Science
Education

RESEARCH ARTICLES

Answering the Questions of Whether and When Learning Occurs: Using Discrete-Time Survival Analysis to Investigate the Ways in Which College Chemistry Students' Ideas About Structure–Property Relationships Evolve

SONIA M. UNDERWOOD, DAVID REYES-GASTELUM, MELANIE M. COOPER
Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States

Received 29 July 2014; revised 21 April 2015; accepted 12 May 2015
DOI 10.1002/sce.21183
Published online 14 July 2015 in Wiley Online Library (wileyonlinelibrary.com).

ABSTRACT: Longitudinal studies can provide significant insights into how students develop competence in a topic or subject area over time. However, there are many barriers, such as retention of students in the study and the complexity of data analysis, that make these studies rare. Here, we present how a statistical framework, discrete-time survival analysis, can help overcome these barriers to longitudinal assessment studies using data from our research on students' understanding of structure–property relationships in chemistry. In the study presented, we administered the *Implicit Information from Lewis Structures Instrument* (IILSI)—an instrument designed to elicit from students what information can be predicted using a Lewis structure—to three cohorts of students at five time points over a two-year period, throughout their general chemistry and organic chemistry courses, to determine the ways in which student ideas about structure–property relationships evolved.

Correspondence to: Sonia M. Underwood; e-mail: under108@msu.edu
Contract grant sponsor: NSF.
Contract grant numbers: 1341987 and 1420005.
Supporting Information is available in the online issue at wileyonlinelibrary.com.

Using survival analysis, we were able to identify both *whether* and *when* learning occurred over the two-year period. The single model also allowed us to construct trajectories to determine the ways in which students developed the ideas that underlie structure–property relationships in chemistry. *Sci Ed* **99:**1055–1072, 2015

INTRODUCTION AND MODEL

The National Research Council (NRC) report on Discipline-Based Education Research (DBER) states that longitudinal studies on teaching and learning are necessary to help achieve long-term goals such as understanding "how people learn the concepts, practices, and ways of thinking of science and engineering" and "the nature and development of expertise in a discipline" (2012). Unfortunately, many contributing factors, including availability of resources, attrition of students in the study, and the difficulties associated with using appropriate analysis measures (Bauer, 2004; White & Arzi, 2005), have led to a dearth of longitudinal studies of student learning in science education (Tytler, 2009; White & Arzi, 2005) while even fewer studies have been replicated (White & Arzi, 2005). In a similar vein, the DBER report notes that resources such as "time and money are required to develop and refine measurement instruments and to conduct longitudinal studies" (NRC, 2012). Tracking students across courses or disciplines can also become problematic since students do not tend to move as a uniform cohort. This means that the number of students in the study tends to decrease over time, making it increasingly difficult to use traditional statistical measures.

The logistical problems inherent in longitudinal studies can push researchers to instead conduct cross-sectional studies with different groups of students (e.g., general chemistry and organic chemistry) to develop an understanding of student ideas over shorter periods of time. Current approaches to analyzing cross-sectional data include examining differences between two or more groups (e.g., independent *t*-test, analysis of variance, or chi squared analyses) or investigating change within a group of students (e.g., paired *t*-test, McNemar, or Wilcoxon signed rank). While these traditional statistical approaches help generate "snapshots" of student knowledge at specific points in time, these measures typically fail to provide information about the ways in which student ideas develop over periods of time greater than one semester.

Change-over-time measurements typically generate multifaceted data sets that require more complex analysis methods. The choice of appropriate analysis techniques must emerge from the research questions; for example, measurements of how an outcome changes over time and what factors contribute to differences between groups might be analyzed using multilevel models such as hierarchical linear models (Porter & Umbach, 2001) or growth models (Singer & Willett, 2003). However, the questions of identifying *whether* and *when* an outcome occurs require survival analysis (Singer & Willett, 2003) since these models were specifically developed to analyze data for which there is an interest in studying the length of time needed for an event to occur.

The question of *when* learning occurs we believe is as important as identifying *whether* learning has occurred. Identifying *when* becomes essential if foundational knowledge is required to develop a robust understanding of future material. There is a great deal of evidence to support the idea that for many learning goals, the sequence of concepts and skills must be well scaffolded in a learning progression (Corcoran, Mosher, & Rogat, 2009; Jin, Zhan, & Anderson, 2013; Stevens, Delgado, & Krajcik, 2010). If students do not have necessary prior knowledge, they may be forced to resort to memorization and surface level reasoning rather than meaningful learning (Bretz, 2001; Novak, 1998). Therefore, in this paper we present how a statistical analysis framework called survival analysis can be used to analyze longitudinal assessment data for identifying both *whether* and *when* student learning occurs.

Survival analysis is a type of regression analysis that is specifically designed for longitudinal studies where the length of time for an event to occur is identified (Willett & Singer, 2012). As with most analysis frameworks, there are various types of survival analysis models available ranging from parametric to nonparametric (e.g., Weibull, Cox proportional hazards, and Kaplan–Meier [Liu, 2012]). To determine which survival analysis model is appropriate, the following aspects should be taken into consideration: (i) Is the event of interest repeatable or nonrepeatable? (ii) Are there multiple aspects of the event that are important? (e.g., if marriage is the event of interest, is it enough to know when the participant marries or is it important to distinguish a first marriage from a second marriage), and (iii) Are the data parametric or nonparametric? (Allison, 2014).

Originating from mortality tables or life tables to determine life expectancies (Miller, 1981), the use of survival analysis is predominantly found in the medical research literature where the event of interest might be drug relapse or death (e.g., Grilli, Oxman, & Julian, 1993; Kimber et al., 2010). This type of analysis has expanded into other fields such as engineering, sociology, and economics (Miller, 1981) under various names, for example, event history analysis in sociology, reliability analysis or failure time in engineering, and duration analysis or transition analysis in economics (Allison, 2010). Examples outside of the medical field include investigating the effectiveness of government programs on long-term unemployment (Firth, Payne, & Payne, 1999) and analyzing warranty information to determine vehicle reliability issues (Zhou, Chinnam, & Korostelev, 2012).

Survival analysis has made an appearance in educational research to examine retention rates for both students and instructors. Singer and Willett used this analysis framework to examine the average length of a special education teaching career (1993), while other researchers used it to understand factors influencing student retention and degree completion rates (Bowers, 2010; Ishitani, 2006; Zwick & Sklar, 2005). Without this approach, these researchers would not have been able to answer the questions of *whether* and *when* an event will occur, which Singer and Willett refer to as the "whether and when test" for using survival analysis (Singer & Willett, 2003). That is, they suggest that survival analysis should be used if a research question contains both of these intentions.

Survival analysis is unique because of its ability to handle censored data (Hosmer, Lemeshow, & May, 2011). A participant's response is considered censored (more specifically right-censored) when that response is incomplete or missing, meaning that it is not known when the event of interest will occur for the participant. A response is incomplete when the event of interest has not occurred either (i) before the participant leaves the study or (ii) when the study was completed. Incomplete responses are particularly problematic in longitudinal studies where it is usually impossible to retain all participants for the length of the study. For example, consider a five-year study on whether and when a person marries. If a person leaves the study after three years and has not married before that time point, the researcher does not know whether the participant will marry or not within the last two years of the study. Therefore, this participant's missing responses for the time points after leaving the study would be considered censored.

Traditional statistical analysis methods are not appropriate or satisfactory for this analysis because of the way they handle censored data (Guo, 2010; Singer & Willett, 1991). That is, some traditional methods treat censored data as missing, meaning that most methods do not use any information that is actually known about the participant and the participant is omitted from the study. While there are methods to impute missing data (e.g., Hsu, Taylor, Murray, & Commenges, 2007), they still require subsequent use of an analysis framework that can be used to analyze the research question of *whether* and *when* learning (i.e., time to event) occurs. Survival analysis simplifies this process since it makes use of the censored data and analysis of the research questions all in one step.

Although survival analysis has been used to analyze some types of educational data, to the best of our knowledge, this type of analysis has not been used to analyze longitudinal assessment of learning. We propose that survival analysis can provide some insight into *the ways in which* learning develops over time, by allowing us to determine *when* students first develop a particular expertise. As previously noted, there is strong emerging evidence that student understanding can best be developed over time with a scaffolded progression of ideas and skills—that is, a learning progression (Corcoran et al., 2009; Jin et al., 2013; Stevens et al., 2010). National reform efforts such as the Next Generation Science Standards (NGSS) are now building on this approach (NGSS Lead States, 2013). While the idea of learning progressions has received less attention at the college level, there are a number of reform efforts underway (e.g., Cooper & Klymkowsky, 2013; Sevian & Talanquer, 2014). The recent emphasis on the reform of the first two years of college, as noted in the President's Council of Advisors on Science and Technology (PCAST) report which stated "the first two years of college are the most critical to the retention and recruitment of STEM majors" (PCAST, 2012), means that an increased emphasis on reform efforts will almost certainly continue. As these curricula reforms take hold it will become even more important to understand not only whether students learn a particular idea, but also when that learning first occurs.

RESEARCH QUESTION

How effective is the framework of survival analysis to analyze longitudinal assessment data for identifying *whether* and *when* student learning occurs?

To address this research question, we use data collected from a study that investigated what information students self-report can be predicted from a chemical structure to illustrate how survival analysis can be applied to longitudinal assessment data.

Context: Student Understanding of Structure–Property Relationships

Although the idea that the molecular level structure of a substance can be used to predict macroscopic properties is essential for understanding chemistry, this connection is known to be quite difficult for students (Cooper, Corley, & Underwood, 2013; Cooper, Grove, Underwood, & Klymkowsky, 2010; Cooper, Underwood, & Hilley, 2012; Cooper, Underwood, Hilley, & Klymkowsky, 2012; Gilbert & Treagust, 2009; Kozma & Russell, 1997). In our previous studies on this topic, we investigated how students use Lewis structures since these structures are the first representation introduced to students that encode enough information to allow them to predict phenomena (Cooper et al., 2010; Cooper et al., 2012a, 2012b). The students' difficulty with this relationship is not surprising when we consider the long series of inferences that students must concatenate when learning to predict a compound's macroscopic properties (Cooper et al., 2012a). Ideally, we would want students to be able to construct the structure and use it to determine the three-dimensional geometry and molecular structure, then identify bond polarities and molecular polarity, before determining the strength and types of molecular interactions that are possible. Using this information, students can then predict properties (Cooper et al., 2012a). If students do not connect this sequence of inferences, they may have to rely on heuristics and rules that may or may not be scientifically useful (Cooper et al., 2013). Therefore, it is important that we develop an understanding of whether and when students become aware of each of these ideas. In this study, we use an instrument designed to measure whether students have made these specific connections between structure and properties as described in the Methods section.

METHODS

Participants

The participants in this study were students enrolled in introductory university level chemistry courses, specifically the four-semester sequence of general chemistry and organic chemistry at a research university located in the southeastern United States (Cohort 1 Fall 2010 – Spring 2012 $N = 1443$). Two additional cohorts of students also participated in the study (Cohort 2 Fall 2011 – Spring 2013 $N = 1313$, Cohort 3 Fall 2012 – Spring 2013 $N = 1329$). For the general chemistry courses taught at the institution of interest, the faculty for the different sections met weekly to discuss the course, covered the same material, and administered common exams. While the organic chemistry courses were not coordinated in this way, the same material was covered. The final exam for all second-semester general chemistry and organic chemistry courses at this institution was the American Chemical Society standardized two-semester examination in either general or organic chemistry as applicable (American Chemical Society Examinations Institute, 2015). Informed consent was obtained from all participants.

Assessment Measure: Implicit Information from Lewis Structures Instrument

The *Implicit Information from Lewis Structures Instrument* (IILSI) was used in this study. The development, validity, and reliability testing of this instrument have been previously reported (Cooper et al., 2012a). The IILSI asks students "What information could you determine using a Lewis structure and any other chemistry knowledge you may have? (Mark all that may apply)." The students can select information from the 16 choices provided such as *element(s) present, polarity, acidity/basicity,* and *reactivity*. The answer choice *no information* is also included since some students do not believe that any information can be predicted. The student responses to the IILSI provide instructors and researchers with a rapid, reliable, and valid way to learn what information students believe can be predicted using a Lewis structure (Cooper et al., 2012a, 2012b).

Study Design

The IILSI was administered five times to each cohort of students over the course of two years: (i) prior to instruction during the first semester of general chemistry (Pre GC1), (ii) at the end of the first semester of general chemistry (End GC1), (iii) at the end of the second semester of general chemistry (End GC2), (iv) at the end of the first semester of organic chemistry (End OC1), and (v) at the end of the second semester of organic chemistry (End OC2). Because of external constraints there were the following differences in the timing and number of administrations of the survey between cohorts: (i) Data from Cohort 3 were collected only for general chemistry (for the first three time points) since the study ended in the middle of their two-year course sequence and (ii) The data for Cohort 2 at the first time point (Pre GC1) were recorded slightly later than for Cohorts 1 and 3, and after instruction on Lewis structures began. However, for the purposes of this paper, we will still refer to the first time point of the data set for all three cohorts as pre-instruction measures (see Discussion section for more details).

The IILSI was administered using an online software platform (*SurveyMonkey* for Cohort 1 and *beSocratic* [Bryfczynski, Pargas, Cooper, Klymkowsky, & Dean, 2013; Cooper, Underwood, Bryfczynski, & Klymkowsky, 2014] for Cohorts 2 and 3), and typically took about five minutes for the students to complete. Students were given unique identifiers to deidentify their responses and were tracked as they progressed through the study. All three cohorts of students completed this activity for participation credit in their laboratory course and were asked to take the assignment seriously, in the same manner as the students who participated in our validation studies (Cooper et al., 2012a). Cohorts 1 and 2 completed the IILSI during their laboratory meeting while Cohort 3 completed the IILSI as a homework assignment. Even though the survey was administered under different circumstances, we found that the nature of the administration did not appear to affect student responses since the students' initial responses to the survey were similar for Cohorts 1 and 3, and all subsequent responses were similar for all three cohorts.

All of the ideas covered by our assessment instrument are initially introduced during the first semester of general chemistry (GC1) and then used in every semester after that through the second semester of organic chemistry (OC2).

RATIONALE FOR USING DISCRETE-TIME SURVIVAL ANALYSIS

Although there are other statistical models that could have been used for this study, we determined that the best choice was survival analysis since (i) the analysis can be performed in a single step (i.e., use a single analysis for the whole data set rather than performing multiple tests) and (ii) we want to know *when* learning occurs (i.e., when students initially connect a structure with particular properties). Although various types of survival analyses are available, we specifically used a discrete-time survival analysis since the time to event is not considered a continuous value. That is, there were only five time points of interest evaluated for this study (pre-instruction and at the completion of each additional semester of instruction). The other main reason for selecting discrete-time survival analysis stems from its ability to provide a more flexible modeling technique than some of the other types of survival analysis, since we had no expectations in terms of the trajectories that the student learning would take over time.

How Discrete-Time Survival Analysis Handles Attrition

In most longitudinal studies, attrition becomes an issue since not all students continue with the same sequence of courses. For example, in this study not all students are required to take two years of chemistry at the institution of interest. In fact, only about 10% of the students who were initially enrolled in GC1 completed *all* four semesters of general and organic chemistry during the two-year period. The issue of attrition considerably limits the available methods of analysis since many statistical models (e.g., linear regression) require the exclusion of participant responses that contain missing data by performing listwise deletion.

It should be noted that students do not continue in chemistry for a variety of reasons, most of which have nothing to do with ability. For example, many students leave not because they are failing, but because of other reasons such as they are not interested or find the introductory science courses unstimulating (Khan, 2005; Lastusaari & Murtonen, 2013; Lewis, Shaw, Heitz, & Webster, 2009; Seymour & Hewitt, 1997; Tai, Sadler, & Loehr, 2005). Others may simply take a semester off, which would remove them from a study such as this. Majors such as engineering (who can make up about 60% of the student population in general chemistry at this institution) do not require four semesters of chemistry, and are responsible for a large percentage of the drop-off in enrollment after general chemistry.

Traditional approaches to address missing data consist of deleting responses that are incomplete (i.e., listwise deletion) or imputing data (Cox, McIntosh, Reason, & Terenzini, 2014). Unfortunately, neither of these approaches for missing data are ideal. Deleting all responses that contain missing data means that we cannot know whether this large cohort of students ever learns. Consider, for example, a student who leaves the study after the first three time points (two semesters) who has not selected a particular property. We know that it would have taken the student at least four time points to select the specific property, but by omitting the student from the study we are losing all the associated data. We could impute the missing data (Hsu et al., 2007), but even then we would have to use an appropriate statistical analysis to answer the questions of *whether* and *when* the student learns. By using survival analysis and its maximum likelihood estimation procedure (Allison, 2010; Hosmer et al., 2011; Lee & Wang, 2003), we can use the available information (for example that the student in question will take at least four time points to learn).

At each time period, only the students who are left in the study contribute to the analysis, but for this to be a valid approach we must assume that the students who left the study were equivalent (as a group) to those who stayed. In our case, this appears to be true: after general chemistry (where the largest drop-off in enrollment occurs), students who continue into organic chemistry and those who do not continue had no significant differences in their earlier responses on the IILSI and similar grades in general chemistry (a B average). Although there was a slight but significant difference between the two groups in their grades, the effect size was small [Mann–Whitney, $U = 72{,}265.0$, $Z = -5.61$, $p < .001$, $r = .19$]. For the purposes of this study, we believe the two groups of students were similar, meaning that we could use available responses for students who continue taking introductory chemistry courses to extrapolate results for the students who do not continue.

Survival Analysis Can Handle Dichotomous Data and Removes the Need for Multiple Comparisons

We chose to use survival analysis for this study because the IILSI requires use of statistical methods designed for dichotomous-dependent variables, meaning that students either select (given a value of 1) or do not select (given a value of 0) that a given piece of information can be predicted. Parametric methods (e.g., linear regression and independent samples *t*-test) and even more sophisticated models such as growth models, which are designed for longitudinal data, are not applicable because of the nature of our dichotomous student responses and our interest in studying the time of its occurrence. If survival analysis was not employed, one possible approach could be to perform many individual McNemar tests to investigate change *within* each cohort for the subsequent semesters and chi-square tests to determine differences *between* the three cohorts for each IILSI administration. Additional ad hoc tests would probably be needed for the chi-square analyses to determine which pairs of groups are significantly different. Therefore, to avoid performing all of these separate analyses with the same set of data (which can lead to questionable inferences [Bland & Altman, 1995]), we selected a single statistical model to analyze the data. As Bland and Altman have written, "if we go on testing long enough we will inevitably find something which is 'significant'" (1995).

While it might be tempting to use a traditional logistical regression model for this study where identification of the property is the dependent dichotomous variable (i.e., information is learned or information is not learned), it could not answer the question of *when* the learning occurred. Even if the dependent variable instead symbolized the length of time for the learning to occur, the model would still not account for the majority of our students who would be censored because they leave the study before learning the concept. Allison has further explained the problems with these traditional approaches using a study investigating the time for inmates to be rearrested after release (2010).

When selecting to use discrete-time survival analysis for these data, we considered its ability to provide a more flexible modeling technique over some of the other types of survival analysis. That is, some types of survival analysis methods assume specific relationships in the resulting curves (i.e., exponential), however discrete-time survival analysis does not. This flexibility was important since we did not know what to expect in terms of shape of the trajectory that would result when investigating student learning.

Discrete-time survival analysis is a type of regression analysis, and as with any regression analysis, an equation model should be identified to capture the different variables of interest and the patterns within the data. The survival analysis performed in this study includes a time variable, a cohort variable, and the interaction between the time and the cohort variables. The interaction term was included in the model since it was assumed that there could be differences in the trajectories between the cohorts. The coding of the discrete time variables was chosen to capture the change between two consecutive time values (e.g., the difference between Pre GC1 and End GC1 was labeled as the Time 1–2). Similarly, the variables Time 2–3, Time 3–4, and Time 4–5 were also created. The cohort variable indicated which group of students was being evaluated: Cohort 1, 2, or 3.

DATA ANALYSIS

The student responses from each semester were matched using their unique identifiers. It is important to note that only responses from students who were originally enrolled in GC1 at the start of each cohort were included in our analysis since we were interested in identifying when cohorts of students (from beginning of Pre GC1 to End OC2) started to connect structure–property relationships. That is, students who moved into the course sequences such as high-school students with Advanced Placement course credit or transfer students from other universities were excluded from this study. If a researcher was interested in using discrete-time survival analysis to include students who enter the study after data collection has begun, the data would need to be coded so that some of the students in this subset are identified as left-censored. For example, if a student enters the study at time point four (first course of organic chemistry) and the student selects a property, then we would not know the accurate time of the event, and we would only know that the accurate time is a value less than or equal to four. We chose not to complicate the analysis at this time.

The data set was then transformed from a "one-person with one-record" format (i.e., person-oriented data set) into a "one-person with multiple-records" format (i.e., person-period data set) using SPSS 21 (SAS or R could be used as suitable alternatives). Singer and Willett provide an example how the data sets are transformed from a person-oriented into a person-period data set (1993). Once the data set is in the proper format, it is important to consider, as mentioned above, if the event can occur more than once during the length of the study (Allison, 2010; Singer & Willett, 1991). While (obviously) death can only occur once, identification of a particular property may fluctuate from semester to semester. Although some types of survival analysis can account for recurring events, the resultant analysis process can become quite complex (Allison, 1984). Therefore, since our initial goal was to determine if survival analysis is a useful analysis framework, we chose discrete-time survival analysis to investigate the *first time* students connect structure–property relationships.

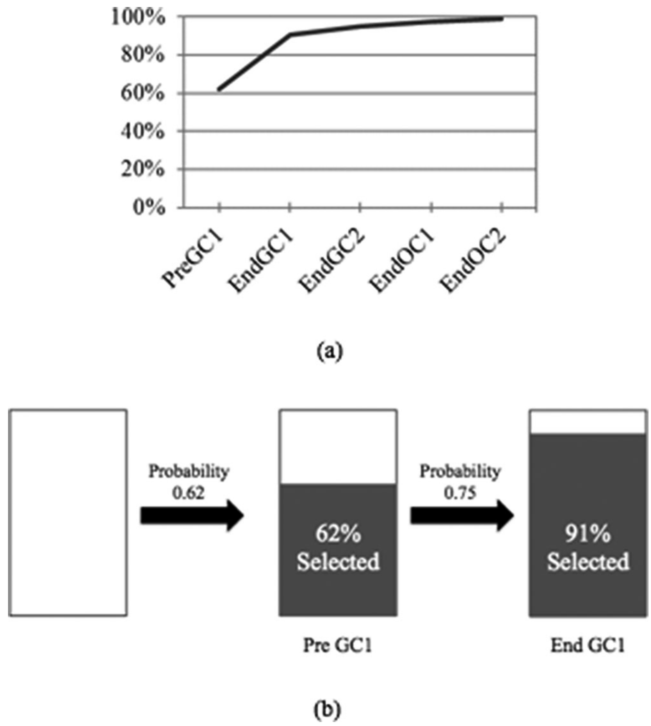

Figure 1. (a) Competency curve for the IILSI item "element(s) present" for Cohort 1 students and (b) example of how competency curve for the IILSI item "element(s) present" for Cohort 1 students is constructed. Note that Pre GC1 is prior to instruction during the first semester of general chemistry, End GC1 is at the end of the first semester of general chemistry, End GC2 is at the end of the second-semester of general chemistry, End OC1 is at the end of the first semester of organic chemistry, and End OC2 is at the end of the second semester of organic chemistry.

This analysis approach is specific for nonrepeatable (single occurrence) events. To use the discrete-time survival analysis, the data must first be restructured to remove repeatability in student responses for the subsequent semesters. The procedure for this process is presented in Supplemental Materials in the Supporting Information.

Finally, the data set was transferred into SAS 9.3 to run the *SurveyLogistic* procedure (S.A.S. Institute, 2012). This procedure within SAS 9.3 allows us to estimate the selected model and also account for the relationship between the multiple records of the same student (i.e., person-period data). The statistical output from the discrete-time survival analysis reports the estimated values in terms of the logarithm of the odds (log odds). Equation 1 helps us determine the odds that an event will occur as related to probability of an event occurring. The probabilities allow us to determine the percent of students who collectively have selected the various IILSI items by the end of each time period. From these probabilities, we can construct competency curves (often referred to as cumulative incidence curves) to identify *when* students first connect a molecular structure with other information that is encoded within the structure (Figure 1a).

$$\text{Odds} = \frac{(\text{Probability of understanding concept})}{(\text{Probability of not understanding concept})} \qquad (1)$$

RESULTS AND DISCUSSION

Here, we describe a simplified explanation of how the model uses the available student data at each time period to extrapolate student ideas for those who do not continue with the four sequence courses. When analyzing the IILSI question "What information could you determine using a Lewis structure and any other chemistry knowledge you may have? (Mark all that may apply)," we found an initial probability of 0.62 or that 62% of the students selected the IILSI item of *element(s) present* prior to instruction in GC1 (Pre GC1), as shown in Figure 1b. Since these students already know that a structure can provide information about the elements present, the model now ignores these students in subsequent analyses.

The analysis continues with the remaining 38% of the population who did not select the IILSI item in Pre GC1, which includes both students who continue and students who will not continue in subsequent semesters. For the next time period, the probability that a student would now select the particular type of information for the first time is 0.75. This means that after instruction in GC1, 75% of the remaining 38% of students (who had not selected the specific type of information during the first administration) now select this choice, meaning that now a total of 91% (i.e., 62% + 29%) of students have selected this item by the end of the first semester of general chemistry. If we continue this analysis for the five administration periods, we can construct a trajectory showing how students' competence develops over time in Figure 1a. More information about this process can be found in the Supplemental Materials in the Supporting Information.

The IILSI has 16 different items that students may choose (excluding the item *no information*), and therefore we were able to generate 16 different competency curves, one for each IILSI item. We examined each curve for similarities and differences. Although all the curves possess a positive slope, meaning that with each additional semester of instruction more students make a connection between structure and the relevant information, the curves had different shapes and slopes meaning that not all of the competencies developed in the same way or over the same time period.

To obtain an unbiased understanding of the similarities among the trajectory to competence curves, two postdoctoral researchers and four graduate students (none of whom are authors on this paper) grouped the unlabeled graphs into sets that appeared to have similar trajectories. This grouping process would not have been relevant had there only been a single question of interest or several unrelated questions of interest instead of the 16 IILSI items, which are very much related.

The first set of IILSI items, unanimously grouped together, consisted of surface level features of the structure. That is, information that can be directly determined from the chemical structure (*number of valence electrons*, *type of bond(s)*, *element(s) present*, and *number of bonds between particular atoms*; Figure 2). In fact, when students enter the first semester of general chemistry about 60% indicate that this surface level information can be determined. After instruction during the first semester of general chemistry, the probability that a student will choose these items significantly increases to about 90% of the population having selected these types of information at least once. This large increase in the probability indicates that an external factor (i.e., instruction) has occurred. That is, the first-semester course in the four-semester sequence has helped students learn that Lewis structures contain information about molecular structure such as number and type of atoms and bond angles. This is not surprising since the topic of Lewis structures is taught during the first semester of general chemistry. This finding is consistent with our prior work where all the students in the study reported that Lewis structures could be used to predict structural information (Cooper et al., 2010).

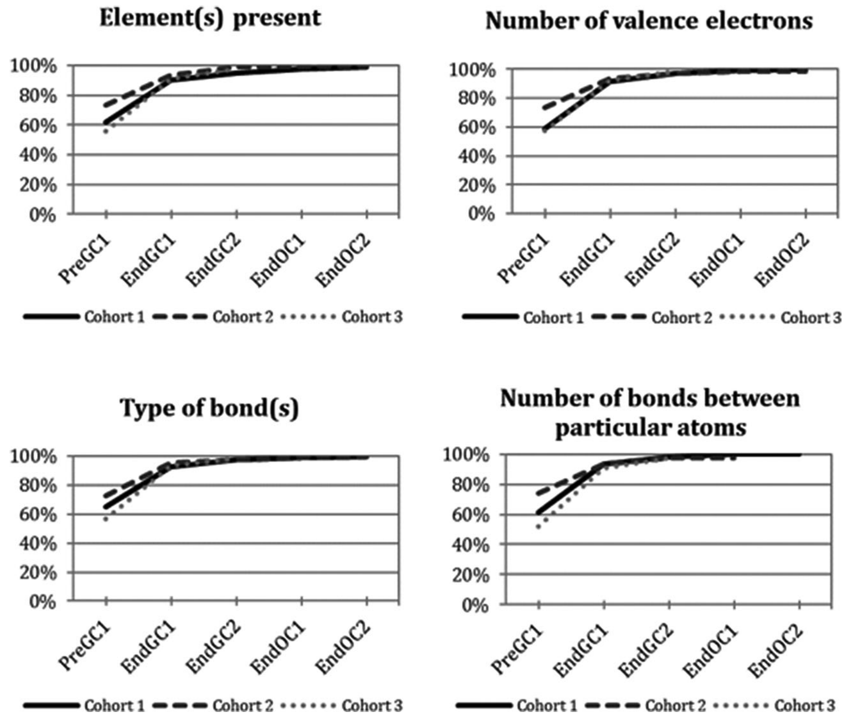

Figure 2. The four IILSI items included in the structural information group.

As noted in the Methods section, the “Pre GC1” data were collected at a slightly later date (after the topic of Lewis structures had been introduced) for Cohort 2. As can be seen in this group of graphs, Cohort 2 students have a higher probability of selecting these items than the other cohorts. However, probabilities at later time periods are similar for all three cohorts. These small, expected differences between the cohorts in the analysis of the data agree both with our previous study that involved the IILSI, and our expectations given the content covered in class. Further discussion and examples for using discrete-time survival analysis are found in Supplemental Materials in the Supporting Information.

The coders were also in complete agreement on a second group of graphs, which included chemical and physical properties (*acidity/basicity*, *reactivity*, *relative boiling point*, *relative melting point*, and *physical properties*; Figure 3). These properties require students to use the Lewis structure and information encoded within them to predict properties. That is, the first group of trajectories was composed of surface-level features that can be determined from examining the Lewis structure alone, whereas this second group requires students to use other knowledge such as polarity and intermolecular forces in addition to the Lewis structure to predict these implicit underlying chemical/physical properties. In the first group, the Lewis structure is a representation, but in the second group the Lewis structure must be used as a model to predict and explain phenomena.

At the beginning of GC1, only about 10% of students indicate that they can predict chemical or physical properties using a structure. The initial low selection rate is understandable since most students have not been taught to use structures in this way. The low selection rate attests to the validity of the responses provided by the students on the IILSI; that is, students do not select categories randomly. At the end of GC1, the number of additional students who chose these items was still very small. While there are significant increases in

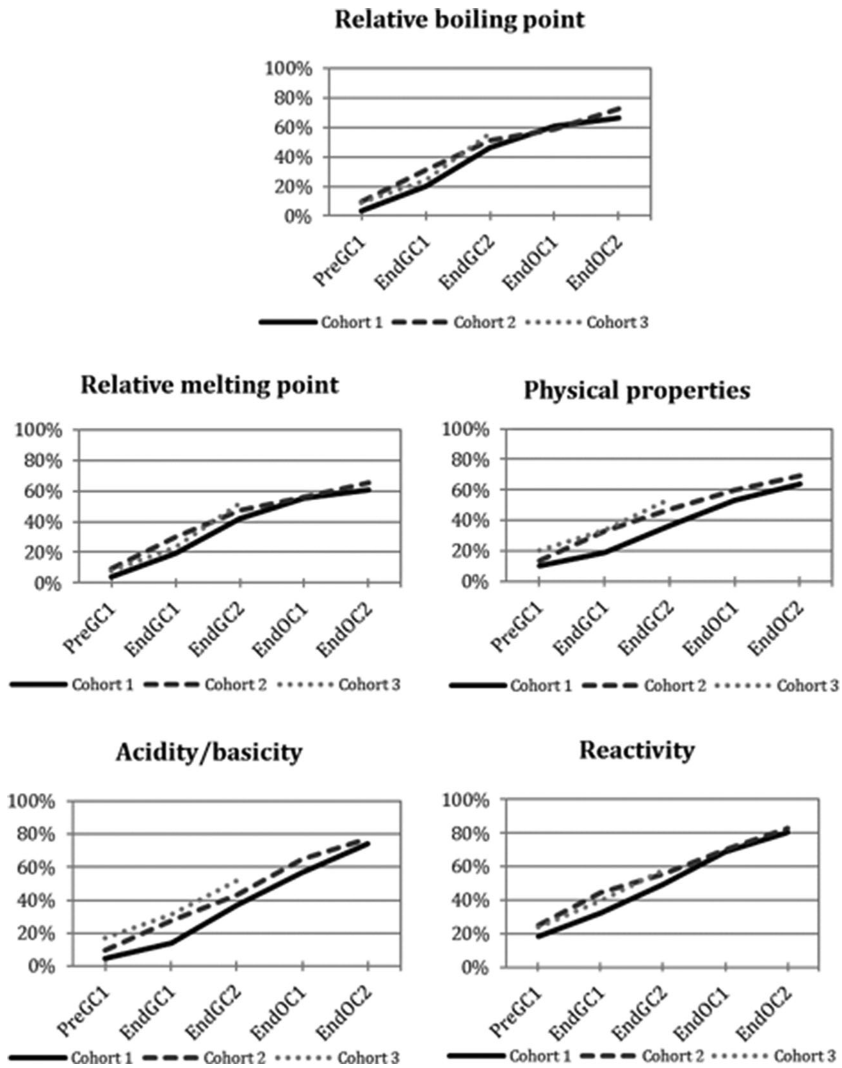

Figure 3. The five IILSI items included in the chemical/physical property information group.

the probabilities for subsequent semesters, the linear trajectory suggests that this skill takes time for many students to develop and there is not one particular semester that appears to help students connect these ideas. Even by the end of OC2 (after two full years of chemistry), there are still many students who have not made the connection between structure and properties. In fact, about 30% of the students never selected some types of physical properties and 20% gave similar responses for chemical properties.

The marked differences in the shapes of the competency curves between the first and second groups of trajectories point out the importance of using this particular type of survival analysis (i.e., discrete-time survival analysis). As previously discussed, some types of survival analysis require prior knowledge about the shape of the curves (i.e., linear, exponential, etc.); therefore, since we had no prior knowledge of the trajectories that would emerge, nor that they would be different for different items, discrete-time survival analysis for these data sets was indeed appropriate and useful. In fact, using a different model for the

TABLE 1
Comparing the Change in Odds for the IILSI Item "Acidity/Basicity" for Cohorts 1 and 2

Semester Comparison	Cohort 1 Change in Odds	Cohort 2 Change in Odds	*p*-Value for comparing Cohort 2's Change in Odds to Cohort 1's Change in Odds
Pre GC1 to End GC1	0.05–0.11	0.11–0.25	.70
End GC1 to End GC2	0.11–0.36	0.25–0.27	<.001
End GC2 to End OC1	0.36–0.45	0.27–0.63	.039
End OC1 to End OC2	0.45–0.68	0.63–0.56	.25

survival analysis would have forced the trajectories into a specific shape and the behavior differences of the second group of curves might have not been identified.

As noted earlier, there are very few longitudinal studies that have been replicated. We compared the findings for Cohort 1 (the reference cohort) to two subsequent cohorts of students (Cohorts 2 and 3). Using survival analysis to compare the multiple cohorts of students provided us with a single analysis method, thus preventing multiple analyses that would have otherwise been necessary, as described above. To determine if the cohorts' results produced similar trends in the competency curves, we compared the change in the probability that a student would select a type of information on the IILSI for each pair of semesters (e.g., Pre GC1 to End GC1) in Cohort 1 to the change in probability for Cohorts 2 and 3 in that same time period. For example, using the output from the analyses on the IILSI responses from Cohorts 1 and 2 for the item *acidity/basicity* (Table 1), we found that there was no significant difference in the change of probability for the two cohorts from Pre GC1 to End GC1 (p-value .70) and from End OC1 to End OC2 (p-value .25). With regard to the changes for the remaining semesters (from End GC1 to End GC2 and from End GC2 to OC1), we found that the two cohorts were statistically different (p-value $< .001$ for End GC1–End GC2 and .039 for End GC2–End OC1). We believe these small differences between the two cohorts may result from changes in how the material was taught during the second semester of general chemistry and first semester of organic chemistry. For example, we know that instructors were aware of our research, and it may be that our initial research findings prompted a somewhat stronger emphasis on structure and acid/base relationships in Cohort 2.

Each pair of time points between Cohort 1 (reference cohort) and Cohort 2 were examined, followed by additional comparisons of Cohorts 1 and 3 for all 16 items. We found that the three cohorts produced very similar trajectories as shown in Figures 2 and 3. There appears to be little difference from year to year, meaning that the results found from Cohort 1 were replicated using discrete-time survival analysis for Cohorts 2 and 3. All of the competency curves for the 16 IILSI items (excluding the answer choice *no information*) can be found in Supplemental Materials in the Supporting Information.

CONCLUSIONS

In this paper, we present one possible approach to using discrete-time survival analysis to evaluate longitudinal assessment data. The main use of this analysis method is to answer the question of *whether* and *when* an event has occurred—in other words, it analyzes the time to an event. Here, we have provided an example study where we determined *whether* students self-report they could determine particular information from a chemical structure in addition to *when* students report this connection. Discrete-time survival analysis provided us a means by which we could build upon our previous research that used cross-sectional studies to examine how students connect structure and property relationships (Cooper et al., 2010; 2012a).

While our previous results indicated that all students enrolled in traditional chemistry courses could predict structural information and only about half of them able to predict chemical/physical properties (Cooper et al., 2010), it was only when we used survival analysis that we could understand *when* multiple groups of students developed this initial connection between structure and property. Specifically while almost all students know that a Lewis structure contains structural information after the first semester of general chemistry instruction (i.e., time point two [End GC1]), the more difficult task of connecting structure with properties takes much longer, and for many students this connection never occurs at all (as evident in the linear shape of the competency curves and their maximum value reached at time point five, at the end of the second semester of organic chemistry [End OC2]). This finding was important since we know that the majority of students enrolled in general chemistry will not go on to take the second semester of organic chemistry, meaning that about 50% of the total number of beginning students leave chemistry without ever recognizing that Lewis structures are models that can be used to predict and explain.

Selecting discrete-time survival analysis over other models for survival analysis proved to be essential since we had no prior knowledge of the shape of the competency curves that would emerge from the analysis. As shown in Figures 2 and 3, the trajectories are quite different for the different sets of data. The first group of answer choices on the IILSI resulted in exponential curves while the second group consisted of linear curves. If we had selected a different model for the survival analysis method, we would have had to predetermine the shape of the curve, rather than letting it emerge from the data.

As previously noted, with most longitudinal studies, the problem of attrition makes following cohorts of students quite difficult. Even in this study, while our initial sample sizes were relatively large (Cohort 1 $N = 1443$, Cohort 2 $N = 1313$, and Cohort 3 $N = 1329$), attrition was a major concern since only 10% of our sample completed all four semesters of introductory chemistry. This analysis framework allowed us to use all the available student responses, whereas if we used more traditional methods much of the available data would have been removed during the analysis process because of missing data. Using the responses from students who continue into subsequent semesters of chemistry, we were able to predict a pattern for the general population. This procedure also allowed us to use all the information available both from the students who completed the sequence of four courses and from those who did not complete the sequence. The analysis and the procedure become especially important with smaller sample sizes (see Supplemental Materials in the Supporting Information), thus providing researchers with a viable method to track smaller cohorts of students ($N = 100$). In fact, our current work focuses on comparisons between much smaller cohorts of students (Underwood, Reyes-Gastelum, & Cooper, Manuscript in preparation).

IMPLICATIONS

A discrete-time survival analysis framework provides researchers with ways to investigate a wider range of longitudinal studies, including those involving multiple cohorts of students for replication purposes, where traditional analysis measures are not appropriate. Although we present here a study with a self-report instrument that contains only one question with 16 answer choices, there is a range of potential scenarios for which this analysis framework might also be appropriate. The limitations on using discrete-time survival analysis are that the event of interest must be dichotomous, that is, the event occurs or does not occur (partial occurrence of the event is not an option). Possible examples include determining when students answer a particular question correctly or reach a certain level of understanding. We present here two examples of how this analysis framework could be used for future studies.

Survival analysis could be used to analyze student responses to multiple-choice questions. For example, concept inventories are often administered to determine student understanding of specific topics (e.g., Hestenes, Wells, & Swackhamer, 1992; Klymkowsky, Underwood, & Garvin-Doxas, 2010; McClary & Bretz, 2012; Mulford & Robinson, 2002; Treagust, 1988). If these inventories are administered over several courses (rather than just pre/post as is often the case), discrete-time survival analysis could be used to determine *whether* and *when* students answer a particular question correctly. While this analysis method would allow researchers to investigate if students eventually answer a question correctly, it will not indicate which answer choice the students selected. If a researcher was interested in tracking all of the responses, however, each response could be analyzed in a similar manner to how we analyzed the IILSI 16 answer choices presented here.

Discrete-time survival analysis could also be used for open-ended assessment questions if the responses were analyzed using rubrics coding for levels of sophistication (e.g., the ChemQuery system [Claesgens, Scalise, Wilson, & Stacy, 2009], or other assessment approaches that use Rasch analysis to produce a construct map [Wilson, 2005])—that is, where each level represents a more sophisticated student response. While the framework cannot be used to measure how student understanding changes over time (increasing or decreasing in levels of sophistication), it could be used to identify whether and when a student reached a certain threshold, thus creating a dichotomous variable to define the event. The idea of reporting when a threshold is reached has been used previously; for example, Singer and Willett report when 50% of the teachers have left the profession (1993). In our present study, we could have examined the time taken for 50% (or 90%) of the students to choose a particular item from the instrument.

LIMITATIONS OF SURVIVAL ANALYSIS AND THIS STUDY

Discrete-time survival analysis is designed to answer the question of *whether* and *when* an event occurs. A consequence of using this method would be that the results of the analysis depend on the definition of the event to be monitored. Therefore, it is important to use a substantive definition that reflects the research question of interest. For example, if we consider the possible use discussed in the implications section for determining when students achieve a certain level of sophistication, the analysis would not be able to indicate whether student learning was increasing or decreasing over time but instead inform the researcher of the length of time it takes for students to meet a specific threshold of sophistication in their knowledge.

In the analysis performed on the IILSI data, survival analysis allowed us to identify when a student *first* became aware of the particular type of information that is encoded within a Lewis structure. However, it did not reveal whether students were actually able to (for example) correctly predict how an acid might react, or predict the relative boiling points for a set of related compounds. We realize this is a limitation of the IILSI itself and not of the method.

We would also like to thank the Center for Statistical Training and Consulting at Michigan State University for its assistance.

REFERENCES

Allison, P. D. (1984). Event history analysis: Regression for longitudinal event data. Newbury Park, CA: Sage.

Allison, P. D. (2010). Survival analysis using SAS: A practical guide. Cary, NC: SAS Institute.

Allison, P. D. (2014). Event history and survival analysis (2nd ed.). Thousand Oaks, CA: SAGE.

American Chemical Society Examinations Institute. (2015). Retrieved from http://chemexams.chem.iastate.edu/.

Association of American Universities. (2014). Undergraduate STEM education initiative. Retrieved from https://stemedhub.org/groups/aau.

Bauer, K. W. (2004). Conducting longitudinal studies. New Directions for Institutional Research, 2004(121), 75 – 90.

Bland, J. M., & Altman, D. G. (1995). Multiple significance tests: The Bonferroni method. British Medical Journal, 310(6973), 170.

Bowers, A. J. (2010). Grades and graduation: A longitudinal risk perspective to identify student dropouts. The Journal of Educational Research, 103(3), 191 – 207.

Bretz, S. L. (2001). Novak's theory of education: Human constructivism and meaningful learning. Journal of Chemical Education, 78(8), 1107 – 1117.

Bryfczynski, S., Pargas, R. P., Cooper, M. M., Klymkowsky, M., & Dean, B. C. (2013). Teaching data structures with beSocratic. Paper presented at Proceedings of the 18th ACM Conference on Innovation and Technology in Computer Science Education (pp. 105 – 110). New York, NY: ACM.

Claesgens, J., Scalise, K., Wilson, M., & Stacy, A. (2009). Mapping student understanding in chemistry: The perspectives of chemists. Science Education, 93(1), 56 – 85.

Cooper, M. M., Corley, L. M., & Underwood, S. M. (2013). An investigation of college chemistry students' understanding of structure – property relationships. Journal of Research in Science Teaching, 50(6), 699 – 721.

Cooper, M. M., Grove, N., Underwood, S. M., & Klymkowsky, M. W. (2010). Lost in Lewis structures: An investigation of student difficulties in developing representational competence. Journal of Chemical Education, 87(8), 869 – 874.

Cooper, M. M., & Klymkowsky, M. W. (2013). Chemistry, life, the universe and everything: A new approach to general chemistry, and a model for curriculum reform. Journal of Chemical Education, 90(9), 1116 – 1122.

Cooper, M. M., Underwood, S. M., Bryfczynski, S. P., & Klymkowsky, M. W. (2014). A short history of the use of technology to model and analyze student data for teaching and research. In R. Cole & D. Bunce (Eds.), Tools of chemistry education research (pp. 219 – 239). Washington, DC: American Chemical Society.

Cooper, M. M., Underwood, S. M., & Hilley, C. Z. (2012a). Development and validation of the implicit information from Lewis structures instrument (IILSI): Do students connect structures with properties? Chemistry Education Research and Practice, 13(3), 195 – 200.

Cooper, M. M., Underwood, S. M., Hilley, C. Z., & Klymkowsky, M. W. (2012b). Development and assessment of a molecular structure and properties learning progression. Journal of Chemical Education, 89(11), 1351 – 1357.

Corcoran, T., Mosher, F. A., & Rogat, A. (2009). Learning progressions in science: An evidence based approach to reform (No. RR-63). New York, NY: Teachers College, Columbia University, Consortium for Policy Research in Education.

Cox, B. E., McIntosh, K., Reason, R. D., & Terenzini, P. T. (2014). Working with missing data in higher education research: A primer and real-world. Review of Higher Education, 37(3), 377 – 402.

Firth, D., Payne, C., & Payne, J. (1999). Efficacy of programmes for the unemployed: Discrete time modelling of duration data from a matched-comparison study. Journal of the Royal Statistical Society: Series A (Statistics in Society), 162(1), 111 – 120.

Gilbert, J. K., & Treagust, D. F. (Eds.). (2009). Multiple representations in chemical education. Dordrecht, the Netherlands: Springer.

Grilli, R., Oxman, A. D., & Julian, J. A. (1993). Chemotherapy for advanced non-small-cell lung cancer: How much benefit is enough? Journal of Clinical Oncology, 11(10), 1866 – 1872.

Guo, S. (2010). Survival analysis. New York, NY: Oxford University Press.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. The Physics Teacher, 30(3), 141 – 158.

Hosmer, D. W. Jr., Lemeshow, S., & May, S. (2011). Applied survival analysis: Regression modeling of time to event data (2nd ed.). Hoboken, NJ: Wiley.

Hsu, C.-H., Taylor, J. M. G., Murray, S., & Commenges, D. (2007). Multiple imputation for interval censored data with auxiliary variables. Statistics in Medicine, 26(4), 769 – 781.

Ishitani, T. T. (2006). Studying attrition and degree completion behavior among first-generation college students in the united states. The Journal of Higher Education, 77(5), 861 – 885.

Jin, H., Zhan, L., & Anderson, C. W. (2013). Developing a fine-grained learning progression framework for carbon-transforming processes. International Journal of Science Education, 35(10), 1663 – 1697.

Khan, S. (2005). Teaching strategis designed to change the undergraduate experience for college women learning chemistry. Journal of Women and Minorities in Science and Engineering, 11(4), 365–388.

Kimber, J., Copeland, L., Hickman, M., Macleod, J., McKenzie, J., De Angelis, D., & Robertson, J. R. (2010). Survival and cessation in injecting drug users: Prospective observational study of outcomes and effect of opiate substitution treatment. BMJ, 341, c3172.

Klymkowsky, M. W., Underwood, S. M., & Garvin-Doxas, R. K. (2010). Biological concepts instrument (BCI): A diagnostic tool for revealing student thinking. arXiv:1012.4501.

Kozma, R., & Russell, J. (1997). Multimedia and understanding: Expert and novice responses to different representations of chemical phenomena. Journal of Research in Science Teaching, 34(9), 949–968.

Lastusaari, M., & Murtonen, M. (2013). University chemistry students' learning approaches and willingness to change major. Chemistry Education Research and Practice, 14(4), 496–506.

Lee, E. T., & Wang, J. W. (2003). Statistical methods for survival data analysis (3rd ed.). Hoboken, NJ: Wiley.

Lewis, S. E., Shaw, J. L., Heitz, J. O., & Webster, G. H. (2009). Attitude counts: Self-concept and success in general chemistry. Journal of Chemical Education, 86(6), 744.

Liu, X. (2012). Survival analysis: Models and applications. Chichester, UK: Wiley.

McClary, L. M., & Bretz, S. L. (2012). Development and assessment of a diagnostic tool to identify organic chemistry students' alternative conceptions related to acid strength. International Journal of Science Education, 34(5), 2317–2341.

Miller, R. G. (1981). Survival analysis. New York, NY: Wiley.

Mulford, D. R., & Robinson, W. R. (2002). An inventory for alternate conceptions among first-semester general chemistry students. Journal of Chemical Education, 79(6), 739–744.

National Research Council. (2012). Discipline-based education research: Understanding and improving learning in undergraduate science and engineering. S. R. Singer, N. R. Nielson, & H. A. Schweingruber, (Eds.), Washington, DC: National Academies Press.

NGSS Lead States. (2013). Next generation science standards: For states, by states. Washington, DC: The National Academies Press.

Novak, J. D. (1998). Learning, creating, and using knowledge: Concept maps as facilitative tools in schools and corporations. Mahwah, NJ: Erlbaum.

PCAST. (2012). Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics.

Porter, S. R., & Umbach, P. D. (2001). Analyzing faculty workload data using multilevel modeling. Research in Higher Education, 42(2), 171–196.

Institute, S. A. S. (2012). SAS/STAT 12. 1 user's guide: Survey data analysis (Book Excerpt). Cary, NC: SAS Institute.

Sevian, H., & Talanquer, V. (2014). Rethinking chemistry: A learning progression on chemical thinking. Chemistry Education Research and Practice, 15(1), 10–23.

Seymour, E., & Hewitt, N. M. (1997). Talking about leaving: Why undergraduates leave the sciences. Boulder, CO: Westview Press.

Singer, J. D., & Willett, J. B. (1991). Modeling the days of our lives: Using survival analysis when designing and analyzing longitudinal studies of duration and the timing of events. Psychological Bulletin, 110(2), 268–290.

Singer, J. D., & Willett, J. B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. Journal of Educational Statistics, 18(2), 155–195.

Singer, J. D., & Willett, J. B. (2003). Applied longitudinal data analysis: Modeling change and event occurrence. New York, NY: Oxford University Press.

Stevens, S. Y., Delgado, C., & Krajcik, J. S. (2010). Developing a hypothetical multi-dimensional learning progression for the nature of matter. Journal of Research in Science Teaching, 47(6), 687–715.

Tai, R. H., Sadler, P. M., & Loehr, J. F. (2005). Factors influencing success in introductory college chemistry. Journal of Research in Science Teaching, 42(9), 987–1012.

Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. International Journal of Science Education, 10(2), 159–169.

Tytler, R. (2009). Longitudinal studies into science learning: Methodological issues. In M. C. Shelley II, L. Yore, & B. Hand (Eds.), Quality research in literacy and science education (pp. 83–105). Dordrecht, the Netherlands: Springer.

Underwood, S. M., Reyes-Gastelum, D., & Cooper, M. M. (Manuscript in preparation). Investigating longitudinal effects of an alternative general chemistry curriculum on student understanding of structure-property relationship.

White, R. T., & Arzi, H. J. (2005). Longitudinal studies: Designs, validity, practicality, and value. Research in Science Education, 35(1), 137–149.

Willett, J. B., & Singer, J. D. (2012). Using discrete-time survival analysis to study event occurrence. In J. Newsom, R. N. Jones, & S. M. Hofer (Eds.), Longitudinal data analysis: A practical guide for researchers in aging, health, and social sciences. (pp. 329–372). New York, NY: Routledge.

Wilson, M. (2005). Constructing measures: An item-response modeling approach. Mahwah, NJ: Erlbaum.

Zhou, C., Chinnam, R. B., & Korostelev, A. (2012). Hazard rate models for early detection of reliability problems using information from warranty databases and upstream supply chain. International Journal of Production Economics, 139, 180–195.

Zwick, R., & Sklar, J. C. (2005). Predicting college grades and degree completion using high school grades and SAT scores: The role of student ethnicity and first language. American Educational Research Journal, 42(3), 439–464.