CrossMark

DEVELOPMENT ARTICLE

Analyzing "real-world" anomalous data after experimentation with a virtual laboratory

Eva Erdosne Toth[1]

Published online: 16 November 2015
© Association for Educational Communications and Technology 2015

Abstract Developing effective pedagogies to help students examine anomalous data is critical for the education of the next generation of scientists and engineers. By definition anomalous data do not concur with prior knowledge, theories and expectations. Such data are the common outcome of empirical investigation in hands-on laboratories (HOLs). These aberrations can be counter intuitive for students when they investigate real-world processes with technology tools, such as virtual laboratories (VRLs), that may project a simplified view of data. This study blended learning with a VRL and the examination of real-world data in two engineering classrooms. The results indicated that students developed new knowledge with the VRL and were able to apply this knowledge to evaluate anomalous data from an HOL. However, students continued to experience difficulties in transferring their newly constructed knowledge to reason about how anomalous data results may have come about. The results provide directions for continued research on students' perceptions of anomalous data and also suggest the need for evidence-based instructional design decisions for the use of existing VRLs to prepare science and engineering students for real-world investigations.

Keywords Science and engineering student cognition · Reasoning about anomalous data · STEM learning and instruction · Instructional use of virtual laboratories

Introduction and rationale

Anomalous data are results of real-world investigations that do not concur with prior knowledge, theories, or anticipated outcomes. These data have various forms (numbers, images and graphs), and are also called outliers, aberrations, error, or noise (Chandola et al.

✉ Eva Erdosne Toth
    eva.toth@mail.wvu.edu

[1] Department of Curriculum and Instruction/Literacy Studies, College of Education and Human Services, West Virginia University, 604L Allen Hall, Morgantown, WV 26505, USA

2009; Gong et al. 2012; Han et al. 2006). Scientific, engineering, and medical malpractice frequently originates from the improper evaluation of and response to anomalous data (Allchin 2001; Dekker 2006; Dorner 1996; Spector and Davidsen 2000; US News 2007). Therefore, the responsible and accurate use of anomalous data during scientific and engineering practice is paramount (Drenth 2006; Martinson et al. 2005; Masnick and Klahr 2003; NGSS 2012). However, science and engineering are no longer limited to hands-on investigation since powerful computer tools are available to examine natural phenomena through data generation, evaluation, and reasoning (Windschitl 2000; Spector 2008). Of these tools, ViRtual Laboratories (VRLs) are software-tools that allow users to design repeated experiments to test the effects of variables. This is similar to learning with hands-on laboratories (HOLs) but in a shorter amount of time, with increased safety, and at a reduced cost (Ma and Nickerson 2006; Toth 2009; Zacharia et al. 2008). A variety of VRLs are now available from research organizations (HHMI, n.d.; NASA, n.d.), higher education institutions (GSLC 2013; MyDNA 2003) as well as educational vendors. Yet, the processes of developing software-supported learning environments are complex (Reiser 2004) and there is a continued need to support practitioners (teachers, instructors, professors) in effectively using technology for teaching and learning (Kirschner and Wopereis 2003). This study responds to this need by evaluating the effectiveness of a pedagogical approach that asked students to learn basic concepts with a VRL and use these to analyze anomalous data from a real-world HOL.

Prior research

A birds-eye-view of the history of the integration of software tools for science learning and teaching indicates progress through microcomputer-based laboratories (Jackson et al. 1996; Nakhleh and Krajcik 1993), microworlds (Rieber 1992), instructional simulations (De Jong and Van Joolingen 1998), and weblabs (Dori and Barak 2003; Spector et al. 2013), as well as VRLs (Balamuralithara and Woods 2009; Chen 2010; de Jong et al. 2013; Toth et al. 2009, 2012a, b; Wolf 2010). Current tools allow users to manipulate sophisticated, remote equipment (Corter et al. 2011; Ma and Nickerson 2006; Nedic et al. 2003), and to interact with "virtual worlds" (Dede et al. 1997, 2004; Cobb et al. 2009; Dickey 2011).

A deeper-level examination of the instructional effectiveness of these tools indicates that software design features can support learning (Ainsworth 2008; Jonassen 2006; Kali 2008; Kozma 2003; Quintana et al. 2004; Underwood et al. 2005). Educationally effective software tools can scaffold the process of inquiry by structuring complex tasks and by automating routine ones (Kali 2008; Quintana et al. 2004). They can also constrain the users' activity to important tasks, thereby reducing cognitive load (Ainsworth 2008). Furthermore, properly designed software tools can assist users in inspecting the properties of data (Quintana et al. 2004). Not all educational software tools employ effective design features however, and many even pose challenges for users (Reiser 2004). Therefore, practitioners and instructional designers need to make evidence-based instructional design decisions to reduce potential harm caused by ineffective software designs (Kirschner and Wopereis 2003; Toth 2009). The results of this study support practitioners' instructional design efforts by describing how students learn from repeated experiments with a VRL, and how they use this learning to evaluate and reason about real-world, anomalous data. Two lines of prior research supported the development of

this approach: studies that compared students' learning with VRLs and HOLs and those that examined learning with anomalous data.

Prior research on learning with VRLs and HOLs

Two overarching research perspectives exist on combining or blending VRLs and HOLs. One aims to align students' activities with the two environments in order to document which is more productive, and the other aims to support learning by utilizing differences in each tool. Comparative studies of effectiveness examined goals, activities and assessments and generally indicated that VRLs are more effective for learning (Apkan and Strayer 2010; Ma and Nickerson 2006; Toth et al. 2012b; Zacharia 2007) or are at least equivalent to learning with HOLs (Triona and Klahr 2003; Zacharia and Constantinou 2008).

Studies on the perceptual differences between HOLs and VRLs (Olympiou and Zacharia 2012; Olympiou et al. 2013; Toth et al. 2009, 2012a) are concerned with the nature of students' learning interactions with different tools (Jonassen 2006; Ainsworth 2008). For example, Toth et al. (2012a, b) observed that HOLs focus students' thinking on the manipulation of physical equipment, while VRLs direct users' attention to variables. Similarly, when evaluating data from HOL-experiments, students' focus on the outcomes and not the interaction of variables producing the outcomes (Toth et al. 2012a, b, p. 13). With well-designed VRLs, visual clues may direct attention on the processes by which results come about (Toth et al. 2012a, b, p. 13). Differences in the perceptual features of VRLs and HOLs might explain prior results on students' difficulties in conceptually connecting their VRL-work to the "real-world" (Nedic et al. 2003), especially when learning with VRLs that present a simplified view of investigation and do not generate anomalous data outcomes. Consequently, there is a continued need to examine the effectiveness of instructional methods that combine or blend VRLs and HOLs.

Prior research on learning from anomalous data

A critical aspect of experimenting with "real-world" HOLs is the prevalence of error resulting in anomalous data (Allchin 2001; Chandola et al. 2009; Dorner 1996; Masnick and Klahr 2003). The pedagogical value of learning from anomalous data is that, by definition, these data are inconsistent with expected results. This inconsistency generates cognitive dissonance between what is believed (expected) and what is experienced (Schnotz and Rasch 2005; Sweller et al. 1998). Unexpected outcome can lead to new knowledge construction (Driver 1989; Limón 2001; Piaget 1985) if learners modify their existing, often incorrect, assumptions (Dreyfus et al. 1990; Limon and Carretero 1997; Posner et al. 1982). Using these processes, even very young children can have a basic understanding of variance in data outcome (Masnick and Klahr 2003) and hold sophisticated understandings of scientific investigation (Carey and Smith 1993; Smith et al. 2000; Zimmerman 2007).

Cognitive dissonance between prior knowledge (prior belief) and anomalous data can also have negative consequences and may result in students excluding anomalous data from their reasoning, re-interpreting findings to eliminate anomaly, and even rejecting anomalous data rather than changing prior (incorrect) theories (Chinn and Brewer 1993, 1998; Mason 2001). A possible explanation is that individuals formulate internal

representations or *models of data* that influence their evaluation and reasoning (Chinn and Brewer 2001). Recent research indicates that even college students in science and engineering experience difficulties when evaluating the results of repeated experiments (e.g. Renken and Nunez 2013) and can respond to anomalous data with the same, ineffective strategies documented by Chinn and Brewer (1993).

Students' difficulties can originate from a variety of problems in transferring learning with a VRL to real-word application. For example, students could experience difficulties in knowledge construction in the first place (Salomon and Perkins 1989). Alternatively, students may fail to recognize the explicit and implicit similarities between the learning environment and the application environment (Nokes-Malach and Mestre 2013; Rebello et al. 2005; Magin and Kanapathipillai 2000) and may fail to connect learning with a VRL and real-world data (Nedic et al. 2003). Therefore, this study examined whether university engineering students (a) learn important concepts with a VRL, (b) make a conceptual connection between their learning with the VRL and the evaluation of data from an HOL and whether they (c) use their learning with the VRL when they reason about real-world anomalous data. The research questions were as follows.

1. Does investigation with a VRL assist engineering students in constructing new knowledge?
2. In what ways do students use their learning with a VRL to evaluate anomalous data from an HOL?
3. In what ways do students use the variables they examined with the VRL to reason about anomalous data from an HOL?

Methods

The overall approach was constructivist in nature and it centered on the primacy of students' knowledge development processes (Guba and Lincoln 1994). This approach is also termed "naturalistic inquiry" in the early literature (Lincoln and Guba 1985). From this perspective, knowledge is subject to continuous revision as students engage with the context of learning and form "ever more informed and sophisticated constructions" (Guba and Lincoln 1994, p. 114). Given the focus of this study on using concepts and processes learned in the context of a VRL to be applied for examining anomalous data from the real-world, we provided a learning environment where "relatively different constructions are brought into juxtaposition" (Guba and Lincoln 1994, p. 113).

Instructional context

The setting of instruction was a newly developed course on "Cellular Machinery," designed and taught by an engineering faculty at a large research university in the United States. This semester-long, biomedical engineering course incorporated traditional lecture-discussion with problem-solving projects, group activities, and project presentations. One of the projects was used in this study. Students' learning goal was to develop an in-depth understanding of the concepts and processes of DNA fragment separation by gel-electrophoresis with the MyDNA VRL (2003).

Instructional tools: gel-electrophoresis with the MyDNA virtual laboratory

Molecular biology is an interdisciplinary field of science and gel-electrophoresis is a common laboratory procedure, designed to separate different-sized fragments of DNA—Deoxyribonucleic acid, the genetic materials in living cells. The procedure can be used to compare DNA samples from different individuals for identification, paternity testing, and disease detection. The science of the process rests on the fact that the DNA molecule is negatively charged; thus, in a porous medium, fragments migrate to the positive pole under electric current. The porous medium (often an agarose gel) functions as a sieve and separates fragments by size with larger ones captured close to the loading points (wells on the top of the gel), and smaller ones captured farther down the gel, closer to the positive pole.

Students in biomedical engineering are required to be familiar with the variables, processes, and possible anomalous outcomes of this common laboratory technique. The traditional teaching approach uses a hands-on laboratory that can produce unexpected, anomalous results. A website titled "Hall of Shame" by Rice University (Caprette 1996) hosts a substantial database of anomalous outcome gel images—to the delight and pedagogical benefit of students and faculty. While anomalous outcomes are common with HOLs, determining the sources of anomaly by experimentation (by changing variable settings and producing repeated tests), is not feasible due to cost and lack of time.

With the MyDNA VRL (2003), students could easily design and run repeated experiments and study the effects of different variables (concentration, voltage, time of run) on the distance DNA fragments travel. They are able to stop the process and study the movement of different fragments over time. However, this particular tool does NOT allow users to generate anomalous results and it presents a rather simplified world of experimentation (Chen 2010). To help students relate the idealized world of this VRL to working in the real world, the instruction used existing images of anomalous outcomes from the "Hall of Shame" website instead of performing costly and time-consuming HOLs.

Participants

The participants were graduate and undergraduate students majoring in chemical, mechanical and aerospace or computer engineering and biology. The "Cellular Machinery" course was an elective and students participated based on their interests.

In study one, during the first year, 21 students participated (six females, 15 males; 16 undergraduate students and five graduate students). In study two, during the second year, 17 students participated (eight females, nine males; 13 undergraduate students and four graduate students). All students were required to complete the classroom work, but they could voluntarily withhold their results from inclusion in this study. As a result, 16 students in study one and 15 students in study two participated. The university's Institutional Review Board approved the study.

Instructional activities

The same professor was the course instructor in year one (study one) and year two (study two). The instructional activities were also the same and took place in five steps (Table 1), lasting for 1 week. First, all students participated in a brief (15-min-long) lecture/

Table 1 The study had five steps and used two instruments for data collection: the same test was used before and after investigation with the VRL and the error survey was used at the end

INSTRUCTIONAL STEPS	INSTRUMENTS	MEASURES	DATA ANALYSIS
1. Lecture on DNA in modern science [~15 min]	None	None	None
2. Pre-Assessment of Knowledge [~15 min]	Pre-Test	Pre-Knowledge Score	
3. Learning by VRL experimentation [~30 min]	None	None	Wilcoxon SRT
4. Post-Assessment of Knowledge [~15 min]	Post-Test	Post-Knowledge Score	
5. Error-Survey [~15 min]	Error Survey	Evaluation Score Reasoning Score	Theme and Frequency

discussion about the current uses of DNA in modern biological and computational applications. The instructor did not detail the specific processes of gel-electrophoresis. Next, students individually completed a short (15-min-long) Pre-Test, probing their knowledge of DNA and gel-electrophoresis concepts and processes. Subsequently, students individually worked with the My-DNA VRL for 30 min. They created repeated experimental trials in order to determine how different variables influence the distance DNA fragments travel. Students used worksheets to record the settings of voltage, concentration, and time of run as well as the distance different-sized DNA fragments traveled. They expressed the distance as a ratio of the actual travel from the starting point and the maximum distance possible. After this learning experience, students completed the Post-Test: the same, 15-min-long, instrument used as the Pre-Test. Finally, students completed an Error-Survey that presented images of anomalous gel outcomes (Caprette 1996) and asked them to evaluate and explain the results.

Design

Because a randomized assignment of students to control and treatment conditions was not possible, a quasi-experimental approach involving pre-post-instruction tests was used that examined learning from a constructivist perspective with a mix of quantitative and qualitative data (Clark Plano and Creswell 2008). The assessment protocol was the same in both studies and it focused on students' learning with the VRL followed by their application of newly constructed knowledge to evaluate and reason about anomalous results.

Data sources and measures

The study employed two instruments: a test of students' knowledge (before and after work with the VRL) and an error-survey of students' analysis of anomalous data (at the end of the work).

The test instrument

The author developed this instrument and used it in prior studies (Toth et al. 2012b). Collaborating scientists and an engineering professor examined this instrument for face validity, and established the content validity. The same test instrument was used before and after instruction (Table 1). Figure 1 illustrates the nature of the test items. The test yielded the *"Knowledge Score"* with three components: (1) knowledge of DNA characteristics (charge, direction of DNA-move in charged medium, differential move of small, medium and large DNA fragments etc.), and (2) knowledge of agarose gel characteristics (porous nature, function as a sieve, size of pores depending on concentration, etc.), as well as (3)
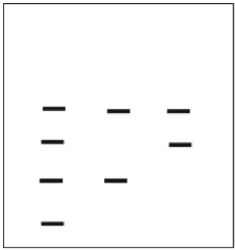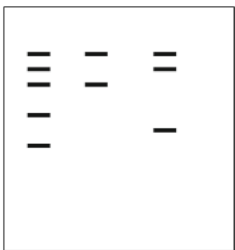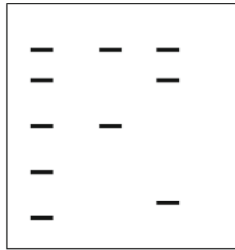
1a. Example "fill-in-the-blank" question that was used to determine students' Conceptual Knowledge Score

a. The 0.6 % agarose gel is _____ (more / less) concentrated than the 2.0 % agarose gel.

b. The gel pores of the 06% agarose gel are _____ (smaller / bigger) than the gel pores of the 2.0% agarose gel.

c. The smaller the pores of a gel, the _____(further/closer) the LARGE DNA fragments will travel down the gel?

Explain your answer : _____

1b. An example multiple-choice question that was used to determine students' Knowledge Application Score

Select the correct ending to formulate a scientifically correct statement:

If we stop a gel electrophoresis process five minutes earlier than the recommended 60-minute-long run,
a. the DNA will not be visible

b. the different DNA bands will be visible but travel less

c. the different DNA bands will be visible but travel more

Explain your answer: _____

1c. You used three gels to evaluate the presence of different sized DNA bands. You ran each gel at 100 V for 60 minutes. The resulting gels are illustrated below. Explain what may have happened to gel B.

Gel A   Gel B   Gel C

Fig. 1 Three example test-items from the pre- and post-tests. Questions such as *1a* and *1b* tested students' knowledge of DNA fragment gel-electrophoresis concepts and processes. Images such as those in *2c* tested students' ability to use their knowledge to evaluate outcome data
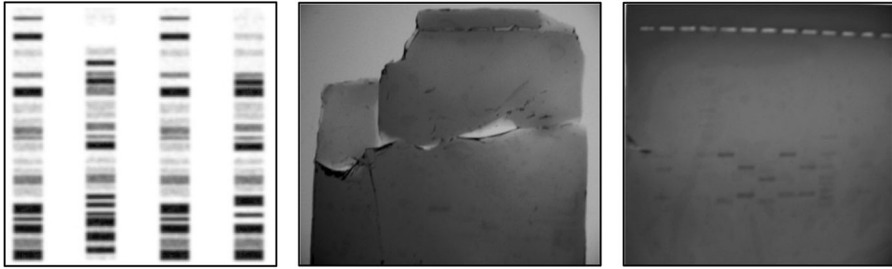
Fig. 2 Three images illustrating gel electrophoresis results, similar to those used on the Error Survey. The *first image* illustrates what an ideal outcome may look like. The *second image* is an example of a broken gel that does not support the reliable assessment. The *third image* is difficult to evaluate due to inadequate visibility. The images that were actually used in the Error-Survey were made available for educational purposes by Rice University and *David R. Caprette (caprette@rice.edu)*. Those images are not shown here and are available at http://www.ruf.rice.edu/~bioslabs/studies/sds-page/sdsgoofs.html

knowledge of the processes of gel-electrophoresis in response to a variety of errors (such as mixed order of loading samples, failure to stop the process in time, incorrect concentration, instrument malfunction and random environmental factors etc.). The 18-item test instrument employed true/false, multiple choice and fill-in-the-blank questions with correct answers worth one point each. Incorrect answers were scored zero. Cronbach's alpha indicated that the 18 items of the test instrument were internally consistent (0.83 and 0.75 in the first and second study, respectively).

The error survey

The author developed this instrument for the purpose of this research. It was the last step of assessment (Table 1) to probe students' ability to analyze anomalous data outcomes from HOLs. The content validity of this instrument was established by the collaborating engineering professor and the data coding assistant, who had master's level training in biology (forensic science). Survey items were edited until full agreement by the content experts was obtained. This instrument used images from the "Hall of Shame" website (Caprette 1996) and included simplified images similar to those the students saw when using the VRL (Fig. 2).

All images on the survey indicated anomalous outcomes so the question under each image asked students to (a) detect and describe (evaluate) the nature of anomaly and to (b) reason about how these aberrations may have come about. For each of six images on the survey, students were asked, "What makes the DNA band outcome hard to interpret from this gel?" and "What do you think may have caused this anomalous outcome?" Students responded to the questions with open-ended text that was either typed or handwritten. These responses were coded and quantified to yield the *Evaluation Score,* and the *Reasoning Score*.

Data coding and analysis

The first analysis examined quantitative data on students' learning of gel electrophoresis concepts and processes: the Knowledge Score from the Pre- and Post-Tests. This analysis used a Wilcoxon signed rank test (WSRT), a non-parametric equivalent of the paired-t test,

designed for studies with small participant numbers. The analysis established the Wilcoxon Z score separately for study one and study two by using the software SPSS (2012). It calculated the effect size by dividing the Z-score of each WSRT with the square root of the participant number, as suggested by Field (2005). This analysis answered the first research question about participants' knowledge construction.

The second analysis coded, quantified, and analyzed the qualitative data from students' open-ended answers on the six questions of the Error Survey. The coding and analysis centered on the interpretation of the meaning of students' statements. To establish inter-rater reliability the researchers discussed the coding and made modifications until reaching 100 % agreement. Based on this agreed-upon meaning, the researchers assigned codes to each evaluation and reasoning statement, then quantitatively described the characteristics of coded data. This method for the quantification and analysis of qualitative data follows protocols described by Chi (1997), Miles and Huberman (1994), and Miles et al. (2013).

To arrive at the *evaluation score*, an evaluation statement that mentioned the focal measure, the distance DNA fragments traveled, at least once, received a score of one. Evaluative statements that did not mention DNA bands yielded a score of zero for each image, thus the maximum score for evaluation on each image was one. *Reasoning scores* were quantified with a similar method. Statements about anomalous outcomes that included reference to the variables studied with the VRL (concentration, voltage, time of run) received a score of one. No mention of these variables resulted in the score of zero for each image; thus the maximum reasoning score on each image was one. With this scoring method, the maximum evaluation score and the maximum reasoning score were six and the minimum scores were zero. Cronbach's alpha indicated that the reliability of the evaluation score was 0.60 and 0.64 in study one and study two respectively. The reliability of the reasoning score was 0.45 and 0.64 in study one and study two respectively. The reliability scores of 0.6 and above are in the acceptable range for social science research (Field 2005). However, reliability is associated with the number of items in a survey, thus the moderate to low reliability of the six-item reasoning scores should be considered in this context. The analysis of the evaluation and reasoning scores answered research questions two and three, extending the findings of a prior publication (Toth et al. 2012a) with a deeper insight on how students' use their learning with the VRL to the analyze anomalous data from an HOL.

Results

Does investigation with a VRL assist engineering students in constructing new knowledge?

The analysis of the students' Knowledge Score with a WSRT revealed significant increase in students' knowledge of gel-electrophoresis concepts and processes after work with the VRL. The effect sizes (r) were 0.85 (in study one) and 0.63 (in study two) as illustrated in Table 2. Leech et al. (2008) suggest that these effect sizes are larger than typical.

In what ways do students use their learning with a VRL to evaluate anomalous data from an HOL?

The analysis found that the majority of evaluations focused on the same outcome measure as the VRL-work: the movement of DNA bands. The top data row in Table 3 documents

Table 2 Students' mean knowledge scores of gel-electrophoresis concepts and processes before and after work with a virtual laboratory indicates significant increases in both studies

	Time	Study one (N = 16)				Study Two (N = 15)			
		Mean	Z	r	p	Mean	Z	r	p
Knowledge score [max = 18]	PRE	9.44	3.4	0.85	0.001 **	11.6	2.44	0.63	0.01*
	POST	16.19				14.5			

* Indicates significance at the confidence level <0.05 and ** indicates significance at the confidence level <0.01

Table 3 The large majority of students' evaluations of anomalous data focused on DNA-bands, but only half of the students' reasons focused on the variables of the VRL

	Percent of answers in study one (N = 16) (%)	Percent of answers in study two (N = 15) (%)
Evaluations focused on DNA-bands	81	87
Reasoning with variables studied	60	53

that 81 % of all answers in the first study and 87 % in the second study focused on DNA bands. The mean evaluation score was M = 4.75 (SD = 1.34) and M = 4.73 (SD = 1.39) in study one and study two respectively, indicating that students were not always consistent in using DNA-bands in answering each question.

The analysis found that, students noted two important concepts that connected their HOL-data evaluation to their learning with the VRL. First, they referred to the separation of different DNA bands using statements such as "The bands aren't well defined, they run together, difficult to identify" and "The bands seem to have shadows, it is hard to tell where the bands should be." Second, students' evaluative statements also demonstrated focus on DNA-band pattern across the gel, as seen while working with the VRL. Sample statements showing this focus included "None of the bands line up," "The bands are not perfectly horizontal," and "the bands are not in line." Statements without focus on DNA bands also appeared in evaluations such as "I am not sure what this means" and "It looks like a pirate ship."

In what ways do students use the variables they examined with a VRL to reason about anomalous data from an HOL?

Table 3 indicates the frequency of those statements that correctly referred to the variables of the VRL: concentration, voltage, and the time of run. Only 60 % of rationales in study one and 53 % of rationales in study two used these variables. The reasoning scores were M = 3.63 (SD = 1.36) and M = 4.73 (SD = 1.44) in study one and study two respectively.

Two themes of variable-based reasoning emerged from students' written rationales: simple reference to the variables studied with the VRL, and reference to variables with their mode of action, as learned from the VRL. For example, reasoning that simply referenced the variables of the VRL included "The voltage was too high" or "The concentration was too low." "Reasoning statements that provided a mode-of-effect included "The high voltage heated the equipment," "The high voltage warped the gel," "Too low concentration and longer fragments did not separate," and "The gel is not concentrated enough

for good separation." However, rationales also included statements with focus on experimental error such as "the gels were damaged," "Broken wells/no wells," "Not proper procedures were followed", "Somebody messed up," and "the gel was polluted."

Discussion of results

Does investigation with a VRL assist engineering students in constructing new knowledge?

An analysis that is powerful for nonparametric data and small sample sizes indicated a significant change in students' knowledge (Table 2). Students did learn the characteristics of the gel-medium and the DNA molecule as well as the roles of concentration, voltage, and the time of run from their work with a simplified, virtual environment. These results helped eliminate the lack of knowledge as an obstacle of transfer (Salomon and Perkins 1989). The results also corroborated prior evidence on the benefits of learning with VRLs in a variety of contexts (Apkan and Strayer 2010; Triona and Klahr 2003; Zacharia 2007). Furthermore, this finding confirmed the result of a prior study from a university science classroom (Toth et al. 2012b) and suggested that the instructional approach can be replicated in engineering classrooms, as well. Having documented students' basic conceptual understanding, the next focus was on whether students used this knowledge to evaluate and reason about anomalous data.

In what ways do students use their learning with a VRL to evaluate anomalous data from an HOL?

This question examined one critical indicator of making a conceptual connection between the two learning environments, the use of the same outcome measure. The majority of students' evaluation statements focused on the characteristics of DNA bands, the same outcome measure that was at the center of students' learning with the VRL (Table 3). This finding suggests that students were successful in making a conceptual "leap" from learning with the VRL to analyzing anomalous data—a skill students did not study with the VRL. Mentioning the *separation* of different length DNA bands, as well as the *pattern* of small, medium and long bands across the gel support this conclusion. These results suggest that, the perceptually apparent, surface-features of this particular VRL provided enough support for students to make a connection between experimenting with the VRL and analyzing anomalous data from the real world (Jonassen 2006). Students' mean evaluation scores as well as their written explanations point to their successes in evaluating anomalous data by applying what they learned with the VRL. This result is in contrast to prior findings that higher education students conceptually separate virtual investigations and real-world results (Nedic et al. 2003). In our case, students were able to make such connection between the simplified variable manipulations permitted by the VRLs and the evaluation of real-world data, without the design contradictions mentioned in the prior literature (Reiser 2004).

In what ways do students use the variables they examined with the VRL to reason about anomalous data from an HOL?

In contrast to the positive findings on students' ability to transfer their learning with the VRL to evaluate anomalous data, the analysis of students' reasoning pointed to the

cognitive complexity of the task (Chinn and Brewer 2001). The disconnect between work with the VRL and real-world, HOL-data was apparent in three aspects of students' reasoning: the low frequency of referring to changes in variable levels, the limited explanation of the mode-of-effect of variables, and students' reference to general experimental error instead of explaining the role of variables in the unanticipated outcome.

The frequency of referring to changes in variables was 60 and 53 % in the two studies respectively (Table 3). Given that students did have accurate content knowledge and that they were able to refer to the variables studied with the VRL during their data evaluations, this frequency of reasoning with variables is low. Furthermore, students in at least one of the studies were quite inconsistent in their reasoning with the variables. Both the low mean reasoning score and the low reliability of the reasoning score suggest the need to continue examining the difficulties students experience in reasoning about anomalous data. For example, the limited examination of the mode-of-effect of variables suggests at least some level of disconnect between learning about variables with repeated experimental designs in the VRL and applying this knowledge to the analysis of anomalous data from an HOL. Therefore, instructional design additions that draw students' attention to the implicit and explicit similarities between this particular VRL and the real-world HOL processes could be fruitful in the future. Such explanation prompts were successful in prior studies (Sandoval and Millwood 2005; Sandoval and Reiser 2004) and a recent literature review published in this journal also point to other effective guidance methods, albeit with attention to some of the shortcomings of each method (Zacharia et al. 2015).

Students' struggles in using the mode-of-effect of variables may have also contributed to their simple reference to human error or instrument malfunction, instead of focusing on the specific variables that contributed to the outcome. Therefore, in order to support students' investigations in environments that blend VRLs and HOLs, the description and recognition of error types as related to working with anomalous data needs further investigation. However, the topic alone requires further study, since the definition of error varies widely by context and discipline (Allchin 2001). In the context of DNA gel electrophoresis for example, the results have personal consequences for human subjects. Therefore, a keen attention to overcoming experimental errors is not unusual and is even commendable. Nevertheless, the results are aligned with prior data on students' use of alternative/general causal factors when they cannot offer specific reasons for why the outcome may deviate from expected results (Chinn and Brewer 2001). These results suggest the need for the more detailed examination of combining VRLs and HOL in different contexts with respect to the sources of anomalous data.

Practical implications

The results provide evidence for practitioners and instructional designers on the benefits of learning with VRLs by testing the effects of different variables with repeated experimental trials. However, there continues to be a need for instructional design decisions that minimize the disadvantage of VRLs that do not yield anomalous data. While working with such VRLs, students may require assistance in making the necessary conceptual leap between the simplified results of a VRL and the analysis of real-world data. For example, future instructional interventions may use explanation prompts, found to be beneficial by prior research (Sandoval and Millwood 2005; Sandoval and Reiser 2004). Alternatively, practitioners can employ discussion and debate to help students internalize these critical

processes of scientific and engineering practice (Duschl 2008; Toth et al. 2002) while helping to reduce cognitive complexity (Sweller et al. 1998; Schnotz and Rasch 2005). In this study, there were no starter prompts in order to document the "nascent" thinking processes of these higher education students.

Limitations and further research directions

The limitations of this work are largely the artifacts of educational research in classroom settings. Research in formal classrooms is limited in opportunities for the development of randomized trials, the use of control groups, the time available for assessment and sometimes even in the number of participants available for a study. Randomized trials are especially challenging in university settings where students often self-select for entry to courses. Furthermore, university instructors are concerned about ethical issues associated with teaching different course-sections with different pedagogies. Instructional time is at premium in these classes and the number of students is usually low. This limitation may have contributed to the limited generalizability of the findings and to the low reliability of the reasoning score in study one. Furthermore, in higher education settings, students may learn concepts by responding to the questions on the pre-test. Due to the above reasons, the results of this, and other quasi-experimental studies, are best to (a) complement existing studies, and to (b) provide suggestions for further, controlled exploration (Shadish et al. 2002).

Despite the developmental nature of the work, the results draw practitioners' attention to critical components of students' difficulties in working with anomalous data. Future studies with larger participant numbers may explore the specific learning patterns of students with different prior experiences (i.e. graduate or undergraduate students). Future qualitative studies may employ "think-aloud" approaches and document the details of students' interpretations of anomalous images. Continued studies on students' work with anomalous data in complex, socially significant settings are also needed to examine the cognitive, behavioral, and attitudinal factors of preparing students to handle anomalous data. These studies may use anomalous data as a motivator before VRL work or could focus on how students evaluate their own anomalous data results from an HOL and examine whether students learn to recognize sources of error when they apply the procedures they learned with the VRL during hands-on work. Much needed research should assist practitioners in helping students conceptualize aspects of anomalous data that are random aberrations in the results of repeated experiments, results that are due to experimental error such as human action in the design, setup, and execution of experiments or instrument malfunction or random change in the investigation environment. The current literature indicates that the definition of error is context and discipline dependent (Allchin 2001) and several conceptualizations of anomalous data continue to exist today including outliers, aberrations, error, or noise (Chandola et al. 2009; Gong et al. 2012; Han et al. 2006). With further clarifications of key constructs and research-evidence on students' learning, scholars may contribute to the development of evidence-based pedagogical approaches and consequently to reducing the frequency of misconduct and malpractice due to incorrect data handling (Dorner 1996; Mason 2001; Spector and Davidsen 2000).

References

Ainsworth, S. (2008). How do animations influence learning? In D. Robinson & G. Schraw (Eds.), *Current innovations in educational technology that facilitate student learning* (pp. 37–67). Charlotte, NC: Information Age Publishing.

Allchin, D. (2001). Error types. *Perspectives on Science, 9*(1), 37–58.

Apkan, J., & Strayer, J. (2010). Which comes first the use of computer simulation of frog dissection or conventional dissection as academic exercise? *Journal of Computers in Mathematics and Science Teaching, 29*(2), 113–138.

Balamuralithara, B., & Woods, P. C. (2009). Virtual laboratories in engineering education: The simulation lab and remote lab. *Computer Applications in Engineering Education, 17*(1), 108–118.

Caprette, D. (1996). Rice University STS-Page Hall of Shame. Available at http://www.ruf.rice.edu/~bioslabs/studies/sds-page/sdsgoofs.html. Retrieved April 29, 2015.

Carey, S., & Smith, C. (1993). On understanding the nature of scientific knowledge. *Educational Psychologist, 28*(3), 235–251.

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys, 41*(3), 15.1–15.58.

Chen, S. (2010). The view of scientific inquiry conveyed by simulation-based virtual laboratories. *Computers and Education, 55*(3), 1123–1130.

Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of the Learning Sciences, 6*(3), 271–315.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research, 63*, 1–49.

Chinn, C. A., & Brewer, W. F. (1998). An empirical test of taxonomy of responses to anomalous data in science. *Journal of Research in Science Teaching, 35*(6), 623–654.

Chinn, C. A., & Brewer, W. F. (2001). Models of data: A theory of how people evaluate data. *Cognition and Instruction, 19*(3), 323–393.

Clark Plano, V. L., & Creswell, J. W. (2008). *Mixed methods reader*. Thousand Oaks: Sage.

Cobb, S., Heaney, R., Corcoran, O., & Henderson-Begg, S. (2009). The learning gains and student perceptions of a second life virtual lab. *Bioscience Education*. doi:10.3108/beej.13.5.

Corter, J. E., Esche, S. K., Chassapis, C., Ma, J., & Nickerson, J. V. (2011). Process and learning outcomes from remotely-operated, simulated, and hands-on student laboratories. *Computers and Education, 57*(3), 2054–2067.

de Jong, T., Linn, M. C., & Zacharia, Z. C. (2013). Physical and virtual laboratories in science and engineering education. *Science, 340*(6130), 305–308.

De Jong, T., & Van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research, 68*(2), 179–201.

Dede, C., Brown-L'Bahy, T., Ketelhut, D., & Whitehouse, P. (2004). Distance learning (virtual learning). *The Internet Encyclopedia*. doi:10.1002/047148296X.tie047.

Dede, C., Salzman, M., Loftin, R. B., & Ash, K. (1997). Using virtual reality technology to convey abstract scientific concepts. *Learning the sciences of the 21st century: Research, design, and implementing advanced technology learning environments*. Hillsdale, NJ: Lawrence Erlbaum.

Dekker, S. (2006). *The field guide to understanding human error*. Burlington: Ashgate Publishing.

Denzin, N. K., & Lincoln, Y. S. (2005). *Handbook of qualitative research* (2nd ed.). Sage: Thousand Oaks.

Dickey, M. D. (2011). The pragmatics of virtual worlds for K-12 educators: Investigating the affordances and constraints of Active Worlds and Second Life with K-12 in-service teachers. *Educational Technology Research and Development, 59*(1), 1–20.

Dori, Y. J., & Barak, M. (2003). A web-based chemistry course as a means to foster freshmen learning. *Journal of Chemical Education, 80*(9), 1084–1092.

Dorner, D. (1996). *The logic of failure. Recognizing and avoiding error in complex situations*. New York: Metropolitan Books.

Drenth, P. J. D. (2006). Responsible conduct in research. *Science and Engineering Ethics, 12*(1), 13–21.

Dreyfus, A., Jungwirth, E., & Eliovitch, R. (1990). Applying the "cognitive conflict" strategy for conceptual change. Some implications, difficulties and problems. *Science Education, 74*, 555–569.

Driver, R. (1989). Students' conceptions and the learning of science. *International Journal of Science Education, 11*(5), 481–490.

Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of research in education, 32*(1), 268–291.

Field, A. (2005). *Discovering statistics: Using SPSS* (2nd ed.). Sage: Singapore.

Genetic Science Learning Center (GLSC). (2013). *Virtual labs*. Available at http://learn.genetics.utah.edu/. Retrieved April 29, 2015.

Gong, M., Zhang, J., Ma, J., & Jiao, L. (2012). An efficient negative selection algorithm with further training for anomaly detection. *Knowledge-Based Systems, 30*, 185–191.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. *Handbook of Qualitative Research, 2*, 163–194.

Han, J., Kamber, M., & Pei, J. (2006). *Data mining: Concepts and techniques*. San Francisco, CA: Elsevier.

Howard Hughes Medical Institute, HHMI (n.d.). BioInteractive: Free resources for science teachers and students. Available at http://www.hhmi.org/biointeractive. Retrieved April 29, 2015.

Jackson, S. L., Stratford, S. J., Krajcik, J. S., and Soloway, E. (1996). *Model-It: A case study of learner-centered software design for supporting model building*. ERIC Clearinghouse.

Jonassen, D. H. (2006). *Modeling with technology: Mindtools for conceptual change*. Pearson: Upper Saddle River, NJ.

Kali, Y. (2008). The design principles database as means for promoting design-based research. In A. E. Kelly, et al. (Eds.), *Handbook of design research methods in education: Innovations in science, technology, engineering, and mathematics learning and teaching* (pp. 423–438). New York: Rutledge.

Kirschner, P., & Wopereis, I. G. (2003). Mindtools for teacher communities: A European perspective. *Technology, Pedagogy and Education, 12*(1), 105–124.

Kozma, R. (2003). The material features of multiple representations and their cognitive and social affordances for science understanding. *Learning and Instruction, 13*(2), 205–226.

Leech, N., Barrett, K., & Morgan, G. (2008). *SPSS for intermediate statistics: Use and interpretation* (3rd ed.). New York: Psychology Press.

Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: a critical appraisal. *Learning and Instruction, 11*(4), 357–380.

Limon, M., & Carretero, M. (1997). Conceptual change and anomalous data: A case study in the domain of natural sciences. *European Journal of Psychology of Education, 13*(2), 213–230.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry* (Vol. 75). Sage.

Ma, J., & Nickerson, J. V. (2006). Hands-on, simulated, and remote laboratories: A comparative literature review. *ACM Computing Surveys, 38*(3), 7. doi:10.1145/1132960.1132961.

Magin, D., & Kanapathipillai, S. (2000). Engineering students' understanding of the role of experimentation. *European Journal of Engineering Education, 25*(4), 351–358.

Martinson, B. C., Anderson, M. S., & De Vries, R. (2005). Scientists behaving badly. *Nature, 435*(7043), 737–738.

Masnick, A. M., & Klahr, D. (2003). Error matters: An initial exploration of elementary school children's understanding of experimental error. *Journal of Cognition and Development, 4*(1), 67–98.

Mason, L. (2001). Responses to anomalous data on controversial topics and theory change. *Learning and Instruction, 11*, 453–483.

Miles, M. B., & Huberman, M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Sage: London.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2013). *Qualitative data analysis: A methods sourcebook*. SAGE Publications, Incorporated.

MyDNA. (2003). *Discovery module 2: Sorting DNA molecules*. Available at http://mydna.biochem.umass.edu/modules/sort.html. Retrieved April 29, 2015.

Nakhleh, M. B., & Krajcik, J. S. (1993). A protocol analysis of the influence of technology on students' actions, verbal commentary, and thought processes during the performance of acid-base titrations. *Journal of Research in Science Teaching, 30*(9), 1149–1168.

National Aeronautics and Space Administration (NASA). (n.d.). *The virtual lab educational software*. Available at http://www.nasa.gov/offices/education/centers/kennedy/technology/Virtual_Lab.html Retrieved April 29, 2015.

Nedic, Z., Machotka, J., and Nafalski, A. (2003). Remote laboratories versus virtual and real laboratories. *Frontiers in Education, 2003*. Vol. 1, pp. T3E–1.

Next Generation Science Standards (NGSS). (2012). Available at http://www.nextgenscience.org/next-generation-science-standards. Rretrieved April 29, 2015.

Nokes-Malach, T. J., & Mestre, J. P. (2013). Toward a model of transfer as sense-making. *Educational Psychologist, 48*(3), 184–207.

Olympiou, G., & Zacharia, Z. C. (2012). Blending physical and virtual manipulatives: An effort to improve students' conceptual understanding through science laboratory experimentation. *Science Education, 96*(1), 21–47.

Olympiou, G., Zacharias, Z., & de Jong, T. (2013). Making the invisible visible: Enhancing students' conceptual understanding by introducing representations of abstract objects in a simulation. *Instructional Science, 41*(3), 575–596.

Piaget, J. (1985). *The equilibration of cognitive structures: The central problem of intellectual development*. Chicago: University of Chicago Press.

Posner, G. J., Strike, K. A., Hewson, P. V., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66*, 211–227.

Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., & Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *The Journal of the Learning Sciences, 13*(3), 337–386.

Rebello, N. S., Zollman, D. A., Allbaugh, A. R., Engelhardt, P. V., Gray, K. E., Hrepic, Z., & Itza-Ortiz, S. F. (2005). Dynamic transfer: A perspective from physics education research. In J. P. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective*. Greenwich: IAP.

Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *The Journal of the Learning Sciences, 13*(3), 273–304.

Renken, M. D., & Nunez, N. (2013). Computer simulations and clear observations do not guarantee conceptual understanding. *Learning and Instruction, 23*, 10–23.

Rieber, L. P. (1992). Computer-based microworlds: A bridge between constructivism and direct instruction. *Educational Technology Research and Development, 40*(1), 93–106.

Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanism of a neglected phenomenon. *Educational psychologist, 24*(2), 113–142.

Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction, 23*(1), 23–55.

Sandoval, W. A., & Reiser, B. J. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education, 88*(3), 345–372.

Schnotz, W., & Rasch, T. (2005). Enabling, facilitating, and inhibiting effects of animations in multimedia learning: Why reduction of cognitive load can have negative results on learning. *Educational Technology Research and Development, 53*(3), 47–58.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Smith, C. L., Maclin, D., Houghton, C., & Hennessey, M. G. (2000). Sixth-grade students' epistemologies of science: The impact of school science experiences on epistemological development. *Cognition and Instruction, 18*(3), 349–422.

Spector, J. M. (2008). Cognition and learning in the digital age: Promising research and practice. *Computers in Human Behavior, 24*(2), 249–262.

Spector, J. M., & Davidsen, P. I. (2000). Designing technology enhanced learning environments. In B. Abbey (Ed.), *Instructional and cognitive impacts of Web-based education* (pp. 241–261). Hershey, PA: Idea Group Publishing.

Spector, J. M., Lockee, B. B., Smaldino, S., & Herring, M. (Eds.). (2013). *Learning, problem solving, and mind tools: Essays in honor of David H. Jonassen*. Upper Saddle River, NJ: Routledge.

SPSS Statistics for Windows. Version 21(2012) IBM Corporation: Armonk, NY.

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296.

Toth, E. E. (2009). "Virtual Inquiry" in the science classroom: What is the role of technological pedagogical content knowledge? *International Journal of Communication Technology in Education*, 5(4), 78–87.

Toth, E. E., Brem, S. K., & Erdos, G. (2009). "Virtual inquiry": Teaching molecular aspects of evolutionary biology through computer-based inquiry. *Evolution: Education and Outreach*, 2(4), 679–687.

Toth, E. E., Dinu, C. Z., McJilton, L., & Moul, J. (2012a). Interdisciplinary collaboration for educational innovation: Integrating inquiry learning with a virtual laboratory to an engineering course on "cellular machinery". In T. Amiel & B. Wilson (Eds.), *Proceedings of EdMedia: World conference on educational media and technology 2012* (pp. 1949–1956). Association for the Advancement of Computing in Education.

Toth, E. E., Morrow, L., & Ludvico, L. R. (2012b). Blended inquiry with hands-on and virtual laboratories: The role of perceptual features during knowledge construction. *Interactive Learning Environments*. doi:10.1080/10494820.2012.693102.

Toth, E. E., Suthers, D. D., & Lesgold, A. (2002). Mapping to know: The effects of representational guidance and reflective assessment on scientific inquiry skills. *Science Education*, *86*(2), 264–286.

Triona, L. M., & Klahr, D. (2003). Point and click or grab and heft: Comparing the influence of physical and virtual instructional materials on elementary school students' ability to design experiments. *Cognition and Instruction, 21*(2), 149–173.

Underwood, J. S., Hoadley, C., Lee, H. S., Hollebrands, K., DiGiano, C., & Renninger, K. A. (2005). IDEA: Identifying design principles in educational applets. *Educational Technology Research and Development, 53*(2), 99–112.

UsNews and World Report (2007). Dennis Quaid's Newborns Are Given Accidental Overdose: Medical Mistakes Are Not Uncommon in U.S. Hospitals. Nov 1, 2007.

Windschitl, M. (2000). Supporting the development of science inquiry skills with special classes of software. *Educational Technology Research and Development, 48*(2), 81–95.

Wolf, T. (2010). Assessing student learning in a virtual laboratory environment. *Education, IEEE Transactions, 53*(2), 216–222.

Zacharia, Z. C. (2007). Comparing and combining real and virtual experimentation: an effort to enhance students' conceptual understanding of electric circuits. *Journal of Computer Assisted learning, 23*(2), 120–132.

Zacharia, Z. C., & Constantinou, C. P. (2008). Comparing the influence of physical and virtual manipulatives in the context of the physics by inquiry curriculum: The case of undergraduate students' conceptual understanding of heat and temperature. *American Journal of Physics, 76*(4), 425–430.

Zacharia, Z. C., Manoli, C., Xenofontos, N., de Jong, T., Pedaste, M., van Riesen, S. A. N., et al. (2015). Identifying potential types of guidance for supporting student inquiry when using virtual and remote labs in science: a literature review. *Educational Technology Research and Development, 63*(2), 257–302.

Zacharia, Z. C., Olympiou, G., & Papaevripidou, M. (2008). Effects of experimenting with physical and virtual manipulatives on students' conceptual understanding in heat and temperature. *Journal of Research in Science Teaching, 45*(9), 1021–1035.

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review, 27*, 172–223.

Eva Erdosne Toth received her biology and chemistry teaching certificate from the Eotvos Lorand University, Budapest, Hungary and her M. Ed. and Ph. D. degrees from the University of Illinois at Urbana-Champaign. Her research focuses on the applications of technology for science teaching and learning, and on the social implications of scientific and engineering practices.