International Journal of Science Education

ISSN: 0950-0693 (Print) 1464-5289 (Online) Journal homepage: http://www.tandfonline.com/loi/tsed20

Learning science content through socio-scientific issues-based instruction: a multi-level assessment study

Troy D. Sadler, William L. Romine & Mustafa Sami Topçu

INTERNATIONAL JOURNAL OF SCIENCE EDUCATION, 2016
http://dx.doi.org/10.1080/09500693.2016.1204481

Learning science content through socio-scientific issues-based instruction: a multi-level assessment study

Troy D. Sadler[a], William L. Romine[b] and Mustafa Sami Topçu[c]

[a]ReSTEM Institute, University of Missouri-Columbia, Columbia, MO, USA; [b]Department of Biology, Wright State University, Dayton, OH, USA; [c]Elementary Science Education Department, Faculty of Education, Yıldız Technical University, Istanbul, Turkey

ABSTRACT
Science educators have presented numerous conceptual and theoretical arguments in favor of teaching science through the exploration of socio-scientific issues (SSI). However, the empirical knowledge base regarding the extent to which SSI-based instruction supports student learning of science content is limited both in terms of the number of studies that have been conducted in this area and the quality of research. This research sought to answer two questions: (1) To what extent does SSI-based instruction support student learning of science content? and (2) How do assessments at variable distances from the curriculum reveal patterns of learning associated with SSI-based instruction? Sixty-nine secondary students taught by three teachers participated in the study. Three teachers implemented an SSI intervention focused on the use of biotechnology for identifying and treating sexually transmitted diseases. We found that students demonstrated statistically and practically significant gains in content knowledge as measured by both proximal and distal assessments. These findings support the claim that SSI-based teaching can foster content learning and improved performance on high-stakes tests.

ARTICLE HISTORY
Received 2 January 2015
Accepted 17 June 2016

KEYWORDS
Curriculum; biotechnology; socio-scientific issues; learning; multi-level assessment

Socio-scientific issues (SSI) are complex societal issues with conceptual, procedural, and/or technological associations with science. These challenging issues are likely to be confronted within peoples' daily lives (Kolstø, 2001) and provide students with a context to support their active examination of the relationships among science, their lives in society, and technology (Driver, Newton, & Osborne, 2000; Kolstø, 2006; Zeidler, Sadler, Applebaum, & Callahan, 2009). In recent years, science education researchers have recognized the significance of SSI as a means of engaging students in inquiries related to science as well as their own lived experiences (Topcu, Sadler, & Yilmaz-Tuzun, 2010; Zeidler, 2003). Development of learners' abilities to understand, negotiate, and ultimately make decisions regarding SSI is critical to scientific literacy (see Roberts, 2007). The importance of integrating student practices relative to the negotiation,

CONTACT Troy D. Sadler sadlert@missouri.edu University of Missouri-Columbia, 321C Townsend Hall, 65211 Columbia, MO, USA

discussion, and analysis of complex SSI is recognized internationally (American Association for the Advancement of Science [AAAS], 1990; Australian Curriculum, Assessment and Reporting Authority [ACARA], 2015; Ministry of National Education of Turkey, 2013; National Research Council, 1996).

Multiple models for SSI-based instruction have been developed (e.g. Eilks, 2010; Presley et al., 2013; Saunders & Rennie, 2013), but they share common features: (1) use of a complex, socially relevant issue as a central theme; (2) engagement of learners in higher order thinking processes, (3) and explicit attention to the science and social dimensions of the issue. We have argued elsewhere that SSI-based instruction needs to provide learning contexts with opportunities for students to explore conceptual connections to science as well as engage in discussion, critical thinking and decision-making (Klosterman & Sadler, 2010). This instruction focuses on evidence-based reasoning and attention to problems that require negotiation, engagement, and analysis of multiple perspectives. Through SSI, students are challenged to consider scientific principles underlying the issues and analyze scientific data that can inform the negotiation of issues (Zeidler et al., 2009). The scientific knowledge that develops as a result of discussion and negotiation of SSI becomes personally relevant and socially shared. Advocates of SSI-based instruction have argued that the approach can support several desired student learning outcomes including students' development of interest and motivation in science (Albe, 2008; Bulte, Westbroek, De Jong, & Pilot, 2006; Dori, Tal, & Tsaushu, 2003; Harris & Ratcliffe, 2005; Parchmann et al., 2006; Romine & Sadler, 2014), understandings of nature of science (NOS) (Eastwood et al., 2012; Khishfe & Lederman, 2006; Walker & Zeidler, 2007), and reasoning skills (Lee & Erdogan, 2007; Pedretti, 1999; Tal & Hochberg, 2003; Tal & Kedmi, 2006; Yager, Lim, & Yager, 2006; Zeidler et al., 2009; Zohar & Nemet, 2002; Romine, Sadler, & Kinslow, 2016).

Although interest, motivation, NOS, and critical thinking are important components of scientific literacy and their development is crucial to science education, measures of these variables are rarely assessed in state, national, or international examinations (Orpwood, 2001). Instead, these examinations place emphasis on content knowledge. Many educators argue (or at least implicitly assume) that developing increasingly sophisticated understandings of scientific content is the primary goal of science education (see Roberts, 2007). Given the prioritization and representation of basic science concepts in national standards and standardized assessments, this view is reified in the field's most important policy documents (Klosterman & Sadler, 2010). While researchers have offered ample conceptual and theoretical justifications for the inclusion of SSI as central elements of science teaching (Sadler, 2011a; Zeidler, 2003), actual implementation of SSI approaches has been relatively limited (Sadler, 2011b). At least part of the challenge in widespread dissemination of SSI is an assumption on the part of some teachers, administrators, and policy-makers (Hughes, 2000; Pouliot, 2008) that SSI-based teaching dilutes student exposure to basic science ideas and principles and may inhibit student learning of the kinds of science knowledge that are most valued in today's educational system (Orpwood, 2001). Therefore, in order to advance the field of SSI practice and research, empirical investigations that test the hypothesis that SSI-based teaching supports science content learning are necessary. *In this study, we explore the efficacy of SSI-based instruction for supporting student learning of content knowledge related to molecular biology and genetics at the high school level.*

Review of literature

Learning content in the context of SSI

Some research points to the effectiveness of SSI-based instruction in supporting science content learning. Venville and Dawson (2010) studied student learning of genetics in the context of an instructional unit related to genetic technologies. Students participating in the SSI intervention demonstrated significantly better learning than a comparison group, which studied genetics through a more traditional approach. The authors characterized the differences observed as modest but significant. This study supports the contention that SSI-based instruction fosters learning of science content, but the necessarily limited scope (the study focuses on one unit) leaves the question of student learning in the context of SSI only partially answered. Other researchers have also addressed the question, but limitations in the design of the studies and instrumentation have led to some criticism. For example, some research in this area relies on post-test designs that cannot account for variance in prior knowledge (e.g. Bulte et al., 2006; Zohar & Nemet, 2002). Zohar and Nemet (2002) investigated the effects of a 12-week SSI intervention involving four 9th grade classes from two Israeli schools. The intervention focused on the explicit incorporation of argumentation and related the instruction to genetic engineering, applied human genetics, and social issues associated with these topics. A comparison group received conventional instruction and followed a traditional textbook approach with no special attention paid to processes of argumentation. Since the main focus of the study was improving students' argumentation skills, students' content knowledge was not explored prior to the SSI instruction. Therefore, the effect of this SSI instruction on students' learning of science content was not as clear as the investigation of argumentation skills. Similarly, Bulte et al. (2006) used a chemistry curriculum about water quality to investigate the potential impact of a context-based curriculum. They reported that 70% of the students answered questions correctly following completion of the curriculum, but they did not report on students' knowledge prior to the intervention. Most recently, Dawson and Venville (2013) investigated the effect of an argumentation-focused SSI unit on students' argumentation and informal reasoning capacities. Using a quasi-experimental approach, they found that improvement in argumentation and informal reasoning was significantly greater in students who engaged in argumentation activities around SSI than students who did not.

Another group of studies used pre-post designs to investigate content knowledge gains following the implementation of SSI-based instruction (Barab, Sadler, Heiselt, Hickey, & Zuiker, 2007; Dori et al., 2003; Klosterman & Sadler, 2010). Dori et al. (2003) investigated high school students' learning associates with implementation of SSI-based instruction including biotechnology, the environment, and related issues. The pre- and post-test results showed that students improved their understanding of science content across the instruction. Barab et al. (2007) and Sadler, Klosterman, and Topcu (2011) also studied student learning in the context of SSI-based units and documented learning gains.

An important critique of the extant literature related to the question of how effective SSI-based instruction is for supporting student learning of science content relates to the validity of the measures employed. For example, Yager et al. (2006) explored student learning in the context of a local environmental issue. Comparisons of student performance between an issue-based class and a more traditional class were made based on assessments of student learning derived from teacher-generated tests. The items were well aligned with the instructional contexts, which certainly make sense from a pedagogical perspective, but such alignment presents difficulty for the interpretation of empirical results; in fact, the What Works Clearinghouse (Institute of Educational Sciences, 2014) cautions researchers about the negative impact of overalignment on the generalizability and usefulness of clinical findings. Zeidler, Applebaum, and Sadler (2011) provide another example in which student learning was assessed through an instrument developed for classroom purposes. In this case, student's content learning was not the primary research focus, but the study has been cited as evidence for learning outcomes generated by SSI-based instruction.

Multi-level assessment

In the investigation of the effects of focused teaching interventions on content learning, we face the need to obtain measures that are both sensitive to change and generalizable. This is difficult to do with one assessment. Assessments that are closely aligned to a particular curriculum are subject to critique for limited generalizability and comparability. On the other hand, assessments that are distanced further from a curriculum may document growth that is more generalizable but are less likely to show change in students' understanding of relevant content, and are therefore less powerful. For instance, international test scores such as the Programme for International Student Assessment (PISA) are unlikely to change significantly in response to a 2-week intervention because PISA exams cover such a broad array of content relative to what could be covered over a focused unit.

Multi-level assessment (Ruiz-Primo, Shavelson, Hamilton, & Klein, 2002) offers a conceptual framework for responding to these challenges. Assessments can be thought of as occupying variable distances from a curriculum. The 'closest' assessments are embedded within instruction. These kinds of assessments likely provide critical information for teachers that can be used to inform instructional decisions, but they are not likely to provide the kind of information needed to build an empirical argument around the efficacy of an intervention (Hickey & Pellegrino, 2005). Assessments such as unit tests are not as 'close' to the curriculum as embedded assessments but are still closely aligned with the context of instruction. End-of-course exams are positioned at a greater distance from the curriculum, and international comparative exams, like PISA, are among the most distant measures commonly used in educational contexts. Research designs are strengthened and have potential to offer more information when measures are taken at variable distances from the target curriculum (Hickey & Pellegrino, 2005).

Research questions

Given the status of the field's understandings of SSI-based teaching, critiques that have been applied to the SSI research base, and insights gleaned from the multi-level assessment framework, we designed a study based on the following research questions:

(1) To what extent does SSI-based instruction support student learning of science content?
(2) How do assessments at variable distances from the curriculum reveal patterns of learning associated with SSI-based instruction?

Methods

Participants

Sixty-nine secondary students taught by three teachers participated in the study. The science learning goals for the classes related to genetics and molecular biology, and instruction incorporated a focus on biotechnology. The instructional intervention was implemented over approximately three weeks. Teachers used an SSI curriculum that highlighted the use of biotechnology for identifying and treating sexually transmitted diseases. The learning context is described in greater detail below.

The three participating teachers taught science in different high schools across the state of Florida. Two of the teachers taught biology, and one taught integrated science. The integrated science teacher was a biology teacher by training but her school district had adopted an integrated science sequence. The course she taught had a heavy focus on the life sciences, and the primary audience was students in their sophomore year of high school, which was the same group that took biology in the other schools involved in the study. Of the 69 participating students, roughly half were female.

Intervention

In creating the SSI intervention, we drew from an empirically based framework for operationalizing and planning SSI instruction (Presley et al., 2013; Sadler, 2011b). Key aspects of this framework include the following: (a) framing instruction around a compelling issue and explicitly featuring this issue at the outset of instruction; (b) focusing on student learning of core disciplinary ideas; (c) challenging students to collect and/or analyze data related to the issue; (d) providing opportunities for learners to negotiate social dimensions of the issue featured; and (e) using media and information and communication technologies (ICT) as a means of student collection and/or dissemination of information regarding the issue. In this particular intervention, instruction was framed around a narrative case involving the emergence of a novel strain of a sexually transmitted viral disease. The issue of how a new strain of a sexually transmitted disease impacts individuals and communities and how biotechnology can be used to identify disease-causing agents were the central organizing features of the unit; this corresponds to aspect A in the framework highlighted above. The science content learning goals for the unit related to genetics and molecular biology and were aligned with state science standards (consistent with aspect B in the SSI instructional framework). Our research team considered the standards addressed in four 'bundles:' (1) deoxyribonucleic acid (DNA) structure and replication, (2) transcription, translation and protein structure, (3) genetic technologies, and (4) pathogens and immune responses. The eight state standards that correspond to these bundles are presented in Table 1. As the unit unfolded, students followed the work of a scientist as she used biotechnology tools and procedures to identify the disease-causing agent featured within the case as human papilloma virus (HPV). This created multiple opportunities for students to examine and analyze scientific data including electrophoretic data to determine the presence and quantity of DNA and results from real-time polymerase chain reactions (PCRs) (aspect C from the framework). Social dimensions of the issue included class discussions of societal responsibilities regarding the prevention of disease including sexually transmitted viruses. This led to consideration of vaccination policies, which pit societal health interests

Table 1. Florida state science standards aligned with the intervention.

Thematic set	Standards
DNA structure and replication	SC.912.L.16.3: Describe the basic process of DNA replication and how it relates to the transmission and conservation of the genetic information. SC.912.L.16.4: Explain how mutations in the DNA sequence may or may not result in phenotypic change. Explain how mutations in gametes may result in phenotypic changes in offspring.
Transcription, translation and protein structure	SC.912.L.16.5: Explain the basic processes of transcription and translation, and how they result in the expression of genes. SC.912.L.18.4: Describe the structures of proteins and amino acids. Explain the functions of proteins in living organisms. Identify some reactions that amino acids undergo. Relate the structure and function of enzymes.
Genetic technologies	SC.912.L.16.11: Discuss the technologies associated with forensic medicine and DNA identification, including restriction fragment length polymorphism (RFLP) analysis. SC.912.L.16.12: Describe how basic DNA technology (restriction digestion by endonucleases, gel electrophoresis, PCR, ligation, and transformation) is used to construct recombinant DNA molecules (DNA cloning).
Pathogens and immune responses	SC.912.L.14.52: Explain the basic functions of the human immune system, including specific and nonspecific immune response, vaccines, and antibiotics. SC.912.L.16.7: Describe how viruses and bacteria transfer genetic material between cells and the role of this process in biotechnology.

against individual rights (aspect D). In terms of using media and ICT (aspect E), students used several different forms of current media that they accessed through the Internet to identify major characteristics of HPV; compare symptoms, effects, and biological mechanisms of HPV with human immunodeficiency virus (HIV); and identify ways in which HPV and HIV affect individuals and society. As a part of these learning experiences, students were scaffolded to support critical examination of media sources, intended messages, and potential biases. A description of the primary strategies employed to support student access and analysis of media is presented in Klosterman, Sadler, and Brown (2012).

Within the SSI unit, classes engaged in a variety of teaching and learning activities. There were numerous small group activities and discussions that encouraged students to work with their peers to develop understandings of the issue under consideration, the science principles related to this issue, biotechnology tools used in examination of the issue, and social dimensions of the issue. Information was also presented through some teacher lectures, a wet laboratory experience (DNA extraction), and a multi-day laboratory simulation (PCR) including an opportunity for students to analyze data from actual real-time PCR results. Throughout the unit, students completed journaling tasks that challenged them to make sense of new science ideas relative to the issue under consideration. The average implementation time among the three SSI teachers was 14 hours of classroom instruction.

Instruments

Our focus in this study was on student learning of science concepts and the extent to which an SSI curriculum and instructional approach can support learning defined in this way. Advocates of SSI teaching contend that the approach can support a variety of science learning goals beyond understandings of core content; for example, opportunities to negotiate complex SSI can promote specific reasoning competencies (Sadler, Barab, & Scott, 2007). However, while the students in the current study may have developed reasoning competencies and other desirable learning outcomes (e.g. procedural knowledge related to biotechnology, better understandings of the epistemic commitments of science, etc.), the focus of our research was student content knowledge and therefore our research instrumentation assessed student understanding of science concepts. In the introductory section, we justified this attention to science content knowledge because of the significance placed on student content learning, as measured through standardized tests, by policy-makers, school administrators, and teachers.

We employed a multi-level assessment framework (Hickey & Pellegrino, 2005; Klosterman & Sadler, 2010; Ruiz-Primo et al., 2002) to evaluate gains in content knowledge related to genetics and molecular biology. Through the lens of multi-level assessment, proximal and distal assessments were administered to students before and after the interventions, and both instruments focused on assessing student understandings of relevant science concepts. The proximal assessment contained multiple-choice items directly tied to the science content covered in the SSI intervention and was similar to an end-of-unit content examination. The items were created by our research and development team, and they were consistent in language and context with ideas and formats presented in the instructional environment. An expert panel consisting of a biotechnology education coordinator, an experienced high school biology teacher, a PhD biochemist, and a psychometrician reviewed the instrument and provided suggestions for improving face validity. We also pilot tested the instrument with a sample of 32 students not involved in the current study. Based on results from the expert review and the pilot test, we produced a finalized version of the proximal exam; it contained 19 items. Reliability of the instrument, as measured by Cronbach's alpha, was appropriate for the temporal comparisons made: $\alpha_{pre} = .686$, $\alpha_{post} = .827$.

The distal assessment was aligned with the eight content standards on which the interventions were based (see Table 1). The items for the distal assessment were derived from publically released portions of several U.S. state (California, Florida, New York, and Texas) and national (International Baccalaureate and National Assessment of Educational Progress) exams. The distal items focused on science concepts addressed in the instruction but they presented representations of the content abstracted from the instructional contexts. A distal assessment of this sort can be considered a proxy for standardized tests (Hickey & Pellegrino, 2005). It addressed a restricted range of content as opposed to most standardized tests which tend to cover the full range of ideas in a discipline or course, but it presented assessment tasks similar to the kinds of concept-oriented items and format used in high-stakes testings. Thirty-nine items from standardized instruments were identified as aligning with target standards, and an expert panel including two biology teachers, a biotechnology researcher, two science teacher professional development providers, and two science education researchers reviewed the items for scientific accuracy and alignment. An initial version of the exam was piloted with 128 students not involved in the current study. We used classical test theory and item response theory to analyze pilot test data. Based on these results and feedback from the expert panel, we reduced the exam to 18 multiple-choice items. The items demonstrated satisfactory collective reliability on the pre- and post-tests: $\alpha_{pre} = .839$, $\alpha_{post} = .868$.

Data analysis

Statistical significance of proximal and distal gains was evaluated using paired t-tests at the 95% confidence level. Since proximal and distal measures were both related to genetics, we

Table 2. *t*-tests and Cohen's *D* effect sizes for gains.

Test	Pre-test mean (SD)	Post-test mean (SD)	$t_{(df = 68)}$	D_{Gain}
Proximal	7.30(2.93)	14.12(5.41)	10.21**	1.22
Distal	7.20(3.53)	9.76(4.31)	5.52**	0.66

**$p \ll .001$.

employed the Bonferroni correction for inflation of Type 1 error due to multiple tests. Cohen's *D*, a sample size independent measure of gains on a scale of standard deviations, was used as measure of practical significance.

Results

Average student scores on the proximal post-test nearly doubled relative to performance on the pre-test (from 38% to 75%). A paired *t*-test suggests that this gain was statistically significant: $t_{df = 68} = 10.21$, $p \ll 0.001$ (see Table 2). The effect size calculated with Cohen's *D* ($D_{gain} = 1.22$) indicates that the change observed was large and practically significant.

Average student scores on the distal test increased from 38% to 54%. A paired *t*-tests indicated that these gains were statistically significant ($t_{df = 68} = 5.52$, $p \ll .001$), and the effect size ($D_{gain} = 0.66$) was moderate. These results suggest that students demonstrated statistically and practically significant gains on both the proximal and distal measures of biological content knowledge following the SSI intervention. However, the magnitude of gain for the proximal test was twice that for the distal test.

Discussion

We begin our discussion of the study results and its implications with a presentation of its limitations. At the outset of our work, we had to make decisions about what to compare as a means of attaining our goal to explore the efficacy of SSI-based instruction. From a strict research design perspective as outlined in the What Works Clearinghouse Guidelines (Institute of Educational Sciences, 2014), implementing a randomized controlled trial whereby we randomly assign teachers to an intervention or control group would be ideal. However, this approach does not recognize the complex realities of modern schooling; nor does it value the professional roles adopted and decisions made by teachers. It was not practically or ethically feasible for us to implement a true controlled experiment. Furthermore, we were faced with the decision of whether or not to use a comparison group, and if so, what type of comparison to use. We worked with a fourth teacher (with 19 students) who decided to develop and implement her own curriculum addressing the standards in Table 2, and so we initially considered using this group as a comparison. However, we found that results on the proximal assessment were confounded because it was designed to be aligned to the SSI curriculum, and not the comparison curriculum. Furthermore, making comparisons between curricula assumes that the mean is a valid estimate of the population and the variances are equal between treatments; these depend on the assumption that the students in the comparison teacher's classroom are similar in prior knowledge and experience as the students in the SSI group. Indeed, Dawson and Venville (2013) address this by using the same teachers and schools for experimental and comparison groups to ensure that both groups could be modeled with the mean. While some of these concerns could be addressed using robust non-parametric methods, analyisis of covariance (ANCOVA), or multi-level modeling, our most pressing concern was that comparison of a curriculum that was carefully designed in a research context to another which was developed by a solitary teacher in a relatively short time frame may not be meaningful.

Given the concerns above, we ultimately found it best to study the effects of the SSI-based intervention as a pre-post design without drawing upon a potentially confounded comparison group. We recognize that these design decisions limit the generalizability of the findings, but we agree with the arguments presented by Shaffer and Serlin (2004) that employing designs featuring non-random sampling can add important depth to educational research even when generalizability is limited. Given these qualifications, we contend that the results of our study make an important contribution to a growing body of literature (e.g. Barab et al., 2007; Bulte et al., 2006; Dawson & Venville, 2013; Dori et al., 2003; Klosterman & Sadler, 2010; Venville & Dawson, 2010) that provides empirical evidence for the use of SSI-based instruction to support student learning of science content knowledge.

SSI-based instruction to foster science content knowledge

Although researchers have presented many conceptual and theoretical justifications about the integration of SSI into science classrooms to increase students' science content knowledge (Sadler, 2011a; Zeidler, 2003), empirical investigations of the effects of SSI-based teaching on students' science content learning have been relatively limited. Investigations of the effects of SSI-based instruction for supporting student learning have been criticized on the basis of using research designs that do not account for learners' prior knowledge (Bulte et al., 2006; Zohar & Nemet, 2002). This has led to questions regarding the evidence base for using SSI-based teaching as a strategy for teaching science concepts effectively.

Results from our study show that students experiencing the SSI intervention demonstrated statistically significant gains in content understandings as measured by both proximal and distal assessments. Venville and Dawson (2010) provide similar conclusions from a quasi-experimental study of students' learning in the context SSI-based instruction. The Venville and Dawson study and our own investigation share similar constraints: the investigations involve a limited number of classes, and each investigates a particular SSI unit. However, the accumulation of evidence across multiple studies helps to build an empirical case for the efficacy of SSI approaches. Our study and the Venville and Dawson (2010) investigation offer evidence from independently developed SSI-based learning experiences. Common elements include framing instruction around a compelling issue and explicitly featuring the issue at the outset of instruction; focusing on student learning of core disciplinary ideas; challenging students to collect and/or analyze data related to the issue; and providing opportunities for learners to negotiate social dimensions of the issue featured. Combined, these two studies provide compelling evidence that SSI-based instruction, particularly with respect to interventions designed around genetics-oriented SSI, can support significant science content learning.

Multi-level assessment of science content knowledge

The other important contribution of this study corresponds to the second research question, which is based on the multi-level assessment framework employed in the study.

Conceptualizing and implementing assessments at variable distances from curricular interventions can provide useful insights into the effects of interventions that would not be detectable with approaches that adopt more simplistic perspectives on assessment (Ruiz-Primo et al., 2002). For example, in a recent study of a technology-based science intervention, Sadler et al. detected differential impacts of a game-based SSI intervention for students with varying academic levels (Sadler, Romine, Stuart, & Merle-Johnson, 2013). In this study, a research design relying on a single assessment would have obscured interesting and practically significant inter-relationships between student ability levels and the effectiveness of the technology-based learning tools designed to support science learning. Over the past few years, science education researchers have begun to feature multi-level assessment within research designs of empirical studies (e.g. Houle & Barnett, 2008; Nordine, Krajcik, & Fortus, 2011; Romine, Miller, Knese, & Folk, in Press; Zuiker & Whitaker, 2014); however, only a few studies focusing on SSI-based interventions have adopted the approach. In the first application of multi-level assessment in the context of SSI research, Barab et al. (2007) documented statistically significant gains on a proximal assessment of student's science understandings in the context of a technology-enhanced SSI learning experience. However, there were no significant changes in student performance on a distal assessment. This result suggested that students learned significant science content directly aligned with the SSI intervention, but that this learning did not impact student performance on a more distanced instrument that assessed more generalized science ideas removed from the immediate context of instruction. Klosterman and Sadler (2010) employed a multi-level assessment design in the study of an SSI unit related to climate change. They documented statistically significant gains in student understandings on both proximal and distal assessments, although, the effect sizes for the distal assessment were modest ($D = .41$ and $.49$ for two different samples).

In the current study, we have documented statistically significant gains in student performance on proximal and distal assessments. The gains on the proximal assessment were quite large from an effect size perspective as measured in standard deviations ($D = 1.22$). We expect to see evidence of student learning on the proximal assessment because it is directly aligned with the learning experiences in the intervention. The average implementation duration for the SSI unit was 14 hours; we expect that learners will develop new understandings given the fact that they and their teachers spent approximately 3 weeks of classroom time on the unit, although an increase of 1.22 standard deviations – equivalent to moving a student from the 50th percentile to the 88th percentile – is large even considering the duration of instruction.

As expected for a test tied more closely to the curriculum, the language and examples used on the proximal instrument were familiar to the SSI students, whereas, this was not the case for the distal instrument. For example, the SSI intervention focused on real-time PCR and the assessment instrument provided opportunities to interpret real-time PCR data in a question that called for the application of a genetics idea. In contrast to the proximal assessment, the goal of the distal assessment was to measure students' understandings of scientific formalisms that underlie ideas explored within the instructional units. These formalisms are identified within the state standards for which the intervention was designed to address. Student understanding of these formalisms were assessed through a series of items derived from standardized achievement tests and therefore did not draw from specific language, examples or contexts that the students might have experienced in the curriculum. The distal exam is not as far removed from the curriculum as an assessment such as an end-of-course (EOC) exam or a national or international achievement test. EOCs and achievement tests will cover a wide range of science standards that span content that is much broader than can be addressed and explored in a single unit. However, distal exams of the sort employed in this study, that focus on a limited range of content and use representations of the content not aligned with instructional contexts, can serve as valid proxies for high-stakes tests such as EOCs and achievement tests (Barab et al., 2009).

The effect size for gains on the distal exam were moderate ($D = .66$) (Cohen, 1988, 1992). The fact that the distal effect size was approximately half the proximal effect size is consistent with what one would expect from multi-level assessment. Our research team interprets this effect size – equivalent to moving a student from the 50th percentile to the 75th percentile – as important and substantial given the context of our analyses. We see this as evidence that SSI-based instruction can support student learning of science concepts such that student understandings of those concepts are sophisticated enough to transfer that knowledge from the immediate context of their learning experiences to abstracted assessment contexts. Given the push for accountability and the reliance on achievement tests as the most important (or at least the most widely used) metric of student learning (Anderson, 2012) and, in some cases, teacher performance (Papay, 2011), the evidence provided in this study is important for establishing an empirical rationale for employing SSI-based instruction in formal science education contexts.

The data and analyses presented in this study support the assertion that SSI-based teaching can foster student learning of science content and that these gains can lead to improved performance on high-stakes tests. By definition, SSI-based instruction focuses on both a specific issue and its related science content; therefore, it would be unrealistic to expect that a single unit would change student performance on high-stakes exams that cover a wide range of disciplinary content in ways that would be detectable. However, by using a proxy that restricts the range of content covered, we gain some insights into the potential of SSI-based instruction in helping students meet selected state and national standards.

Conclusion

Given positive findings from studies of integrating SSI within a variety of contexts, the evidence presented in this study offers additional support for the use of SSI as a curricular vehicle for students' learning of important science content. This study supports the position that integration of SSI is a way to help our classrooms meet state and national accountability measures while making science relevant and interesting for students.

Acknowledgments

Michelle Klosterman, Julie Bokor, Julie Brown, Tamara Mandel, and Mary Jo Koroly contributed to the creation of the SSI intervention materials and the teacher professional development program. We appreciate the many schools, teachers and students who participated in the project.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This material is based upon work supported by the National Science Foundation under [grant no. 0833521]. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Notes on contributors

Troy Sadler is a Professor of Science Education and Director of the ReSTEM Institute at the University of Missouri.

William Romine is an Assistant professor of biology education at Wright State University. He studies applied educational measurement in the context of formal science learning.

Mustafa S. Topçu is an Associate professor of science education at the Yildiz Technical University. His research interests are socioscientific issue-based teaching and learning, and argumentation.

References

Albe, V. (2008). When scientific knowledge, daily life experience, epistemological and social considerations intersect: Students' argumentation in group discussions on a socio-scientific issue. *Research in Science Education*, *38*(1), 67–90.

American Association for the Advancement of Science (AAAS). (1990). *Science for all Americans*. New York, NY: Oxford University Press.

Anderson, K. J. B. (2012). Science education and test-based accountability: Reviewing their relationship and exploring implications for future policy. *Science Education*, *96*(1), 104–129.

Australian Curriculum, Assessment and Reporting Authority (ACARA). (2015). *The Australian curriculum*. Retrieved from http://www.australiancurriculum.edu.au.

Barab, S. A., Sadler, T. D., Heiselt, C., Hickey, D. T., & Zuiker, S. (2007). Relating narrative, inquiry and inscriptions: Supporting consequential play. *Journal of Science Education and Technology*, *16*, 59–82.

Barab, S. A., Scott, B., Siyahhan, S., Goldstone, R., Ingram-Goble, A., Zuiker, S. J., & Warren, S. (2009). Transformational play as a curricular scaffold: Using videogames to support science education. *Journal of Science Education and Technology*, *18*(4), 305–320.

Bulte, A. M. W., Westbroek, H. B., De Jong, O., & Pilot, A. (2006). A research approach to designing chemistry education using authentic practices as contexts. *International Journal of Science Education*, *28*(9), 1063–1086.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

Dawson, V., & Venville, G. (2013). Introducing high school biology students to argumentation about socioscientific issues. *Canadian Journal of Science, Mathematics and Technology Education*, *13*(4), 356–372.

Dori, Y. J., Tal, R., & Tsaushu, M. (2003). Teaching biotechnology through case studies: Can we improve higher order thinking skills of nonscience majors? *Science Education*, *87*, 767–793.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, *84*(3), 287–312.

Eastwood, J. L., Sadler, T. D., Zeidler, D. L., Lewis, A., Amiri, L., & Applebaum, S. (2012). Contextualizing nature of science instruction in socioscientific issues. *International Journal of Science Education*, *34*(15), 2289–2315.

Eilks, I. (2010). Making chemistry teaching relevant and promoting scientific literacy by focusing on authentic and controversial socio-scientific issues. Presentation at the meeting for the Society for didactics in chemistry and physics. Potsdam.

Harris, R., & Ratcliffe, M. (2005). Socio-scientific issues and the quality of exploratory talk – what can be learned from schools involved in a 'collapsed day' project? *Curriculum Journal*, *16*, 439–453.

Hickey, D. T., & Pellegrino, J. W. (2005). Theory, level, and function: Three dimensions for understanding transfer and student assessment. In J. P. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 251–293). Greenwich, CT: Information Age.

Houle, M. E., & Barnett, G. M. (2008). Students' conceptions of sound waves resulting from the enactment of a new technology-enhanced inquiry-based curriculum on urban bird communication. *Journal of Science Education and Technology*, *17*, 242–251.

Hughes, G. (2000). Marginalization of socioscientific material in science-technology-society science curricula: Some implications for gender inclusivity and curriculum reform. *Journal of Research in Science Teaching*, *37*, 426–440.

Institute of Education Sciences. (2014). *What works clearinghouse*. Retrieved December 2, 2014, from http://ies.ed.gov/ncee/wwc/

Khishfe, R., & Lederman, N. G. (2006). Teaching nature of science within a controversial topic: Integrated versus nonintegrated. *Journal of Research in Science Teaching*, *43*, 395–418.

Klosterman, M., Sadler, T., & Brown, J. (2012). Viral news: Media literacy for the 21st century. *Science Scope*, *35*(9), 61–69.

Klosterman, M. L., & Sadler, T. D. (2010). Multi-level assessment of scientific content knowledge gains associated with socioscientific issues based instruction. *International Journal of Science Education*, *32*, 1017–1043.

Kolstø, S. D. (2001). Scientific literacy for citizenship: Tools for dealing with the science dimension of controversial SSI. *Science Education*, *85*, 291–310.

Kolstø, S. D. (2006). Patterns in students' argumentation confronted with a risk-focused socio-scientific issue. *International Journal of Science Education*, *28*(14), 1689–1716.

Lee, M.-K., & Erdogan, I. (2007). The effect of science-technology-society teaching on students' attitudes toward science and certain aspects of creativity. *International Journal of Science Education*, *29*(11), 1315–1327.

Ministry of National Education of Turkey (MONE). (2013). *Science and technology curriculum of elementary schools (3th–8th grades)*. Ankara: Board of Education.

National Research Council (NRC). (1996). *National science education standards.* Washington, DC: National Academy Press.

Nordine, J., Krajcik, J., & Fortus, D. (2011). Transforming energy instruction in middle school to support integrated understanding and future learning. *Science Education*, *95*, 670–699.

Orpwood, G. (2001). The role of assessment in science curriculum reform. *Assessment in Education*, *8*, 135–151.

Papay, J. P. (2011). Different tests, different answers: The stability of teacher value-added estimates across outcome measures. *American Educational Research Journal*, *48*(1), 163–193.

Parchmann, I., Gräsel, C., Baer, A., Nentwig, P., Demuth, R., & Ralle, B. (2006). Chemie im Kontext – A symbiotic implementation of a context-based teaching and learning approach. *International Journal of Science Education*, *28*(9), 1041–1062.

Pedretti, E. (1999). Decision making and STS education: Exploring scientific knowledge and social responsibility in schools and science centers through an issues-based approach. *Journal of School Science and Mathematics*, *99*(4), 174–181.

Pouliot, C. (2008). Students' inventory of social actors concerned by the controversy surrounding cellular telephones: A case study. *Science Education*, *92*(3), 543–559.

Presley, M. L., Sickel, A. J., Muslu, N., Merle-Johnson, D., Witzig, S. B., Izci, K., & Sadler, T. D. (2013). A framework for socio-scientific issues based education. *Science Educator*, *22*(1), 26–32.

Roberts, D. A. (2007). Scientific literacy/science literacy. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 729–780). Mahwah, NJ: Lawrence Erlbaum Associates.

Romine, W. L., Miller, M. E., Knese, S., & Folk, W. R. (2016). Multi-Level assessment of middle school students' interest in the health sciences: Development and validation of a new measurement tool. *CBE Life Sciences Education*, *15*(2), 1–13.

Romine, W. L., & Sadler, T. D. (2014). Measuring changes in interest in science and technology at the college level in response to two instructional interventions. *Research in Science Education, 12*, 1–19.
Romine, W. L., Sadler, T. D., & Kinslow, A. (2016). Meaningful Assessment of New Generation Science Literacy in the Context of Socio-scientific Issues: Development and Validation of the Quantitative Assessment of Socio-scientific Reasoning (QuASSR). Manuscript submitted for publication.
Ruiz-Primo, M. A., Shavelson, R. J., Hamilton, L., & Klein, S. (2002). On the evaluation of systemic science education reform: Searching for instructional sensitivity. *Journal of Research in Science Teaching*, *39*(5), 369–393.
Sadler, T. D. (2011a). Situating socioscientific issues in classrooms as a means of achieving goals of science education. In T. Sadler (Ed.), *Socio-scientific issues in the classroom: Teaching, learning and research* (pp. 1–9). New York, NY: Springer.
Sadler, T. D. (2011b). Socio-scientific issues based education: What we know about science education in the context of SSI. In T.D. Sadler (Ed.), *Socio-scientific issues in the classroom: Teaching, learning and research* (pp. 353–367). New York, NY: Springer.
Sadler, T. D., Barab, S. A., & Scott, B. (2007). What do students gain by engaging in socioscientific inquiry? *Research in Science Education*, *37*, 371–391.
Sadler, T. D., Klosterman, M. L., & Topcu, M. S. (2011). Learning science content and socio-scientific reasoning through classroom explorations of global climate change. In T. D. Sadler (Ed.), *Socio-scientific issues in the classroom: Teaching, learning and research* (pp. 45–77). New York: Springer.
Sadler, T. D., Romine, W. L., Menon, D., Ferdig, R. E., & Annetta, L. (2015). Learning biology through innovative curricula: A comparison of game- and nongame-based approaches. *Science Education*, *99*, 696–720.
Saunders, K. J., & Rennie, L. J. (2013). A pedagogical model for ethical inquiry into socioscientific issues in science. *Research in Science Education*, *43*(1), 253–274.
Shaffer, D. W., & Serlin, R. C. (2004). What good are statistics that don't generalize? *Educational Researcher*, *33*, 14–25.
Tal, R., & Hochberg, N. (2003). Assessing high order thinking of students participating in the 'WISE' project in Israel. *Studies in Educational Evaluation*, *29*, 69–89.
Tal, T., & Kedmi, Y. (2006). Teaching socioscientific issues: Classroom culture and students' performances. *Cultural Studies in Science*, *1*, 615–644.
Topcu, M. S., Sadler, T. D., & Yilmaz-Tuzun. (2010). Preservice science teachers' informal reasoning about socioscientific issues: The influence of issue context. *International Journal of Science Education*, *32*, 2475–2495.
Venville, G. J., & Dawson, V. M. (2010). The impact of a classroom intervention on grade 10 students' argumentation skills, informal reasoning and conceptual understanding of science. *Journal of Research in Science Teaching*, *47*, 952–977.
Walker, K. A., & Zeidler, D. L. (2007). Promoting discourse about SSI through scaffolded inquiry. *International Journal of Science Education*, *29*(11), 1387–1410.
Yager, S. O., Lim, G., & Yager, R. (2006). The advantages of an STS approach over a typical textbook dominated approach in middle school science. *School Science and Mathematics*, *106*, 248–260.
Zeidler, D. L. (2003). *The role of moral reasoning and discourse on socioscientific issues in science education*. Dordrecht: Kluwer.
Zeidler, D. L., Applebaum, S. M. & Sadler, T. D. (2011). Enacting a socioscientific issues classroom: Transformative transformations. In T. D. Sadler (Ed.), *Socio-scientific issues in science classrooms: Teaching, learning and research* (pp. 277–306). New York: Springer.
Zeidler, D. L., Sadler, T. D., Applebaum, S., & Callahan, B. E. (2009). Advancing reflective judgment through socioscientific issues. *Journal of Research in Science Teaching*, *46*, 74–101.
Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching*, *39*, 35–62.
Zuiker, S., & Whitaker, R. (2014). Refining inquiry with multi-form assessment: Formative and summative assessment functions for flexible inquiry. *International Journal of Science Education*, *36*, 1037–1059.