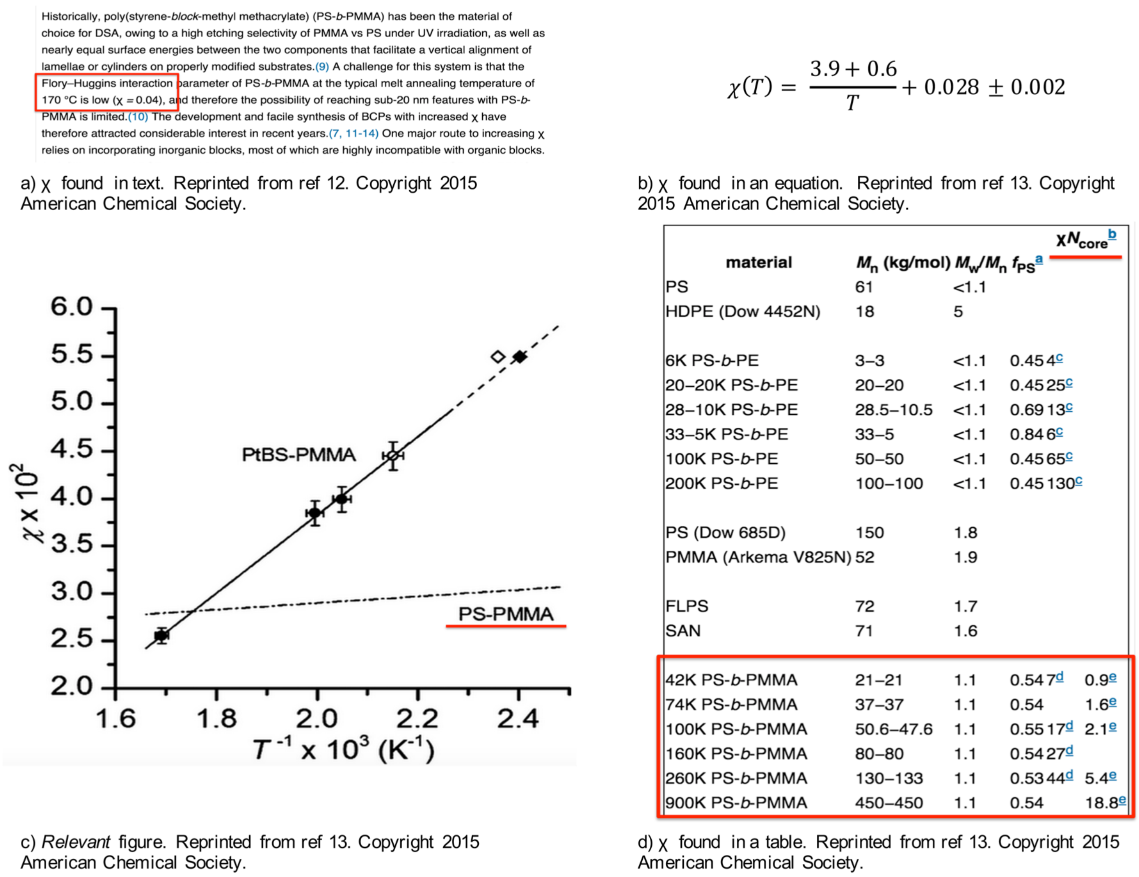
■ INTRODUCTION

The current explosion of digital materials information makes it ever more important to construct and maintain databases of physical properties, databases that will, ideally, be organized to permit efficient querying by both humans and machines.[1] The amount of scientific literature published every year is growing at an astounding rate. Some studies place the number of scientific journals at more than 28,000, and the number of articles published each year at 1.8 million.[2] The amount of information, including data, embedded within these articles is overwhelming, and reading and extracting pertinent information from full-text articles have become unmanageable tasks for scientists and engineers. Access to a structured, searchable database of all materials properties would facilitate the design and model validation of new substances, improving efficiency by enabling scientists and engineers to more quickly discover, query, and compare properties of existing compounds. However, without a concerted effort to generate such databases, this problem will only become larger with time, hindering not only today's but also tomorrow's engineers and scientists.

We address the challenge of creating such databases by engaging students via a specially designed course. The vision for this course is to expose them to the polymer science literature while solving the problem of missing online databases of polymer properties. Previous work suggests that finding and using information to understand a problem with precise instructions helps students develop information literacy skills.[3] In the course outlined here, we therefore sought to emphasize two distinct components. The first involved a formal classroom setting, where students were exposed to fundamental polymer science and context for that knowledge emphasizing real-world applications of both their efforts and related topics. The second, more hands-on, component allowed students to use their knowledge, along with software, to create an entirely new database of a polymer property named the Flory−Huggins (χ) parameter, thereby involving them directly in a project with importance in both academic and industrial realms.

Received: January 11, 2016
Revised: July 1, 2016

DOI: 10.1021/acs.jchemed.5b01032
J. Chem. Educ. XXXX, XXX, XXX−XXX



Historically, poly(styrene-*block*-methyl methacrylate) (PS-*b*-PMMA) has been the material of choice for DSA, owing to a high etching selectivity of PMMA vs PS under UV irradiation, as well as nearly equal surface energies between the two components that facilitate a vertical alignment of lamellae or cylinders on properly modified substrates.(9) A challenge for this system is that the Flory–Huggins interaction parameter of PS-*b*-PMMA at the typical melt annealing temperature of 170 °C is low (χ = 0.04), and therefore the possibility of reaching sub-20 nm features with PS-*b*-PMMA is limited.(10) The development and facile synthesis of BCPs with increased χ have therefore attracted considerable interest in recent years.(7, 11-14) One major route to increasing χ relies on incorporating inorganic blocks, most of which are highly incompatible with organic blocks.

a) χ found in text. Reprinted from ref 12. Copyright 2015 American Chemical Society.

$$\chi(T) = \frac{3.9 + 0.6}{T} + 0.028 \pm 0.002$$

b) χ found in an equation. Reprinted from ref 13. Copyright 2015 American Chemical Society.

c) *Relevant* figure. Reprinted from ref 13. Copyright 2015 American Chemical Society.

material	M_n (kg/mol)	M_w/M_n	f_{PS}[a]		χN_{core}[b]
PS	61	<1.1			
HDPE (Dow 4452N)	18	5			
6K PS-*b*-PE	3–3	<1.1	0.45	4[c]	
20–20K PS-*b*-PE	20–20	<1.1	0.45	25[c]	
28–10K PS-*b*-PE	28.5–10.5	<1.1	0.69	13[c]	
33–5K PS-*b*-PE	33–5	<1.1	0.84	6[c]	
100K PS-*b*-PE	50–50	<1.1	0.45	65[c]	
200K PS-*b*-PE	100–100	<1.1	0.45	130[c]	
PS (Dow 685D)	150	1.8			
PMMA (Arkema V825N)	52	1.9			
FLPS	72	1.7			
SAN	71	1.6			
42K PS-*b*-PMMA	21–21	1.1	0.54	7[d]	0.9[e]
74K PS-*b*-PMMA	37–37	1.1	0.54		1.6[e]
100K PS-*b*-PMMA	50.6–47.6	1.1	0.55	17[d]	2.1[e]
160K PS-*b*-PMMA	80–80	1.1	0.54	27[d]	
260K PS-*b*-PMMA	130–133	1.1	0.53	44[d]	5.4[e]
900K PS-*b*-PMMA	450–450	1.1	0.54		18.8[e]

d) χ found in a table. Reprinted from ref 13. Copyright 2015 American Chemical Society.

Figure 1. Four examples of how the Flory−Huggins χ parameter for the same pair of compounds may be found in the literature in various forms.

We suggest that this combination of theoretical and applied work could lead to better prepared future scientists and engineers. New generations of students rely extensively on the web as a source of information. In the case of polymer thermophysical data, that information is not always directly available. As they are confronted with the challenge of extracting data from the literature, they face the challenge of discerning which data are relevant, how they were measured and validated, and the manner in which they were analyzed, summarized, and ultimately published. Developing the skills to perform these tasks is an integral part of becoming a scientist.

The objective of this work is to present the outcomes and educational lessons learned in the development of a digital collection of χ values, with our long-term goal being to use student feedback to inform solutions for the automated collection and rationalization of polymer properties in databases assembled from publications available online. We motivate this work via the absence of a comprehensive database of polymer properties. We introduce the course structure and describe the software through which students interacted with the literature in the third section and present our findings and statistics accumulated from our course in the fourth section. We summarize and provide a few general observations in the final section.

■ MOTIVATION: CREATING A DATABASE OF POLYMER PROPERTIES

While there exist databases for hard[4] and metallic[5] materials, creating a database for polymers blends is challenging. Indeed, such a database needs to accommodate differences in polymer names due to the chemical structure and blend composition. For example, PolyInfo[6] is a database for polymeric materials extracted from publications satisfying the criteria that chemical structures of constitutional units are clearly determined by the authors. The extraction of these properties requires careful screening of publications by polymer specialists.[6] Our approach integrates the in-class education practice with the ultimate goal of automating the extraction of polymer properties from the literature using students' feedback.

We start with a particularly challenging property, not included in existing databases such as PolyInfo, the so-called Flory−Huggins (χ) parameter,[7] which characterizes the miscibility of polymer blends and polymer solutions. Since polymeric materials both are ubiquitous and typically consist of several polymeric components, which are generally incompatible, the χ parameter represents a key property for design of next-generation materials. Specifically, the χ parameter, which depends on the temperature and the types of polymer(s) or solvent(s) involved, is widely used to characterize the thermodynamic properties, including phase behavior, of polymer blends. Consequently, many experimental methods have been developed to extract χ and its temperature dependence, and representative values are often tabulated in standard polymer data handbooks.[8,9] However, such tables are rarely up to date with recent findings.

While there are thousands of published values of χ, there is little consensus regarding the validity or meaning of different numbers. Different measurement methods often yield different values, and different authors have at times reported different values. Part of this variability is due to inherent deficiencies

DOI: 10.1021/acs.jchemed.5b01032

within Flory−Huggins theory, which states that χ is inversely proportional to temperature. However, experimental evidence suggests a more complicated dependence, such that published χ values are often labeled as "effective" values in order to acknowledge these deficiencies. A database of χ values and associated measurement context will allow researchers to make informed judgments as to which χ values and thermodynamic analysis to use when predicting and understanding the phase behavior of multicomponent polymeric materials.

A possible solution to this problem, toward which we are moving in this work, would be to circumvent the need for manually curated paper copies of materials, which are compiled at considerable cost every several years, by extracting facts directly from scientific papers to be stored in efficient, human- and machine-readable databases. Ideally, since machines are capable of processing large volumes of text faster than humans, a fitting solution would involve computers "reading" thousands of papers and outputting structured content for human consumption. However, while computer-based solutions have improved significantly over the past few decades, extraction of structured data from unstructured documents remains a challenging task that continues to require human supervision. Computer scientists are investigating whether machine-learning techniques can "learn" from a minimal number of knowledgeable human curators and evolve to automatic extraction of "scientific facts" from publications, but a complete solution does not yet exist.[10,11] In building this database, we want to leverage a wealth of relevant information in published research articles. However, mining the literature for loosely structured scientific entities such as χ values, which are inevitably encoded in different forms in manuscripts of various formats (see Figure 1), is a challenging task.[15] A parameter such as χ is not typically captured as a common metadata element, as are, for example, title, authors, and publication date. Nor is it always found in a standard paper element, such as figure, table, or equation. Therefore, mining publications for χ requires extracting values from nonstandardized text, tables, equations, and figures—a challenging task involving encoding, formatting, and other processing activities.[16] Beyond the challenges of locating and extracting the χ parameter for a given paper element, we must consider that this parameter is often reported under various temperature-dependent forms. Moreover, identifying and storing the Flory−Huggins parameter only makes sense if the corresponding polymers, solvents, molecular masses, methods, errors, and other measurement information are also captured. For these reasons, we believe that the population of such a database currently requires hybrid human−computer methods.

More generally, our view is that if we succeed in creating semiautomated tools for database creation in the context of χ, which is arguably one of the most challenging properties to collect and categorize due to the variety of measurement methods, it will be easier to translate such tools to create databases for other important properties.

■ CLASS STRUCTURE

Given the ubiquity of polymers, it has been suggested that every chemist should have at least an introductory course on the subject.[17] Similarly, information literacy is recognized as a topic of the utmost importance in the formation of undergraduate chemistry students.[18,19] Engagement theory[20] and constructive theory[21] hold that students benefit from meaningful involvement in interactive and worthwhile tasks and that learning is most effective when students are active in knowledge formation.

Furthermore, previous work has consistently indicated that the use of supplemental computer-based instructional methods in chemistry has a positive influence upon student performance.[22] Finally, more recent studies suggest that collaborative learning positively changes students' attitude toward chemistry.[23,24] We aimed to incorporate and integrate these findings and teaching philosophies into the class structure.

Our topical course—not a required core course—entitled "Materials Database Creation for MacroMolecules" began with an introduction to both polymer science and databases. After the students learned the basics, which occurred after six 90 min classes, the lectures alternated between application of their knowledge through database development and the teaching of additional polymer physics. The class size was kept small in order to allow the instructors to interact closely with all students. Specifically, the class included a small group of undergraduate students (four freshmen, two juniors, and three seniors) with diverse backgrounds including chemistry, physics, and biology.

Course Content

The first portion of the course opened with a broad overview that described the context for the students' role. This overview included an introduction to commercial and industrial products that rely on polymeric materials. The importance of their involvement was emphasized by outlining the scale of the polymer industry, for example, the fact that the global production of plastics amounts to more than one hundred million metric tons per year.[25] Students were asked to find the chemical structures for common polymers and were taught some key single-polymer characteristics, including molecular mass and configurational statistics. Working on chemical structures prepared them to understand some of the polymers that they would later encounter in the literature. A discussion of polymer characteristics created the groundwork for understanding Flory−Huggins theory, from which the χ parameter is derived.

This general introduction was followed by a general derivation of the Flory−Huggins theory, starting from elementary principles, so that students could develop an appreciation for the meaning of the χ parameter. The discussion of the Flory−Huggins theory was accompanied by examples outlining its application to the calculation of phase diagrams as a function of χ. Such examples illustrated the importance of developing a χ parameter database, and also equipped students with an understanding of the various methods that are used experimentally to determine the χ parameter. The presentation of Flory−Huggins theory was followed by discussion of the osmotic pressure of polymer solutions and its applications. This particular topic was chosen so that students could analyze tractable experimental data and extract basic information about the magnitude of the χ parameter that is typically encountered in such systems.

The next phase of the course introduced students to modern databases. These lectures provided an overview and history of databases, their purpose and current applications, and basic technical details. For example, approaches for data modeling and concepts such as primary keys and joins were presented. The database lectures provided students with the skills to browse and understand the database they would populate, thus presenting a behind-the-scenes look into the software and the

DOI: 10.1021/acs.jchemed.5b01032

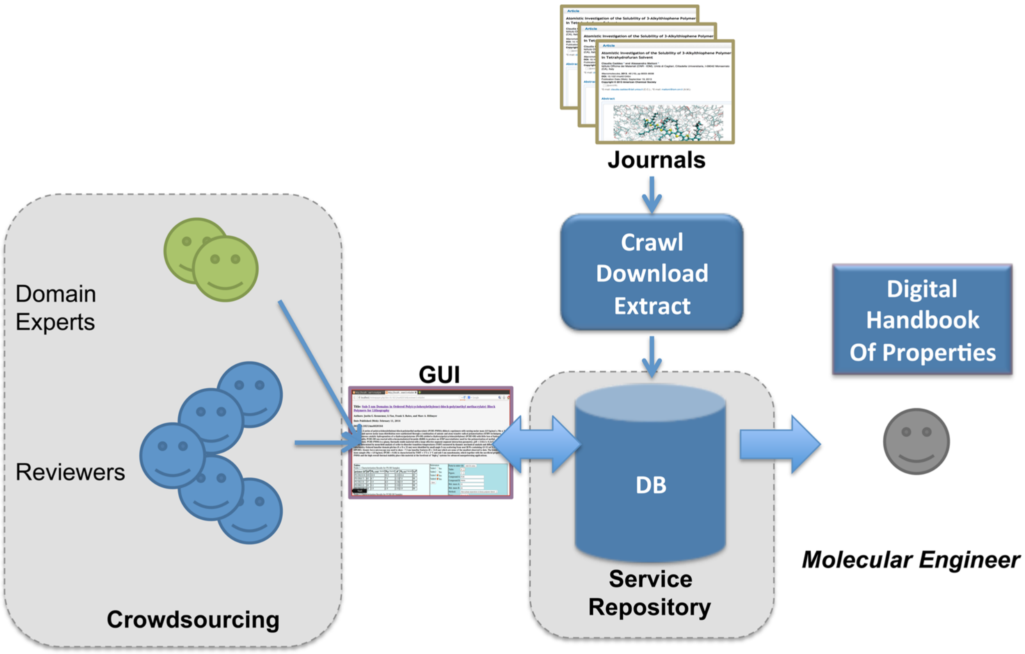
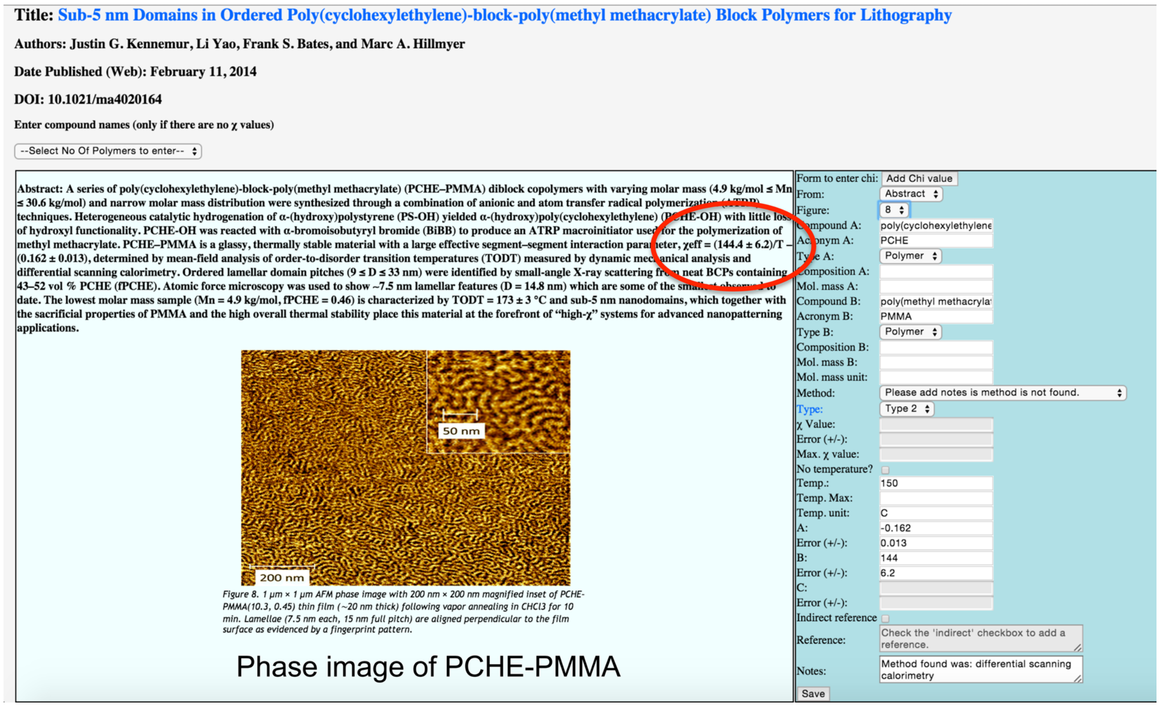


Figure 2. χDB architecture.



Title: Sub-5 nm Domains in Ordered Poly(cyclohexylethylene)-block-poly(methyl methacrylate) Block Polymers for Lithography

Authors: Justin G. Kennemur, Li Yao, Frank S. Bates, and Marc A. Hillmyer

Date Published (Web): February 11, 2014

DOI: 10.1021/ma4020164

Enter compound names (only if there are no χ values)

--Select No Of Polymers to enter--

Abstract: A series of poly(cyclohexylethylene)-block-poly(methyl methacrylate) (PCHE–PMMA) diblock copolymers with varying molar mass (4.9 kg/mol ≤ Mn ≤ 30.6 kg/mol) and narrow molar mass distribution were synthesized through a combination of anionic and atom transfer radical polymerization (ATRP) techniques. Heterogeneous catalytic hydrogenation of α-(hydroxy)polystyrene (PS-OH) yielded α-(hydroxy)poly(cyclohexylethylene) (PCHE-OH) with little loss of hydroxyl functionality. PCHE-OH was reacted with α-bromoisobutyryl bromide (BiBB) to produce an ATRP macroinitiator used for the polymerization of methyl methacrylate. PCHE–PMMA is a glassy, thermally stable material with a large effective segment–segment interaction parameter, χeff = (144.4 ± 6.2)/T – (0.162 ± 0.013), determined by mean-field analysis of order-to-disorder transition temperatures (TODT) measured by dynamic mechanical analysis and differential scanning calorimetry. Ordered lamellar domain pitches (9 ≤ D ≤ 33 nm) were identified by small-angle X-ray scattering from neat BCPs containing 43–52 vol % PCHE (fPCHE). Atomic force microscopy was used to show ~7.5 nm lamellar features (D = 14.8 nm) which are some of the smallest observed to date. The lowest molar mass sample (Mn = 4.9 kg/mol, fPCHE = 0.46) is characterized by TODT = 173 ± 3 °C and sub-5 nm nanodomains, which together with the sacrificial properties of PMMA and the high overall thermal stability place this material at the forefront of “high-χ” systems for advanced nanopatterning applications.

Figure 8. 1 μm × 1 μm AFM phase image with 200 nm × 200 nm magnified inset of PCHE-PMMA(10.3, 0.45) thin film (~20 nm thick) following vapor annealing in CHCl3 for 10 min. Lamellae (7.5 nm each, 15 nm full pitch) are aligned perpendicular to the film surface as evidenced by a fingerprint pattern.

Phase image of PCHE-PMMA

Form to enter chi:	Add Chi value
From:	Abstract
Figure:	8
Compound A:	poly(cyclohexylethylene
Acronym A:	PCHE
Type A:	Polymer
Composition A:	
Mol. mass A:	
Compound B:	poly(methyl methacryla
Acronym B:	PMMA
Type B:	Polymer
Composition B:	
Mol. mass B:	
Mol. mass unit:	
Method:	Please add notes is method is not found.
Type:	Type 2
χ Value:	
Error (+/-):	
Max. χ value:	
No temperature?	□
Temp.:	150
Temp. Max:	
Temp. unit:	C
A:	-0.162
Error (+/-):	0.013
B:	144
Error (+/-):	6.2
C:	
Error (+/-):	
Indirect reference	□
Reference:	Check the 'indirect' checkbox to add a reference.
Notes:	Method found was: differential scanning calorimetry

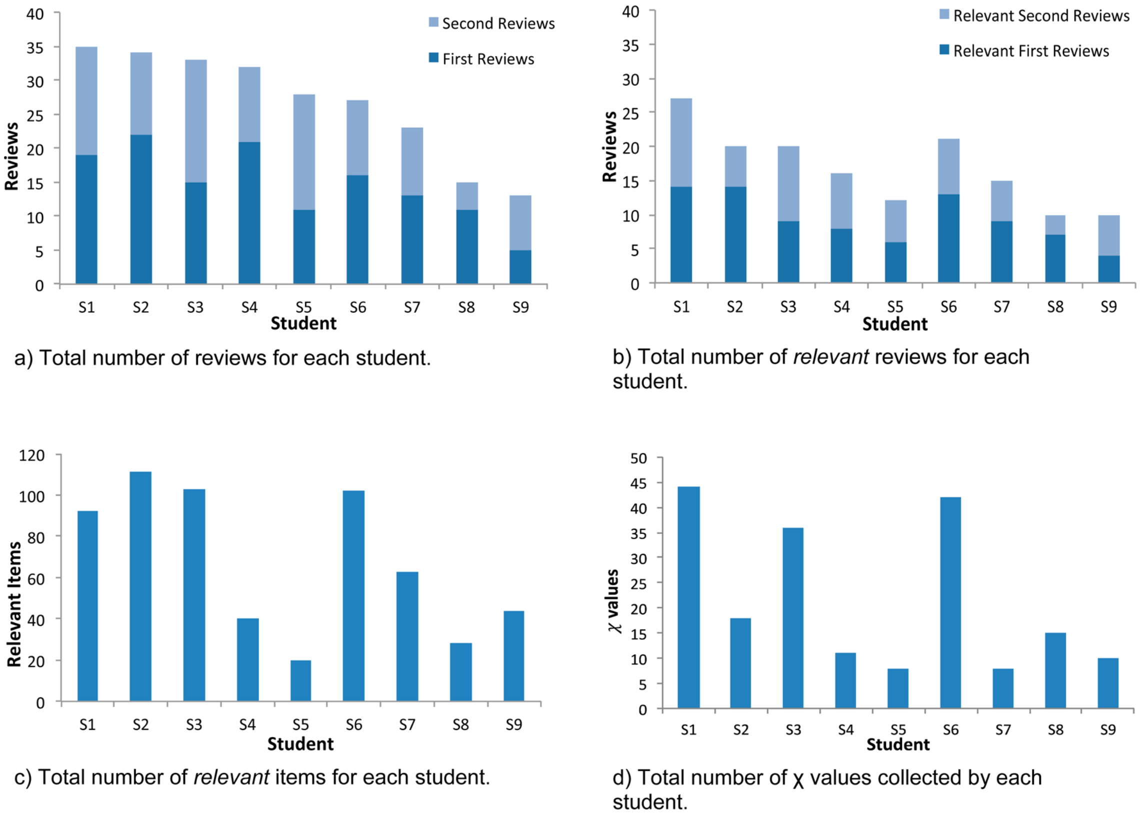
Figure 3. Screenshot of χDB graphical user interface with the χ entry form enabled. Data and figure reprinted from ref 28. Copyright 2014 American Chemical Society.

data organization. Such skills will likely be important in their future as big data continues to grow and databases become ubiquitous across fields.

At this point, it was assumed that students were ready to participate in the creation of the database. For the remainder of the course, lectures alternated between classes on fundamentals and classes on literature review using specialized software, which are discussed in more detail in the next section. We encouraged questions to instructors and to other students during in-class publication reviewing sessions.

Software Assisting Database Creation: χDB

The literature review component of the course leveraged custom software developed by the authors. This software, named χDB,[26] comprises two components that allow students to focus on the science. The first component extracts relevant information from the published literature. The second component presents the previously mined information to students via a Web Interface for review and extraction. Figure 2 shows the χDB software architecture.



a) Total number of reviews for each student.

b) Total number of *relevant* reviews for each student.

c) Total number of *relevant* items for each student.

d) Total number of χ values collected by each student.

Figure 4. Student contributions to the database.

The first component is a Web crawler, coded in the Python programming language,[27] which crawls *Macromolecules* publications and downloads articles that include "Flory−Huggins" as a keyword. We configured the crawler to download all matching articles since January 2010. It then extracts structured metadata such as the title, authors, abstract, and digital object identifier (DOI) by detecting specialized HTML (hypertext markup language) tags. Finally, the crawler downloads the full-text HTML version of the paper, extracts all paper elements that may contain parameter values (in particular, equations, tables, and figures), and registers these data into a MySQL database. While this crawler is currently specific to *Macromolecules*, it can be modularized to support other journals that publish their articles online in HTML.

The χDB literature review graphical user interface (GUI) is a Web site accessible only on the University of Chicago network to ensure that only those with access to the journal are able to perform reviews. The home page follows a three-column layout pattern: the first column lists all papers in the database, the second lists the papers that have been reviewed once, and the third lists the papers that have been fully (twice) reviewed. Clicking on an individual paper leads to the review page.

The review page presents students with the articles (in extracted form) and the associated metadata. It also allows students to mark items (figures, equations, or tables) as *relevant*, meaning that they relate to the Flory−Huggins parameter, for example, because they contain a χ value. Note that not every item identified as *relevant* actually contained a χ value. For example, a figure may be a phase diagram or contain the structure of polymers for which the χ parameter is determined.

When students find a χ value, they click on the "Add Chi Value" button next to the element they are reviewing (e.g., abstract, table, or equation). A form is then generated to enable entry of the name of the polymer(s), the molecular weight(s) of the polymer(s), the method used to measure the χ value, and other information pertinent to the experimentally determined χ value found in the publication. For example, while χ is generally assumed to be independent of the polymer blend composition (φ), this is not always the case, and in practice some authors provide the concentrations of the individual components of the system. Figure 3 shows an example of this form.

To minimize ambiguity in the database, we define a set of minimum required information for successful submission of a χ value. In defining a data model for χ, we rely on the three principal theoretical temperature-dependent representations of χ found in the *Physical Properties of Polymers Handbook*,[8] plus a fourth representation of "Other" for flexibility: type 1, a value reported at a specific temperature; type 2, a linear equation in terms of temperature, $\chi = A + \frac{B}{T}$; type 3, a quadratic equation in terms of temperature: $\chi = A + \frac{B}{T} + \frac{C}{T^2}$; and type 4, other.

We anticipated that some χ values would be embedded directly in the text, rather than only the figures and tables automatically extracted by the χDB Web-crawler. For that reason, the title of each article on the review page was linked to the original (full text) publication, so that students could scan each paper for other χ values. In this scenario, students click on the "Add Chi" button (next to the abstract), and then complete the form to indicate that the value was extracted from the text. Students can also specify whether a reported χ value is obtained from another publication, as opposed to being determined in the work being reviewed. Finally, students can add notes to further describe and support their entries.

As previously mentioned, each publication is reviewed by two students to reduce errors. If the two reviews produce conflicts,

Table 1. Comparison of Survey Results on Closed-Ended Questions by Student Category

	Average Scores or Times and SD Values for Student Responses			
	All Students, $N = 9$		First-year Students, $N = 4$	
Question for Response with Scale or Parameter Indicated	Average	Standard Deviation	Average	Standard Deviation
How many in-class sessions were necessary for you to start reviewing papers on your own?	2.5	1.3	3.0	1.4
How long did it take you to review a relevant paper? (time in minutes)	14.3	5.6	17.0	5.4
Do you think being a college student in the physical sciences/engineering should be a requirement for being a reviewer? (1 = yes; 0 = no)	0.88	0.35	0.80	0.44
Would you be interested in reviewing papers in the future? (1 = yes; 0 = no)	0.62	0.53	1.00	0
On a scale of 1 to 5, how much background is required for the literature review?	3.12	0.99	3.60	0.89
On a scale of 1 to 5, how important was the motivation for the database in motivating you as a reviewer?	3.12	0.64	3.20	0.84
On a scale of 1 to 5, beyond the in-class sessions, how important is the access to an expert molecular engineer to answer specific questions about the publications?	3.50	0.93	4.00	0.71
On a scale of 1 to 5, how comfortable are you with the web interface?	4.00	0.76	4.00	0.71

the publication is flagged for review by an expert. Initially, publications were only reviewed in class so that students could ask questions and provide suggestions. As the course continued, students were asked to make suggestions for improvement, which were incorporated into the software throughout the course. This approach allowed the students to have a meaningful impact on the final database. For example, some χ values were reported in publications as valid for a range of temperatures rather than a single point at a specific temperature. We modified the form to include minimum and maximum temperatures to accommodate such values. The instructors and students shared an online document to report and address problems; this form was especially active early in the class. After initial concerns were addressed, students also reviewed publications outside of the class.

Student Assessment

In addition to the lectures, six homework exercises were assigned to emphasize various aspects of the course material. One such exercise asked students to extract a χ parameter value from small angle neutron scattering data. The average grade in this exercise was 80%. In another exercise, students correctly computed the limits of stability for phase separation using Flory−Huggins theory with an average grade of 96%. Student progress was evaluated through an in-class midterm exam and a take-home final exam. As part of the final exam, students were asked to use Flory−Huggins theory to generate a phase diagram as a function of χ. They then combined their phase diagram with results from a homework exercise, thus arriving at a product that scientists could use to design polymeric materials. This exercise aimed to provide direct experience on how the χ parameters within the database were created and how researchers might use them.

The average grade for the midterm exam was 89%. The average grade for the final exam was 99.4%, and all students were able to derive the phase diagram. Instructors also assigned a final number of publication (first and second) reviews on the χDB Web site, which all students completed. We give an overview of the data collected and present the students' feedback in the next section.

■ RESULTS AND DISCUSSION

Our nine students reviewed 133 papers over an eight-week period, of which 84 were found to be *relevant* as defined earlier. Students marked a total of 360 items as *relevant* and identified 138 χ values including 88 χ values for polymer blends containing 60 unique polymers, as well as 50 χ values for polymer solutions. For comparison, the *Physical Properties of Polymers Handbook*[8] contains χ values for polymer blends containing only 41 unique polymers, less than the students' value of 60.

We also investigated student contributions to the database by considering the number of papers for which they performed first and second reviews, and the number of those papers that contained *relevant* items; these results can be seen in Figure 4a,b, respectively. In order to preserve student anonymity, the students were each assigned a unique identifying number (ordered by the total number of papers reviewed). It is likely that the greater variation across students for *relevant* reviews is a consequence of the significant number of papers that contained no *relevant* items and the random assignment of papers to students, rather than student performance. *Relevant* papers can then be broken down into relevant items, defined as relating to but not necessarily containing χ, and the number of χ values collected as can be seen in Figure 4c,d, respectively. There is no direct correlation between the number of *relevant* items identified in a paper and the number of χ values eventually extracted from that paper. Thus, in the future, we may want to refine the concept of relevancy of items. For example, figures may be *relevant* because they are illustrations of materials or because they are phase diagrams, the latter are more directly related to χ and may be more correlated with the number of χ values extracted. While instructors emphasized phase diagrams, students were also successful in identifying figures that related to the material or the method. This initiative showed a clear understanding of the motivation for their work and the impact of their input on future uses of the database.

We were also able to characterize important aspects of the paper review process by soliciting student feedback through a survey. This survey involved both closed- and open-ended questions. The closed-ended questions are summarized in Table 1. These results, combined with comments during literature reviews in class, represent critical information for determining future modifications of our methods. They can also be used as guidelines for designing similar courses.

We also probed the students' backgrounds and their views on the information required for effective reviewing. One student was auditing and hence not present at the final exam. We found that most students were comfortable reviewing publications independently after only two in-class sessions (see Q1, Table

1), with each publication consuming a moderate time of roughly 15 min (see Q2, Table 1). This experience suggested that their backgrounds in conjunction with course material were the key information required to review publications. Given that only half of the students had any specialized background in materials or polymers as determined from an open-ended question, the content of the course appeared to be successful in preparing the students for their involvement in database creation. The students reported that scientific knowledge was required for reviewing, which seems aligned with their experience in the course. Specifically, seven out of the eight registered students reported that future reviewers should be a college student in the physical sciences or engineering (see Q3, Table 1).

In addition to determining metrics for success, we investigated the students' motivation and enjoyment of their role as reviewers. We found that the freshmen responded more favorably to an open-ended question on their general thoughts regarding the course. The freshmen also all responded that they would be interested in reviewing papers in the future, while none of the upperclassmen were interested (Q4, Table 1). Note that reviewing involves using the software for data entry while reading publications, which more experienced students may have judged to be not sufficiently challenging. Freshmen may also have shown more interest because reviewing literature was likely to be a new experience for them. In general, freshmen also reported a higher desired level of background knowledge prior to reviewing papers (Q5, Table 1), more in-class sessions, and more time to review a paper. While the differences are small, they may indicate that freshmen also found the task more challenging and hence more interesting than did their more senior classmates. Students reported that learning the motivation for the course, while important, was not critical (Q6, Table 1). Students may have simply enjoyed the exposure to the literature or the nontraditional nature of course.

Regarding the software and its usability, the students found that having experts on hand during the in-class sessions was helpful (3.5/5), but not vital (see Q7, Table 1). The expert-flagging feature was added in anticipation of conflicts between reviewers, but in practice the collaborative nature of the in-class sessions, which encouraged students to ask instructors and each other when they were confused, resulted in only one paper being flagged. This factor also might explain why they may not have considered access to experts essential.

Students also reported that they were comfortable using the web interface (see Q8, Table 1). However, there were, and continue to be, improvements that can be made in the software to improve future reviewers' experiences. For example, students reported early in the review process that some χ values were given as a range and that some papers contained *relevant* figures without specific χ values, two factors that had not been considered in the initial design. We thus added the ability to enter a χ value as a range as well as the ability to enter which polymers were mentioned in the paper even when no specific χ values were present. Students had additional suggestions for improvements in the final survey, such as entering several similar χ values as a set instead of entering them one at a time and displaying each user's history and statistics. As coordinators of the χ database, we considered the feedback of the students invaluable to the success of the project as a whole.

■ CONCLUSION

We have reported on a preliminary and ambitious attempt to educate students in polymer science and engineering via engaging them in the population of a database of polymer properties by a human−machine interactive approach. The project involved synergistic efforts from experts in such diverse areas as computational science and polymer physics, along with significant contributions from undergraduates in various scientific disciplines.

Final grades indicate that students learned the material presented during the class. Their ability to identify relevant information in publications beyond what was generally covered by instructors during the classes implies that they benefited from the exposure to the literature and experience extracting properties from a variety of sources.

In general, students reported enjoying their experience, including their interaction with scientific literature, which was an intended goal of the course. Freshmen were particularly enthusiastic, we believe as a result of increased engagement due to the more appropriate level of difficulty of the class. We also asked students if they would like to continue reviewing papers; nearly half of the class has continued to review papers over the summer suggesting that many of the students enjoyed contributing to the solution of a problem facing the scientific community. The students also provided vital insight into the ways in which the Flory−Huggins χ parameter is represented and used in the literature and learned both fundamental polymer physics and how databases are designed, populated, and used.

Our experience suggests that with adequate training and with a sufficiently friendly user interface, undergraduate students can play an important role in tackling the problem of creating scientific databases for both the academic and industrial sectors. In just one academic quarter, nine students were able to identify χ values for 60 polymer blends, including not only 13 (31.7%) of the 41 polymers values found in the *Physical Properties of Polymers Handbook*,[8] but also χ values for an additional 47 polymers not found in the handbook. The large number of new polymers found is not surprising, given that the handbook was published in 2007 and that we analyzed only papers published in *Macromolecules* between 2010 and 2015. Nonetheless, it emphasizes the potential for using our approach to create and maintain a digital database of Flory−Huggins χ parameters that is more up to date than any survey publication. Importantly, the course also trains and involves future scientists in a vital task with minimal financial expense.

We plan to make our software available so that other universities and institutions can offer the same course model. Ultimately, however, our goal is to learn from the students' input and reduce the human component of the system to perhaps a handful of experts. These experts would conduct a number of reviews, which a machine-learning algorithm could then use as ground truth in order to achieve varied level of automatic classification and extraction of polymer properties from publications. Preliminary results in this direction are encouraging though beyond the scope of this paper. We are also in discussions with other journals with a view to expanding the initial data set. The data collected is publically available on the Material Genome Polymer Property and Predictor Project Web site.[29] We are currently evaluating the quality of the collected data and conducting a more detailed and qualitative comparison with the *Properties of Polymers Handbook*.[8]

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available on the ACS Publications website at DOI: 10.1021/acs.jchemed.5b01032.

Raw data for survey data found in Table 1 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: depablo@uchicago.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Jack F. Douglas of the National Institute of Standards and Technology, as well as Karl F. Freed and Jacek Dudowicz of the University of Chicago, for their comments. This work was supported by the NIST through the Center for Hierarchical Materials Design (CHiMaD).

■ REFERENCES

(1) Adams, N.; Schubert, U. S. From data to knowledge: chemical data management, data mining, and modeling in polymer science. *J. Comb. Chem.* **2004**, *6* (1), 12−23.

(2) Mabe, M.; Ware, M. *The STM report: An overview of scientific and scholarly journals publishing*; International Association of Scientific, Technical and Medical Publishers (STM), Prama House: Oxford, United Kingdom, 2009.

(3) Shultz, G. V.; Li, Y. Student Development of Information Literacy Skills during Problem-Based Organic Chemistry Laboratory Experiments. *J. Chem. Educ.* **2016**, *93* (3), 413−422.

(4) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **2013**, *1* (1), 011002.

(5) Spencer, P. J. A brief history of CALPHAD. *CALPHAD: Comput. Coupling Phase Diagrams Thermochem.* **2008**, *32* (1), 1−8.

(6) Otsuka, S.; Kuwajima, I.; Hosoya, J.; Xu, Y.; Yamazaki, M. PoLyInfo: Polymer Database for polymeric materials design. *International Conference on Emerging Intelligent Data and Web Technologies (EIDWT)*; IEEE: Piscataway, New Jersey, United States, 2011; pp 22−29.

(7) Flory, P. J. Thermodynamics of high polymer solutions. *J. Chem. Phys.* **1942**, *10*, 51−61.

(8) Eitouni, H. B.; Balsara, N. P. Thermodynamics of polymer blends. *Physical Properties of Polymers Handbook*; Springer: New York, 2007; pp 339−356.

(9) Brandrup, J.; Immergut, E. H. *Polymer Handbook*, 2nd ed.; John Wiley and Sons: New York, New York, United States, 2003.

(10) Dong, X. L.; Gabrilovich, E.; Heitz, G.; Horn, W.; Murphy, K.; Sun, S.; Zhang, W. From data fusion to knowledge fusion. *Proceedings of the VLDB Endowment* **2014**, *7* (10), 881−892.

(11) Shin, J.; Wu, S.; Wang, F.; De Sa, C.; Zhang, C.; Ré, C. Incremental knowledge base construction using deepdive. *Proceedings of the VLDB Endowment* **2015**, *8* (11), 1310−1321.

(12) Luo, Y.; Montarnal, D.; Kim, S.; Shi, W.; Barteau, K. P.; Pester, C. W.; Hustad, P. D.; Christianson, M. D.; Fredrickson, G. H.; Kramer, E. J.; Hawker, C. J. Poly (dimethylsiloxane-b-methyl methacrylate): A Promising Candidate for Sub-10 nm Patterning. *Macromolecules* **2015**, *48* (11), 3422−3430.

(13) Kennemur, J. G.; Hillmyer, M. A.; Bates, F. S. Synthesis, Thermodynamics, and Dynamics of Poly (4-tert-butylstyrene-b-methyl methacrylate). *Macromolecules* **2012**, *45* (17), 7228−7236.

(14) Bell, J. R.; Chang, K.; López-Barrón, C. R.; Macosko, C. W.; Morse, D. C. Annealing of cocontinuous polymer blends: effect of block copolymer molecular weight and architecture. *Macromolecules* **2010**, *43* (11), 5024−5032.

(15) Hawizy, L.; Jessop, D. M.; Adams, N.; Murray-Rust, P. ChemicalTagger: A tool for semantic text-mining in chemistry. *J. Cheminf.* **2011**, *3* (1), 17.

(16) Kandel, S.; Heer, J.; Plaisant, C.; Kennedy, J.; van Ham, F.; Riche, N. H.; Weaver, C.; Lee, B.; Brodbeck, D.; Buono, P. Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Information Visualization* **2011**, *10* (4), 271−288.

(17) Harris, F. W. Introduction to polymer chemistry. *J. Chem. Educ.* **1981**, *58* (11), 837.

(18) Matthews, F. J. Chemical literature: a course composed of traditional and online searching techniques. *J. Chem. Educ.* **1997**, *74* (8), 1011.

(19) Greco, G. E. Chemical Information Literacy at a Liberal Arts College. *J. Chem. Educ.* **2016**, *93* (3), 429−433.

(20) Kearsley, G.; Shneiderman, B. Engagement Theory: A framework for technology-based teaching and learning. *Educ. Technol.* **1998**, 20−37.

(21) Duffy, T. M., Jonassen, D. H., Eds. *Constructivism and the Technology of Instruction: A Conversation*. Routledge: New York, New York, United States, 2013.

(22) Castleberry, J.; Culp, G. H.; Lagowski, J. J. The impact of computer-based instructional methods in general chemistry. *J. Chem. Educ.* **1973**, *50* (7), 469.

(23) Browne, L. M.; Blackburn, E. V. Teaching introductory organic chemistry: a problem-solving and collaborative-learning approach. *J. Chem. Educ.* **1999**, *76* (8), 1104.

(24) Shibley, I. A., Jr.; Zimmaro, D. M. The influence of collaborative learning on student attitudes and performance in an introductory chemistry laboratory. *J. Chem. Educ.* **2002**, *79* (6), 745.

(25) Plastics: the facts 2014/2015: An analysis of European plastics production, demand and waste data. PlasticsEurope, http://www.plasticseurope.org/, accessed Jun 2016.

(26) Tchoua, R. B.; Chard, K.; Audus, D. J.; Qin, J.; de Pablo, J.; Foster, I. T. A hybrid human-computer approach to the extraction of scientific facts from the literature. *Procedia Computer Science* **2016**, *80*, 386−397.

(27) Sanner, M. F. Python: A programming language for software integration and development. *J. Mol. Graph. Model.* **1999**, *17* (1), 57−61.

(28) Kennemur, J. G.; Yao, L.; Bates, F. S.; Hillmyer, M. A. Sub-5 nm Domains in Ordered Poly (cyclohexylethylene)-block-poly (methyl methacrylate) Block Polymers for Lithography. *Macromolecules* **2014**, *47* (4), 1411−1418.

(29) Material Genome Polymer Property and Predictor Project. http://pppdb.uchicago.edu/ (accessed Jun 2016).