Routledge
Taylor & Francis Group

International Journal of Science Education

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tsed20

Measurement of socio-scientific reasoning (SSR) and exploration of SSR as a progression of competencies

William Lee Romine , Troy D. Sadler , Jenny M. Dauer & Andrew Kinslow

INTERNATIONAL JOURNAL OF SCIENCE EDUCATION
https://doi.org/10.1080/09500693.2020.1849853

Measurement of socio-scientific reasoning (SSR) and exploration of SSR as a progression of competencies

William Lee Romine [a], Troy D. Sadler [b], Jenny M. Dauer [c] and Andrew Kinslow[d]

[a]Department of Biological Sciences, Wright State University, Dayton, OH, USA; [b]School of Education, University of North Carolina, Chapel Hill, NC, USA; [c]School of Natural Resources, University of Nebraska-Lincoln, Lincoln, NE, USA; [d]Rock Bridge High School, Columbia Public Schools, Columbia, MO, USA

ABSTRACT

Socio-scientific reasoning (SSR) is key to helping students take informed positions around socio-scientific issues (SSI). SSR comprises four competencies: recognising complexity of SSI's, multiple perspectives around SSI's, the need for ongoing inquiry around SSI's, and skepticism around different parties' claims made about SSI's. The Quantitative Assessment of SSR (QuASSR) provides a promising measurement framework, but there are still important questions around the ability of this instrument to measure transfer across different scenarios and change in SSR over an intervention. Further, prior work suggests that the four competencies may constitute a progression. We explored the ability of the QuASSR to measure transfer of SSR across three different SSI's using 2-faceted and multi-faceted Rasch models. We used path analysis to test the hypothesis that competencies associated with SSR formed a progression. We found transfer or neartransfer of SSR across the three scenarios, and that the competencies comprise a unidimensional hierarchy. Perspective-taking is a necessary bridge between students' understanding of complexity and the higher-level competencies of inquiry and skepticism. Inquiry and skepticism were found to be conditionally independent upon accounting for perspective-taking, supporting the idea that seeing multiple perspectives around SSI's is central to development of the other SSR competencies.

ARTICLE HISTORY
Received 30 December 2019
Accepted 6 November 2020

KEYWORDS
Environmental education; scientific literacy; reasoning; socio-scientific reasoning

Socio-scientific issues (SSI) are complex social challenges that connect to science and provide contexts for science learning. Using SSI as a teaching context can help students develop scientific concepts, ideas about how science works, and science practices such as argumentation (Khishfe et al., 2017; Romine et al., 2017). These basic outcomes are the primary target of approaches to science education in Vision I science literacy (Roberts & Bybee, 2014). However, SSI teaching has potential to positively impact the way learners come to interact with 'real life' issues and situations that emerge from the intersections of science and society. This leveraging of science understandings and practices along with social, political, economic and ethical perspectives to make decisions or engage in

CONTACT William Lee Romine romine.william@gmail.com Department of Biological Sciences, Wright State University, 3460 Colonel Glenn Highway, Dayton, OH 45435, USA
Supplemental data for this article can be accessed https://doi.org/10.1080/09500693.2020.1849853.

problem solving is central to what Roberts and Bybee (2014) have classified as Vision II scientific literacy (SL). More recently, a Vision III perspective has been proposed, which calls for an increased focus on civic engagement (Rudolph & Horibe, 2016). In line with the Bildung cultural tradition in Northern and Central Europe, Sjöström and Eilks (2018) have proposed that science curricula increase their focus on philosophical values and political issues towards preparing students for responsible citizenship in an increasingly connected world. Socio-scientific reasoning (SSR) represents a set of related competencies employed in the negotiation and resolution of SSI and, therefore, aligns directly with Vision II and Vision III SL frameworks (Romine et al., 2017; Sjöström & Eilks, 2018).

Scientific literacy has always represented an amorphous idea, meant to inspire the pursuit of broad goals for science education (DeBoer, 2000). In contrast, SSR was conceptualised as a construct that could be operationalised and measured (Sadler et al., 2007). As such, SSR was defined in terms of four dimensions, each of which represented a competency necessary for making sense of and taking informed positions on SSI:

- Recognising the inherent **complexity** of issues and therefore not jumping to naïve conclusions;
- Understanding that SSI are subject to ongoing **inquiry** and being able to identify information that is missing;
- Analysing SSI from **multiple perspectives** and appreciating the unique concerns of various stakeholders;
- Exhibiting reflective **skepticism** in the processing and analysis of information about SSI from potentially biased sources.

Initial attempts to measure SSR relied on interviews and surveys with open-ended response items (Sadler et al., 2007), which prohibited collection of large data sets. To remedy this, the Quantitative Assessment of SSR (QuASSR) was developed using ordered multiple choice (OMC) items (Romine et al., 2017). The QuASSR presents students with a scenario that briefly describes a specific SSI and poses a set of 10 questions, each targeting one SSR competency. In initial work with the QuASSR, it was found that items associated with the four SSR competencies could be considered a one-dimensional construct, given a Rasch measurement framework, and that the instrument can produce a reliable measure of this construct. Subsequent QuASSR research has yielded validity evidence for the instrument through expert reviews of the construct, its subdimensions, and alignments between dimensions and individual QuASSR items (Kinslow, 2018; Womack et al., 2019). In the initial study, the QuASSR was used to assess a 1-week SSI-based intervention and did not provide evidence of change in SSR. This result complimented other studies, using qualitative data collection methods, that suggested SSR can be resistant to change especially in response to interventions which are limited in time (Sadler et al., 2011). Additionally, the hierarchy of item difficulty in the QuASSR work suggested relationships among the SSR competencies are indicative of a possible progression (Romine et al., 2017). That is, QuASSR data suggested that the growth in the **Complexity** competency may facilitate growth in **Multiple Perspective-taking**, which is likely necessary for growth with the **Skepticism** and **Inquiry** competencies.

The current study further explores the SSR construct and the QuASSR as a tool for measuring SSR. In this work, we collect a larger QuASSR data set using multiple SSI scenarios, from undergraduate students engaged in a semester-long introductory science course built on SSI learning experiences. In analysing data collected at the beginning and end of the course, we can explore the extent to which students' SSR competencies improve over time. The data also make it possible to test relationships among SSR dimensions hypothesised in previous work. The study is guided by three research questions:

(1) How does introduction of multiple assessment scenarios impact construct validity of the QuASSR, and how consistently do different scenarios measure SSR as a unidimensional construct over time?
(2) How does student SSR, as measured by the QuASSR, change in response to SSI-based instruction?
(3) How do the elements of SSR relate to one another in terms of a possible progression of competencies?

Theoretical framework: transfer and SSR

Extensive evidence suggests that students can learn important outcomes associated with teaching in the context of SSI. For example, SSI-based teaching can support student learning of science content (Sadler et al., 2016), nature of science (Lederman et al., 2014), and scientific practices such as argumentation (Dawson & Venville, 2010) and modelling (Peel et al., 2019). SSI-based teaching can also be helpful in terms of generating student interest in learning science and help students appreciate the relevance of science learning experiences to their broader lives (Ke et al., in press). In light of Vision II SL, SSI based teaching ought to do more for students than support their learning of science content and practices and development of interest. The Bildung concept, aligned with Vision III SL, takes this further, suggesting that SSI-based teaching should prepare students to serve as responsible citizens in a democratic society (Sjöström & Eilks, 2018); this brings home the necessity for understanding how students reason around and negotiate SSI. Transfer theory (Haskell, 2000) offers a helpful framework for considering how SSI-based teaching may benefit learners. Transfer explains how learners use ideas and competencies, experienced in one context, in new contexts with different characteristics and features. Haskell (2000) conceptualises transfer in terms of a set of six progressive stages ranging from relatively simple connections between prior knowledge and a learning opportunity to situations in which the idea or competency to be transferred must be applied in new ways or reinterpreted.

The idea of SSR emerged from consideration of what students might transfer from an SSI learning experience to an opportunity to negotiate a different SSI. Consider, for example, students learning about climate change. Following this experience, the students will have to make sense of other SSI encountered through media, discussion with friends, or as a ballot initiative. The new issue could be something closely related to climate change such as a proposal to impose regulations on CO_2 emissions. In this case, we would expect that ideas learned as a part of the climate-themed SSI experience would influence how students considered the question of CO_2 regulations; hence, students would have demonstrated near transfer. If the new issue dealt with a completely different topic, say labelling of genetically modified foods, students would not be in a position to transfer science content from the climate change experience. However, students may develop competencies in the context of the climate change learning experience that help them to negotiate and make more informed decisions about different issues, such as identifying a need for information or recognising potentially biased sources of information. Research around the operationalisation and measurement of SSR represents an attempt to explore the competencies that learners can develop in one issue context and transfer to other contexts. We would expect that a learner would demonstrate similar SSR competencies across varying issue contexts and that one's SSR would improve in response to opportunities to practice the competencies that constitute SSR.

Review of related literature

General frameworks for reasoning and SSI

Since the inception of SSI research (Fleming, 1986a, 1986b), scholars have been interested in reasoning processes employed by students as they navigate complex issues. Early attempts to frame this kind of work relied on 'informal reasoning' as a construct for describing these thinking processes (Sadler & Zeidler, 2004). However, this proved to be a difficult construct to use for tracking student development of practice because it is conceptualised in several different ways and can be applied to any open-ended problem, regardless of whether it is SSI-based. Even when confined to the analysis of SSI, the construct took on a range of interpretations. For example, Yang and Anderson (2003) classified students' informal reasoning as social or scientific. Sadler and Zeidler (2005) differentiated student informal reasoning patterns as rationalistic, emotive, and intuitive. Wu and Tsai (2007) sorted students' informal reasoning in terms of social, economic, ecological, and scientific/technological orientations. While these classifications helped to qualitatively describe students' thinking, the construct offered limited potential for advancing efforts to systematically quantify reasoning.

Other researchers have framed exploration of student thinking about SSI with other constructs that are not specific to SSI and therefore can be broadly interpreted. For example, argumentation has been used to assess competencies associated with forming and advancing positions on SSI (Dawson & Venville, 2010; Evagorou & Osborne, 2013). Ethical sensitivity has been employed as a framework to explore the extent to which students' reasoning in the context of SSI incorporates the consideration of moral and ethical dimensions of the issues (Lee et al., 2013). Zeidler et al. (2009) studied epistemological development through reflective judgment in the context of SSI teaching. As was the case with informal reasoning; these constructs offered useful tools for exploring teaching and learning but did not provide precise frameworks for measuring changes in student competencies.

Socio-scientific reasoning

SSR was developed as an attempt to more specifically operationalise what students gain through SSI learning experiences and that might transfer to their negotiation of other SSI. The first study of SSR provided qualitative descriptions of ranges of student abilities associated with SSR dimensions (Sadler et al., 2007). Since the first articulation of SSR, several groups have explored other dimensions of the construct. Simonneaux and Simonneaux (2009) suggested that SSR include identification of risk/uncertainty and consideration of cultural and ethical principles. In follow-up work, this group has suggested that argumentation and sustainability should be dimensions of SSR (Morin et al., 2013, 2014). Karahan and Roehrig (2016) added identification of social domains, cost–benefit analyses, and understanding the boundaries of issues. Kahn and Zeidler (2019) conducted a conceptual analysis of perspective taking (one of the SSR competencies) and suggested 'socioscientific perspective taking' as its own construct as opposed to a dimension nested within SSR. In our analysis of this work as well as our perspectives on learner thinking in the context of SSI, we agree with the point that learners need to be able to do more than what is described by the original SSR competencies when negotiating SSI, but some of what has been suggested as additions to the construct (such as argumentation) seem to us to be beyond the scope of what SSR is meant to capture. In other cases, such as Karahan and Roehrig's suggestions, we think that some of what is being called for as an expansion of SSR is subsumed within existing SSR constructs.

A research group led by Eggert and Bögeholz (2010) has proposed the idea of 'socioscientific decision making' and developed instrumentation for assessing it (Eggert et al., 2013). This work focuses on learner analyses of the pros and cons of competing decisions, which is similar to the SSR competency of perspective taking. While there are conceptual connections between SSR and socioscientific decision making, the measurement strategies used and the associated research are distinct relative to one another.

Another thread of SSR research has focused on the extent to which students improve their SSR competencies in response to SSI learning experiences. Relatively short (one to three weeks) SSI units of instruction have not impacted SSR competencies (Romine et al., 2017; Sadler et al., 2011). Longer term interventions are needed. Preservice teachers in a semester long course that focused on how to teach with SSI showed gains on SSR (Cansiz, 2014), and high school students participating in a six-week field-based SSI summer course demonstrated significant gains in SSR (Kinslow et al., 2019). However, these contexts for instruction were very similar to the contexts for assessment. In the Cansiz (2014) study, the course focused extensively on nuclear power and the scenario used to elicit students' SSR was also related to nuclear power. Therefore, while students showed growth in SSR, the study designs did not allow for investigations of transfer. Also, both of these studies feature relatively small sample sizes and qualitative data coded with ordinal rubrics. In order to advance SSR research, measures that can be used with larger sample sizes would be useful.

Methods

Instrumentation

The structure of the QuASSR is described in detail in Romine et al. (2017) Briefly, the QuASSR contains 10 two-tiered questions which are crossed between scenarios. The first tier asks the student to respond to a yes/no question related to complexity of the issue, multiple perspectives on the issue, or the need for skepticism and additional knowledge. For the second tier, we constructed responses based on qualitative data from interviews and open-ended surveys of SSR (Sadler et al., 2007, 2011). The questions within each scenario have the same wording structure, and assess the same content, but relate to the scenario. Two questions measure ability to see the complexity in SSI, two questions measure perspective-taking, three questions measure ability to see the importance of continuous inquiry in the negotiation of SSI, and three questions measure students' skepticism. In this study, we used the scenario on fracking which was developed in the context of an SSI-based instructional experience focusing on fracking (Romine et al., 2017) as well as two additional scenarios: (1) Lake Nothan, which centred on a water management issue, and (2) use of antibiotics, which focused on controversy around whether or not to use antibiotics in agriculture (see the Supplementary Materials). The Lake Nothan scenario was initially developed as a part of the first SSR study, and the use of antibiotics scenario was first created as a part of a study of high school students' learning (Sadler et al., 2007). In both cases, these scenarios were originally developed to elicit open-ended responses, so they were modified to fit the QuASSR format. Items were scored polytomously (0 = low SSR; 1 = moderate SSR, and 2 = high SSR) Table 1.

Instructional context

The study was conducted in an introductory course, *Science and Decision-making for a Complex World* (Alred & Dauer, 2020; Dauer & Forbes, 2016; Dauer et al., 2017) at a large Midwestern university in the United States. The course was required for all STEM and non-STEM students in the agriculture and natural resources college and is described in detail in Dauer et al. (in press). Learning objectives of the course centred on science-informed decision-making, information literacy and systems thinking. During the course, students performed a structured decision-making exercise for four SSI salient to the region. The focus was: (1) Should we conserve prairie dogs? (2) Should we use biofuels? (3) How do we best solve food insecurity? and (4) Should we further restrict the amount of water used for agriculture? As with previous interventions where SSR was studied (Romine et al., 2017; Sadler et al., 2011), the course focused on understanding SSI through the lens of multiple perspectives by examining social factors that contribute to the controversy, and on finding, evaluating and applying scientific data to consider potential solutions to complex SSI. Unique to this intervention was

Table 1. Structure of the QuASSR by scenarios, items, and competencies.

		Fracking	Nothan	Antibiotic
Complexity	Item	1	1	1
	Item	2	2	2
Perspective-taking	Item	3	3	3
	Item	4	4	4
Inquiry	Item	5	5	5
	Item	7	7	7
	Item	8	8	8
Skepticism	Item	9	9	9
	Item	10	10	10
	Item	11	11	11

Note: Item 6 asks for students' position on the SSI, and is not used for measurement of SSR.

its 15-week length and the use of a structured decision-making framework that allowed students to separately consider the role of scientific information and students' priorities for economic, environmental, ethical and cultural outcomes.

The course structure included lecture sections with approximately 120 students, and smaller recitation sections. We collected data from students in three lecture sections, two that were co-taught by one of the authors. Each lecture was characterised by active learning, peer instruction, and group discussion.

Participants

Two hundred seventy-three students (76% of the total enrolled in the three sections) consented to research and completed the three SSR scenarios before and after 15 weeks of instruction. One hundred seventy eight (65%) reported freshman standing, 63 (23%) were sophomores, 19 (7%) were juniors, and 11 (4%) were seniors. Two hundred thirteen (78%) reported pursuing a STEM major, 52 (19%) reported pursuing a non-STEM major, and 8 (3%) did not declare a major.

RQ1: construct validity and dimensionality

The first question was addressed from three perspectives: (1) the correlation; (2) the 2-faceted Rasch model, and (3) the multi-faceted Rasch model. The correlational perspective involved calculating Rasch measures for separate scenarios as if they were separate instruments, and exploring the strength of correlation between the measures. This exploration of inter-scenario reliability is similar in principle to inter-rater reliability: high correlation between measures derived from separate scenarios lends evidence that these are measuring the same thing.

The two-faceted Rasch approach, where person and item measures are mapped onto a common scale, is the approach most commonly used in science education research (i.e. Boone, 2016; Planinic et al., 2019). If used with items from multiple scenarios, this model employs the restrictive assumption that individual items measure SSR independently of the scenario from which they are derived. This assumption can then be falsified through misfit of particular items with the model and inspection of item residuals to detect violations of the unidimensionality and local independence assumptions (Bond & Fox, 2001).

The multi-faceted Rasch approach is an expansion of the 2-faceted model allowing positioning of person ability as well as item and scenario difficulty measures along a common scale. In this model, consistency between items across particular scenarios is evaluated by the displacement of scenario difficulty along the SSR scale (Linacre, 2009). In this context, the assumption that scenarios are of similar difficulty can be falsified by inspecting the differences in the measures of scenario difficulty. Scenario misfit with the Rasch model allows us to detect violations of the expected trend that students with higher SSR are more likely to achieve higher measures on particular scenarios, and conversely, that students tend to express lower levels of SSR on the harder scenarios.

Reliability and correlational analysis

Before undertaking correlational analysis, it was necessary to quantify the reliability of each scenario. Measurement error attenuates the magnitude of the observable correlation

(Spearman, 1904) and is therefore a limiting factor for yielding repeatable inferences regarding group-wise and temporal differences. With a 3-scenario usage in this study, we used Rasch reliability measures to quantify the internal consistency of the individual scenarios as well as use of multiple scenarios. After reliability analysis, we calculated the Pearson correlation, with Spearman's (1904) correction for measurement error, between Rasch logit measures. The higher the correlation between Rasch logit measures, the stronger the case we can make that the respective scenarios measure a similar construct.

2-Faceted Rasch analysis

Exploration of construct validity with respect to the Rasch model entailed two stages. First, we used the Rasch partial credit model (Masters, 1982) to model each separate scenario with WINSTEPS software (Linacre, 2006). Systematic dependence of items on the scenarios was first evaluated by inspecting the correlations of item residuals after fitting the model. Parallel items from separate scenarios with residual correlations above 0.7 indicate practically significant departure from local independence introduced by the scenario context (Linacre, 2006). We used principal components analysis (PCA) on residuals after the model was fit to the items across the three scenarios, on the pre- and post-tests. If the items are measuring a single SSR dimension, then a first eigenvalue at or below 2 is expected (Linacre & Tennant, 2009; Raiche, 2005). In the case of multidimensionality, we inspected items with loadings above 0.5 onto the residual dimension to better understand which items were responsible for the observed multidimensionality. If it happened that these groups of items came from a particular scenario, this would have suggested scenario-induced multidimensionality created by lack of transfer across scenarios.

Validity with respect to the multi-faceted Rasch model

Combining all facets into a single model yields item estimates which are independent of scenarios and testing instance, thereby giving a more generalisable look at the SSR construct and how a multi-scenario QuASSR instrument measures it across time. Multi-faceted Rasch analysis was used with FACETS software (Linacre, 2009) to study the hierarchy and construct validity of items measuring the elements of SSR and their consistency across scenarios and testing instances. We used a partial credit model, meaning that each item was allowed to have its own unique ordinal scale. However, given our focus on item consistency across scenarios, item step thresholds were assumed to be unchanging within parallel items across scenarios.

In a typical two-faceted context, the Rasch model expects that the probability of a student expressing high SSR should be proportional only to the student's ability and the item's difficulty (Bond & Fox, 2001). When the 'scenario' facet is added, it is expected that this likelihood will be affected in the same way by difficulty of the scenario. The same goes with the 'testing instance' facet. For all facets, mean squares infit and outfit were used as an indicator of the extent to which responses to each item, across each scenario and testing instance, fit this expectation. These have expected values of 1, but Wright et al. (1994) suggest that these indices can range between 0.5 and 1.5 for useful items and scenarios.

RQ2: changes in SSR

Our findings regarding the measurement properties of the QuASSR for this sample of students informed the way we used the measures to evaluate the impact of instruction on SSR. Since this was a pre–post design, we used a paired t-test on students' Rasch logit measures for the pre- and post-tests under the null hypothesis of no difference in SSR between the beginning and end of instruction. Rejection of this null hypothesis occurred when a significant difference at the 95% confidence level (2-tailed) was detected.

RQ3: progression of SSR competencies

Evidence presented in previous work (Romine et al., 2017) suggests a potential progression of SSR competencies. Within an instructional context, this suggests that learning with respect to the less difficult competencies like Complexity and Perspective-taking may support mastery of the more difficult competencies like Inquiry and Skepticism. Previous work shows that the easiest aspect of SSR is identification of complexity. Slightly harder is recognition that people with different interests can have unique perspectives on issue-related information and potential solutions even if presented with the same data. The most difficult tasks relate to recognition of the need for additional data (inquiry) and demonstration of skepticism in the face of potentially biased information. In addition to using the Wright map to evaluate the progression of these competencies, we utilised path analysis to evaluate the extent to which changes in the less difficult competencies due to the intervention supported changes in the more difficult competencies. We used Mplus (Muthén & Muthén, 2007) as a tool for evaluating the support that our data lend to this hypothesised progression. The assumptions of the path-analytic approach (Kline, 2015) lend themselves well to evaluation of progressions. The assumption of model existence, and absence of alternative models, is supported by prior qualitative (Kinslow et al., 2019; Sadler et al., 2007) and quantitative work related to SSR (Romine et al., 2017). The assumptions of directional and temporal causality align closely with the idea of a progression – namely that students will tend to identify with simpler ideas before identifying with more complex ideas; or from a pedagogical perspective, that students should master simpler concepts before they can be expected to master those of greater complexity. For these reasons, path analysis has demonstrated effectiveness in evaluating the support data provide for hypothetical learning progressions (Todd & Romine, 2017; Wilson, 2009a, 2009b).

Strength of evidence for the progression was framed in terms of the ability of our proposed model to reproduce the covariance structure among the learning gains within each competency. The model was fit using maximum likelihood estimation. To evaluate fit, we first used the relatively conservative chi-square statistic with respect to the null hypothesis that the model is supported by the data. Since this chi-square statistic tends to be conservative, we also used the root mean square error of approximation (RMSEA), the Tucker-Lewis Index (TLI), the Comparative Fit Index (CFI), and the standardised root mean square residual (SRMR) as additional measures of fit. For a good-fitting model, we expected an RMSEA value below 0.06, TLI and CFI values above 0.95, and an SRMR value below 0.08 (Hu & Bentler, 1999).

Table 2. Correlation between Rasch measures on individual scenarios after correction for measurement error due to lack of internal consistency (reliability indices in Table 3).

Pre-test	Fracking	Nothan	Antibiotic
Fracking	1		
Nothan	0.985	1	
Antibiotic	0.829	0.840	1
Post-test	Fracking	Nothan	Antibiotic
Fracking	1		
Nothan	0.994	1	
Antibiotic	0.931	0.872	1

Results

RQ1: how does introduction of multiple assessment scenarios impact the construct validity of the QuASSR, and how consistently do different scenarios measure SSR as a unidimensional construct?

Reliability and correlational analysis

The items within the individual scenarios measured SSR unidimensionally (see Table 3), and produced scale reliabilities above 0.59. Using multiple scenarios within a single test increased reliability to above 0.7 for two scenarios and above 0.8 for three scenarios.

Correlational analyses of scenario measures at each time point (Table 2) support the hypothesis of transfer across scenarios. When error due to lack of internal consistency was removed, we found that the three scenarios were highly correlated on both the pre- and post-tests. The correlations (on pre- and post-tests) between fracking and Nothan scenarios were particularly high, and while correlations with the antibiotic scenario were lower, the relationships across all scenarios suggested significant associations.

Analysis of item-Level inconsistency with respect to the 2-faceted Rasch model

The 2-faceted Rasch partial credit model indicated that use of all three scenarios within a single testing does lead to some departure from unidimensionality and minor local dependency between items, but the data largely suggested that these inconsistencies were caused at the item level, not at the scenario level. On the pre-test, the highest item residual correlation was 0.4 (between item 10 on the Fracking and Antibiotic scenarios), and on the post-test, the highest item residual correlation was 0.4 (between items 3 and 4 on the Fracking scenario). The former result suggests small scenario-driven item dependency on the pre-test, whereas the latter result suggests small dependency of two perspective-taking items on the fracking scenario. These sit well below 0.7, suggesting

Table 3. Measures of Rasch reliability and unidimensionality for single and multiple-scenario usage of the QuASSR.

		Pre-test		Post-test	
		ρ_{Rasch}	First Eigenvalue	ρ_{Rasch}	First Eigenvalue
Single-Scenario	Fracking	0.61	1.78	0.67	1.68
	Nothan	0.60	1.43	0.65	1.56
	Antibiotic	0.59	1.58	0.67	1.45
Two-Scenario	Fracking + Nothan	0.76	2.01	0.80	2.04
	Fracking + Antibiotic	0.74	2.17	0.80	1.99
	Nothan + Antibiotic	0.72	1.83	0.79	1.81
Three-Scenario	Fracking + Nothan + Antibiotic	0.81	2.46	0.85	2.37

that the scenarios do not introduce significant levels of systematic variance in the QuASSR unrelated to SSR.

PCA on the Rasch residuals corroborated this conclusion (Table S1 in the Online Supplementary Materials). Eigenvalues of 2.46 and 2.37 for the first residual factor on the pre and post-tests, respectively, suggest a minor departure from unidimensionality (Linacre & Tennant, 2009). Analysis of item loadings onto the first residual factor suggested that systematic differences between scenarios is not to blame for this. Rather, on the pre-test, item 9 falls slightly out of the main SSR dimension (fracking = 0.61, Nothan = 0.55, antibiotic = 0.53), whereas on the post-test, item 11 shows some distinction from the main dimension (fracking = 0.50, Nothan = 0.47, antibiotic = 0.62). Both of these items address the *Skepticism* component, suggesting that it may also be affected by latencies related to, but peripheral to SSR. For example, a students' nature of science understanding is a variable that is different from SSR, but an individual's nature of science understanding could impact the extent to which they are able to exhibit skepticism (Table 3).

Multi-faceted Rasch analysis of items, scenarios, and testing instances

The relative contribution of each facet to students' SSR measures is detailed holistically through the multi-faceted Rasch partial credit model (Table 4). We obtained separable and reliable measures for students' ability ($\rho_{Rasch} = 0.89$, separation = 2.83) as well as difficulty of items ($\rho_{Rasch} = 0.99$, separation = 11.79), scenarios ($\rho_{Rasch} = 0.95$, separation = 4.34), and testing instances ($\rho_{Rasch} = 0.77$, separation = 1.82). Item mean squares infit values ranged from 0.88–1.19 while outfit values ranged from 0.82–1.25; both ranges indicative that the items provide useful measures of SSR (Wright et al., 1994). Figure 1 shows that item difficulty measures were evenly-dispersed along the scale (−0.61–0.60), indicating that the items provide useful information about students at multiple levels of SSR competency. However, the spread of item measures is down-shifted relative to the students' ability measures, indicating that the QuASSR provides the most information about students with lower levels of SSR.

Table 4. Rasch measures and mean squares fit indices for the QuASSR's items, scenarios, and testing instances.

	Measure	SE	Infit	Outfit
Item Facet				
Complexity1	−0.61	0.04	1.03	1.03
Complexity2	−0.44	0.04	0.95	0.92
Perspective3	−0.44	0.03	0.90	0.82
Perspective4	−0.05	0.03	0.94	0.96
Inquiry5	0.60	0.03	1.14	1.19
Inquiry7	−0.11	0.03	0.95	0.93
Inquiry9	0.11	0.03	1.01	1.00
Skepticism10	0.19	0.03	0.88	0.87
Skepticism11	0.34	0.03	0.98	0.97
Skepticism12	0.41	0.03	1.19	1.25
Scenario Facet				
Fracking	−0.05	0.02	0.97	0.95
Nothan	−0.06	0.02	0.99	0.98
Antibiotic	0.11	0.02	1.03	1.05
Instance Facet				
Pre	0.03	0.01	1.05	1.06
Post	−0.03	0.01	0.94	0.92

```
+-------------------------------------------------------------------------------+
|Measr|+Student   |-Item                      |-Scenario                |-Time       |
|-----+-----------+---------------------------+-------------------------+------------+
|   2 +           +                           +                         +            +
|     |           |                           |                         |            |
|     | .         |                           |                         |            |
|     | .         |                           |                         |            |
|     | .         |                           |                         |            |
|     | .         |                           |                         |            |
|     | *.        |                           |                         |            |
|     | *.        |                           |                         |            |
|     | *.        |                           |                         |            |
|     | *.        |                           |                         |            |
|   1 + *****.    +                           +                         +            +
|     | **.       |                           |                         |            |
|     | *****.    |                           |                         |            |
|     | ****      |                           |                         |            |
|     | *******   | Inquiry5                  |                         |            |
|     | *******.  |                           |                         |            |
|     | *******.  | Skepticism11              |                         |            |
|     | ******    | Skepticism10              |                         |            |
|     | *****     | Skepticism9               |                         |            |
|     | ********* | Inquiry8                  | Antibiotic              |            |
*   0 * ****.     *                           *                         * Post  Pre  *
|     | ****.     | Perspective4 Inquiry7     | Fracking    Nothan      |            |
|     | ***       |                           |                         |            |
|     | ***.      |                           |                         |            |
|     | ***.      | Complexity2 Perspective3  |                         |            |
|     | *.        |                           |                         |            |
|     | .         | Complexity1               |                         |            |
|     | *         |                           |                         |            |
|     | .         |                           |                         |            |
|     | .         |                           |                         |            |
|  -1 +           +                           +                         +            +
|     | .         |                           |                         |            |
|     |           |                           |                         |            |
|     |           |                           |                         |            |
|     |           |                           |                         |            |
|     |           |                           |                         |            |
|     |           |                           |                         |            |
|     |           |                           |                         |            |
|     |           |                           |                         |            |
|     |           |                           |                         |            |
|  -2 +           +                           +                         +            +
|-----+-----------+---------------------------+-------------------------+------------+
|Measr| * = 3     |-Item                      |-Scenario                |-Time       |
+-------------------------------------------------------------------------------+
```

Figure 1. Person-facet (Wright) map for the QuASSR showing the spread of student, item, scenario, and testing instance (Time) measures on the same scale.

Analysis of difficulty and mean squares fit for the scenario and testing instance facets indicated that scenario context and testing instance had small-to-negligible effects on the difficulty of the QuASSR. Along the scenario facet, we found mean squares fit indices close to 1, indicating that the scenarios are not unduly biased towards high- or low-SSR students. The difficulty spread of the scenarios was relatively small in comparison to the spread of items, corroborating the correlational and 2-faceted Rasch analyses suggesting that differential effects due to inconsistency between scenarios were likely to be relatively small in comparison to the differences in the items embedded within. We did find it interesting that while the fracking and Lake Nothan scenarios where nearly identical in difficulty, it was harder for students to express high SSR for the antibiotic scenario.

RQ2: how does student SSR, as measured by the QuASSR, change in response to SSI-based instruction?

Students' mean SSR was measured at 0.38 logits (SD = 0.58 logits) on the pre-test. This mean increased to 0.45 logits (SD = 0.69 logits) on the post-test. This gain in SSR over the course of instruction was statistically significant (Gain = 0.071 logits, SD = 0.59 logits, $t_{df=272} = 1.98$, $p = 0.049$, Cohen's D = 0.12). The statistical significance just exceeds the 95% confidence threshold (2-tailed) and the effect size is low: however, this shows the ability of the QuASSR to detect changes in SSR in the context of SSI-based instruction.

RQ3: how do the elements of SSR relate to one another in terms of a possible progression of competencies?

The items addressing the complexity of SSR are located at the bottom of the Wright map, indicating that even a student with relatively low SSR is likely to identify that the negotiation of SSI is complex. A more proficient student begins to understand that parties with different interests likely think about the issue differently. Finally, items addressing inquiry and skepticism sit at the top of the scale. This suggests that growth in the lower-level competencies like complexity may buttress growth in the higher-level competencies of inquiry and skepticism. This hypothesis (Figure 2) cannot be rejected by the data ($\chi^2_{df=2} = 2.86$, $p = 0.24$), and models the actual covariance closely (RMSEA = 0.040, CFI = 0.99, TLI = 0.97, SRMR = 0.025).

All directional relationships (Figure 2) are positive and statistically significant, and the structural equations have small-to-moderate effect sizes (Cohen, 1992). Given that the model fits the data well, the direct effects imply that when students learn to identify the complexity in SSI, this supports their ability to learn how to analyse the issue from multiple perspectives ($b = 0.27$, $SE_b = 0.056$, $r^2 = 0.073$). When students increase their ability with respect to perspective taking, then they can learn to exercise skepticism when evaluating how different parties use data to support their positions ($b = 0.31$, $SE_b = 0.055$, $r^2 = 0.096$) and that additional information and data would be helpful in identifying the most effective solutions ($b = 0.35$, $SE_b = 0.053$, $r^2 = 0.125$).

Indirect effects from complexity toward inquiry and skepticism are also significant, indicating that students' learning to identify that SSI's are complex boosts adoption of skepticism ($b = 0.084$, $SE_b = 0.023$) and understanding of the need for inquiry ($b = 0.095$, $SE_b = 0.025$). The correlation between inquiry and skepticism is small and non-significant ($b = 0.015$, $SE_b = 0.061$) after accounting for growth in perspective-taking. This indicates that after the common cause of perspective-taking is taken into account, inquiry and skepticism are largely independent of each other.

Discussion

Measurement and dimensionality of SSR

Learning science content and developing competencies for scientific practices is thought to prepare learners for dealing with complex issues in society that relate to science (Bybee, 2014). However, learning science content and practices is not enough to substantively impact the ways in which learners negotiate SSI (Romine et al., 2017; Sadler et al.,

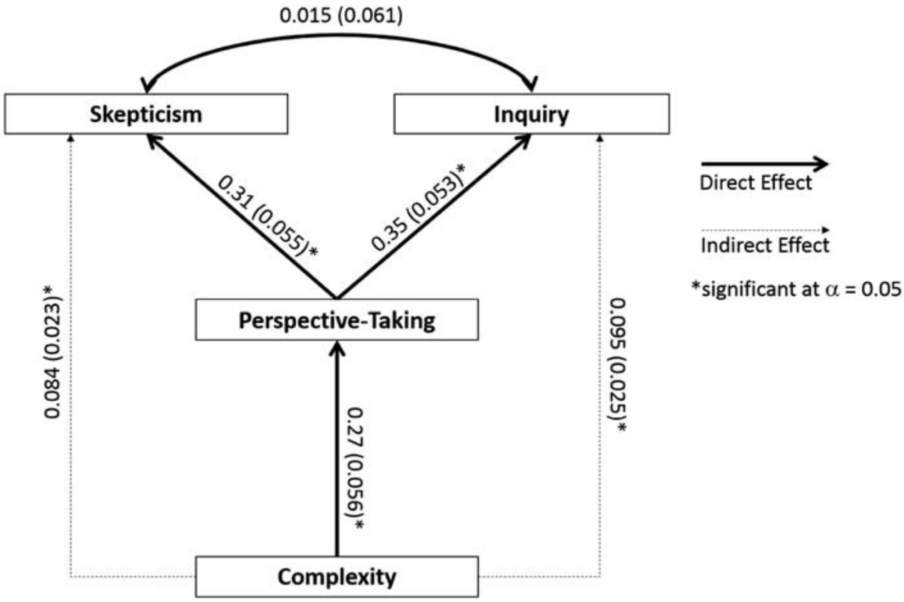

Figure 2. Standardised empirical model for the progression of SSR, suggesting that learning of lower-level SSR competencies supports learning of higher-level competencies in the context of our unit of instruction. Path coefficients (b) with standard errors (SE_b – in parentheses) are reported on the connecting arrows.

2011). While most approaches to science teaching prioritise student learning of science content and practices, the goal of SSI teaching is to support student learning of canonical science as well as the other factors necessary for the negotiation of SSI. The dearth of research in this area is due in part to limited conceptual resources for framing these kinds of investigations and even more limited tools for conducting empirical investigations. SSR was developed as a construct to address the gap in conceptual resources for this work, and the QuASSR was developed as a research instrument for empirical investigations of SSR (Romine et al., 2017; Sadler et al., 2007).

Exploration of the assumptions of unidimensionality and local independence with respect to the 2-faceted Rasch model lent insight into inconsistent parts of the QuASSR. Although local dependency of items was found to be low, we did find some departure from the assumption of unidimensionality when three scenarios were used together (Table S1 in the Online Supplementary Materials). On the pre-test, Item 9 across all scenarios loaded onto the residual dimension (loadings between 0.53 and 0.61). We observed a similar pattern with Item 11 on the post-test (loadings between 0.47 and 0.62). Both Items address the *Skepticism* component of SSR. This finding is interesting in its suggestion that not only does SSR transfer across scenarios, but the measurement behaviour of the QuASSR transfers. Given the position of *skepticism* as one of the higher-level SSR competencies, this repeatable inconsistency leads us to propose that other factors related to SSR, such as nature of science understandings and information literacy competencies, may come into play as students improve their ability to negotiate SSI's. For example, the extent to which a learner understands that conducting and reporting research can be influenced by the researchers' backgrounds, experiences and biases, that is, the subjective nature of science (Khishfe, 2014), can impact the ways in which the learner conceptualises an SSI (Sadler et al., 2004). It seems likely that ideas about this aspect of the nature of science could interact in substantive ways with the ways in which the learner demonstrates skepticism. This leads to a question that is important from both measurement and theoretical perspectives: do we treat *skepticism* as a component of SSR, or do we treat it as a skill that is separate from, but correlated with, SSR? We believe that removing *skepticism* from SSR would add unnecessary complexity and subjectivity to an already complex construct. For example, it seems likely that nature of science ideas and information literacy competencies relate to SSR, but there is not sufficient theoretical work to tease out where SSR begins and these other ideas and competencies end.

Transfer of SSR

SSR represents reasoning competencies that theoretically should transfer from learning opportunities related to one SSI to another SSI. If this holds, then two predictions follow. (1) evidence of SSR elicited in response to different issue contexts should be equivalent. (2) SSR competencies gained through an activity related to one SSI should be detectable in SSR assessments using contexts that differ from the issue explored during instruction. If SSR is an outcome that can be applied across scientific disciplines and types of issues, then adherence to this standard is an important validity criterion for assessments that measure SSR. The correlational perspective shows that measures derived from the three scenarios have high repeatability on both the pre- and post-tests. It is interesting that the consistency between the fracking and Nothan scenarios ($r > 0.9$) exceeds the consistency between these and the antibiotic scenario ($r > 0.8$). The multi-faceted Rasch model suggests that this may be due partly to the fact that it was harder for students to reason around the antibiotic issue than the fracking or water conservation issues.

The fit of the items, scenarios, and testing instances with the multi-faceted Rasch model suggests that excessive bias in terms of student ability did not play a significant role in the attenuation of reliability. However, the relative inconsistency of the antibiotic scenario needs explaining. It is conceivable that a particular scenario context could be biased in favour of students with particular backgrounds. For example, asking a class to reason around a fracking scenario in a community where a portion of the students are personally affected by the issue may result in inconsistency with other issues that the students are less passionate about. This line of reasoning may explain why the students found the antibiotic scenario harder to negotiate. Many of the students participating in this study hailed from agricultural communities which were affected directly by the debate around antibiotic use in farm animals, so it makes sense that this proximity may have increased the tendency of students to see this as a black and white issue. Indeed, educational psychologists make it clear that for students to fully engage with controversial SSI and apply objective reasoning, they must have an ability to regulate emotions as well as detect and diffuse cognitive biases that influence reasoning (Sinatra & Seyranian, 2015; Sinatra et al., 2014). Transfer may vary in relation to the direct relevance of the issue to the students' vested interests.

These implications also extend to SSI instruction. Instructional content that conflicts with students' identity or worldview may produce negative emotions that could hinder engagement in evaluating evidence (Darner, 2019). Those seeking to promote transfer of SSR need to be deliberate in choice of SSI topics, consider multiple topics that are both proximal and distal to the students' vested interests, and use tools to help recognise and manage student emotions. For example, Darner (2019) calls for instruction based on self-determination theory that could fill students' psychological needs around identity, allowing them to orient their learning towards accuracy of evidence rather than self-preservation. Other researchers call for instruction that supports emotion through self-regulated learning (Sinatra & Taasoobshirazi, 2018). While our results, along with others who have documented the role emotions play in students' construal of SSI (Heddy & Sinatra, 2013), point to a possible need for SSI instructors and researchers to create learning experiences that attend to student emotions, few within our subfield have documented explicit efforts on this front.

SSR as a progression

Both in our prior work and in the present study, we observed a progression of difficulty from identifying *complexity* to *perspective-taking*, and then to the higher-level competencies of *inquiry* and *skepticism.* This led us to explore the idea that growth in the lower-level competencies may buttress growth in the higher-level competencies. We find that a path model expressing growth in these competencies in a sequential framework (Figure 2) not only cannot be rejected by the data, but also provides high absolute and relative fit with the data. The direct path coefficients between 0.27 and 0.35 indicate large effects (Keith, 1993) and the r-square values indicate small-to-moderate effects of changes in the lower-level competencies on changes in the upper-level competencies (Cohen, 1992). Although these relationships do not necessarily suggest high predictive power, they do suggest the importance of the sequential connectedness between these competencies. But what does this tell us about how to cultivate SSR in students?

In our view, the most important aspect of the figure is the centrality of *perspective-taking.* We want students to leave our classes understanding that SSI's are complex, but this is just a first step that cannot facilitate the higher-level competencies of SSR directly. Learning experiences must emphasise the multiple perspectives around SSI's before students may be ready to engage in inquiry and skepticism. Science educators have long noted the importance of understanding other's viewpoints; for example, Ratcliffe (1997) included this as one of the key features of well-reasoned decision-making about SSI. Students often approach SSI with reductionist views that are centred in one's own position. Until students are open to multiple perspectives, they may not see SSI as nuanced and solutions as tentative. This may hinder their ability to recognise the need for information to resolve uncertainties about potential courses of action in general, or to be willing to consider specific counter-evidence (Chinn & Brewer, 1993). This suggestion aligns well with recent calls to highlight perspective-taking as a key dimension of SSI-based teaching and learning (Kahn & Zeidler, 2019)

What does it take to improve and measure SSR?

Earlier work investigating college-aged students before and after a 1-week SSI unit found no gains in SSR using the QuASSR instrument (Romine et al., 2017). In this study, we focus on a course which puts science literacy practices as the primary goal, as opposed to most undergraduate science courses that prioritise science content. While SSI was a focus of the course, SSR was not an explicit framework for course design; however, course learning goals centred on science-informed decision-making and information literacy which run parallel to SSR. Additionally, this course focused on science literacy goals throughout a 15-week semester, allowing for learning time and extensive practice, which is important for development of complex skills such as SSR (Cansiz, 2014).

Secondly, an important quality of SSR is that it is measuring competencies that are meant to transfer across SSI contexts. In this study, the three QuASSR scenarios were not used instructionally in the course, indicating that the learning gains were sophisticated enough to be considered *context transfer* or *near transfer* in Haskell's (2000) progressive stages. Toward maximising transfer, the course: (1) connected to students' lived experiences and prior knowledge, (2) required students to practice the same skills of decision-making and finding and applying information multiple times in different contexts allowing them to generalise these skills, and (3) devoted time to meta-level discussion about the role of science and values in SSI, the reasons for complexity in decision-making about SSI, and the values and priorities of multiple stakeholders. All of these course characteristics are known to increase the probability of transfer of skills to new contexts (Bransford et al., 2000).

Limitations and conclusions

We conclude this article by acknowledging limitations of this study, and opportunities these generate for future work. Although we found transfer of SSR between scenarios, the findings are still limited by the structure of the assessment and instruction. We still do not have evidence that students expressing high SSR on the QuASSR are able to apply this generally to everyday interpretation of SSI like those they see in the news. Further, we do not have evidence that a higher QuASSR measure makes a student more likely to engage in SSR during active social interactions such as debates. Looking at these issues around transfer to more authentic settings poses an interesting line of research.

The context experienced by the students in this study also poses a limitation on the extent to which we can generalise findings particularly with respect to skepticism. The framing of SSR and skepticism is consistent with constructivist perspectives which hold that knowledge does not exist independent of the learner (Kuhn, 1970). Further, science is value-laden and the production of scientific knowledge is necessarily dependent on socio-cultural norms and influences (Kuhn, 1970). While constructivism is appealing from a philosophical perspective, it can present challenges when teaching with SSI by obscuring how one idea may be more viable than another, encouraging relativistic thinking (Phillips, 1995). These challenges are particularly relevant in the current post-truth era (McIntyre, 2018) where many people choose 'facts' to fit their beliefs. In response, science instructors may tend toward holding up scientific evidence as a source that holds more objectivity than other kinds of information generated from corporations, political entities or commercial media with targeted audiences.

The structured decision-making process used in the course from which we collected data positions information in a more positivist light than the skepticism dimension of

SSR, in that it is meant to be neutral, objective and predict consequences of actions. Therefore, the course content may have been aligned with lower-level SSR learning goals as measured by the QuASSR. In other words, given the constructivist positioning of SSR, and use of SSI's in the classroom in general, nesting these approaches within a course focused on using science information in structured decision-making may produce different results than if these approaches were nested within a course that actively promoted more constructivist views. It would be interesting to compare similar interventions in courses adopting different epistemologies to explore effects on how students respond to the QuASSR and its ability to measure changes in SSR over time.

Next, use of the QuASSR for middle school, high school, and more senior science students has yet to be explored. Within the epistemological constraints outlined above, our data show promise with regards to the ability of the QuASSR to capture the transfer of SSR across different scenarios. With respect to measurement of change, modelling SSR as a progression of competencies shows the centrality of perspective-taking that needs to be considered as we develop interventions aimed at helping our students reason around SSI. Measurement of SSR and understanding SSR as a construct go hand-in-hand, and using one to understand the other will be a continuous back-and-forth process for some time to come. However, we are in a strong position to start posing testable theoretical questions and experiments regarding the role experience plays in how we learn to negotiate SSI.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

William Lee Romine http://orcid.org/0000-0002-0386-1688
Troy D. Sadler http://orcid.org/0000-0002-9401-0300
Jenny M. Dauer http://orcid.org/0000-0002-1373-8851

References

Alred, A. R., & Dauer, J. M. (2020). Understanding factors related to undergraduate student decision-making about a complex socio-scientific issue: Mountain Lion management. *EURASIA Journal of Mathematics, Science and Technology Education*, *16*(2), em1821. https://doi.org/10.29333/ejmste/113757

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Psychology Press.

Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how? *CBE—Life Sciences Education*, *15*(4), rm4. https://doi.org/10.1187/cbe.16-04-0148

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn* (Vol. 11). National Academy Press.

Bybee, R. W. (2014). NGSS and the next generation of science teachers. *Journal of Science Teacher Education*, *25*(2), 211–221. https://doi.org/10.1007/s10972-014-9381-4

Cansiz, N. (2014). *Developing pre-service science teachers' socio-scientific reasoning through socio-scientific issues focused course* [Doctoral dissertation]. Middle East Technical University.

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), 1–49. https://doi.org/10.3102/00346543063001001

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Darner, R. (2019). How can educators confront science denial? *Educational Researcher*, *48*(4), 229–238. https://doi.org/10.3102/0013189X19849415

Dauer, J. M., & Forbes, C. T. (2016). Making decisions about complex socioscientific issues: A multidisciplinary science course. *Science Education & Civic Engagement: An International Journal*, *8*(2), 5–12.

Dauer, J. M., Lute, M. L., & Straka, O. (2017). Indicators of informal and formal decision-making about a socioscientific issue. *International Journal of Education in Mathematics, Science and Technology*, *5*(2), 124–138. https://doi.org/10.18404/ijemst.05787

Dauer, J. M., Sorensen, A. E., & Jimenez, P. C. (in press). Using a structured decision-making tool in the classroom to promote information literacy in the context of decision-making. *Journal of College Science Teaching*.

Dawson, V. M., & Venville, G. (2010). Teaching strategies for developing students' argumentation skills about socioscientific issues in high school genetics. *Research in Science Education*, *40*(2), 133–148. https://doi.org/10.1007/s11165-008-9104-y

DeBoer, G. E. (2000). Scientific literacy: Another look at its historical and contemporary meanings and its relationship to science education reform. *Journal of Research in Science Teaching*, *37*(6), 582–601. https://doi.org/10.1002/1098-2736(200008)37:6<582::AID-TEA5>3.0.CO;2-L

Eggert, S., & Bögeholz, S. (2010). Students' use of decision-making strategies with regard to socioscientific issues: An application of the Rasch partial credit model. *Science Education*, *94*(2), 230–258.

Eggert, S., Ostermeyer, F., Hasselhorn, M., & Bögeholz, S. (2013). Socioscientific decision making in the science classroom: The effect of embedded metacognitive instructions on students' learning outcomes. *Education Research International*, *2013*, 1–12. https://doi.org/10.1155/2013/309894

Evagorou, M., & Osborne, J. (2013). Exploring young students' collaborative argumentation within a socioscientific issue. *Journal of Research in Science Teaching*, *50*(2), 209–237. https://doi.org/10.1002/tea.21076

Fleming, R. (1986a). Adolescent reasoning in socio-scientific issues: II. Nonsocial cognition. *Journal of Research in Science Teaching*, *23*(8), 689–698. https://doi.org/10.1002/tea.3660230804

Fleming, R. (1986b). Adolescent reasoning in socio-scientific issues, part I: Social cognition. *Journal of Research in Science Teaching*, *23*(8), 677–687. https://doi.org/10.1002/tea.3660230803

Haskell, R. E. (2000). *Transfer of learning: Cognition and instruction*. Elsevier.

Heddy, B. C., & Sinatra, G. M. (2013). Transforming misconceptions: Using transformative experience to promote positive affect and conceptual change in students learning about biological evolution. *Science Education*, *97*(5), 723–744. https://doi.org/10.1002/sce.21072

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Kahn, S., & Zeidler, D. L. (2019). A conceptual analysis of perspective taking in support of socioscientific reasoning. *Science & Education*, *28*(6–7), 605–638. https://doi.org/10.1007/s11191-019-00044-2

Karahan, E., & Roehrig, G. (2016). Use of web 2.0 technologies to enhance learning experiences in alternative school settings. *International Journal of Education in Mathematics Science and Technology*, *4*(4), 272–283. https://doi.org/10.18404/ijemst.32930

Ke, L., Sadler, T. D., Zangori, L., & Friedrichsen, P. J. (in press). Students' perceptions of socio-scientific issue-based learning and their appropriation of epistemic tools for systems thinking. *International Journal of Science Education*. https://doi.org/10.1080/09500693.2020.1759843

Keith, T. Z. (1993). *Causal influences on school learning*. In H. J. Walberg (Ed.), *Analytic methods for educational productivity* (pp. 21–47). JAI Press.

Khishfe, R. (2014). Argumentation instruction in the context of socioscientific issues: An effect on student learning and transfer. *International Journal of Science Education*, *36*(6), 974–1016. https://doi.org/10.1080/09500693.2013.832004

Khishfe, R., Alshaya, F., BauJaoude, S., Mansour, N., & Alrudiyan, K. (2017). Students' understandings of nature of science and their arguments in the context of four socio-scientific issues. *International Journal of Science Education*, *39*(3), 299–334. https://doi.org/10.1080/09500693.2017.1280741

Kinslow, A. T. (2018). *The development and implementation of a heuristic for teaching reflective scientific skepticism within a socio-scientific issue instructional framework* [Doctoral dissertation]. University of Missouri–Columbia.

Kinslow, A. T., Sadler, T. D., & Nguyen, H. T. (2019). Socio-scientific reasoning and environmental literacy in a field-based ecology class. *Environmental Education Research*, *25*(3), 388–410. https://doi.org/10.1080/13504622.2018.1442418

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford Publications.

Kuhn, T. S. (1970). *The structure of scientific revolutions*. University of Chicago Press.

Lederman, N. G., Antink, A., & Bartos, S. (2014). Nature of science, scientific inquiry, and socio-scientific issues arising from genetics: A pathway to developing a scientifically literate citizenry. *Science & Education*, *23*(2), 285–302. https://doi.org/10.1007/s11191-012-9503-3

Lee, H., Yoo, J., Choi, K., Kim, S. W., Krajcik, J., Herman, B. C., & Zeidler, D. L. (2013). Socioscientific issues as a vehicle for promoting character and values for global citizens. *International Journal of Science Education*, *35*(12), 2079–2113. https://doi.org/10.1080/09500693.2012.749546

Linacre, J. M. (2006). *WINSTEPS Rasch measurement computer program*. WINSTEPS.com.

Linacre, J. M. (2009). *Facets Rasch measurement computer program (version 3.65. 0)*. WINSTEPS.com.

Linacre, J. M., & Tennant, A. (2009). More about critical eigenvalue sizes in standardized-residual principal components analysis (PCA). *Rasch Measurement Transactions*, *23*(3), 1228.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174. https://doi.org/10.1007/BF02296272

McIntyre, L. (2018). *Post-truth*. MIT Press.

Morin, O., Simonneaux, L., Simonneaux, J., & Tytler, R. (2013). Digital technology to support students' socioscientific reasoning about environmental issues. *Journal of Biological Education*, *47*(3), 157–165. https://doi.org/10.1080/00219266.2013.821748

Morin, O., Simonneaux, L., Simonneaux, J., Tytler, R., & Barraza, L. (2014). Developing and using an S3R model to analyze reasoning in web-based cross-national exchanges on sustainability. *Science Education*, *98*(3), 517–542. https://doi.org/10.1002/sce.21113

Muthén, L. K., & Muthén, B. O. (2007). *Mplus user's guide* (6th ed.). Muthén &Muthén.

Peel, A., Zangori, L., Friedrichsen, P., Hayes, E., & Sadler, T. (2019). Students' model-based explanations about natural selection and antibiotic resistance through socio-scientific issues based learning. *International Journal of Science Education*, *41*(4), 510–532. https://doi.org/10.1080/09500693.2018.1564084

Phillips, D. Z. (1995). Where are the gods now? In C. I. Accetti (Ed.), *Relativism and religion* (pp. 1–15). Palgrave Macmillan.

Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, *15*(2), 020111. https://doi.org/10.1103/PhysRevPhysEducRes.15.020111

Raiche, G. (2005). Critical eigenvalue sizes in standardized residual principle components analysis. *Rasch Measurement Transactions*, *19*, 1012.

Ratcliffe, M. (1997). Pupil decision-making about socio-scientific issues within the science curriculum. *International Journal of Science Education*, *19*(2), 167–182. https://doi.org/10.1080/0950069970190203

Roberts, D. A., & Bybee, R. W. (2014). Scientific literacy, science literacy, and science education. In D. A. Roberts & R. W. Bybee (Eds.), *Handbook of research on science education* (Vol. II, pp. 559–572). Routledge.

Romine, W. L., Sadler, T. D., & Kinslow, A. T. (2017). Assessment of scientific literacy: Development and validation of the Quantitative assessment of socio-scientific reasoning (QuASSR). *Journal of Research in Science Teaching*, *54*(2), 274–295. https://doi.org/10.1002/tea.21368

Rudolph, J. L., & Horibe, S. (2016). What do we mean by science education for civic engagement? *Journal of Research in Science Teaching*, *53*(6), 805–820. https://doi.org/10.1002/tea.21303

Sadler, T. D., Barab, S. A., & Scott, B. (2007). What do students gain by engaging in socioscientific inquiry? *Research in Science Education*, *37*(4), 371–391. https://doi.org/10.1007/s11165-006-9030-9

Sadler, T. D., Chambers, F. W., & Zeidler, D. L. (2004). Student conceptualizations of the nature of science in response to a socioscientific issue. *International Journal of Science Education*, *26*(4), 387–409. https://doi.org/10.1080/0950069032000119456

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* (pp. 45–77). Springer.

Sadler, T. D., Romine, W. L., & Topçu, M. S. (2016). Learning science content through socio-scientific issues-based instruction: A multi-level assessment study. *International Journal of Science Education*, *38*(10), 1622–1635. https://doi.org/10.1080/09500693.2016.1204481

Sadler, T. D., & Zeidler, D. L. (2004). The morality of socioscientific issues: Construal and resolution of genetic engineering dilemmas. *Science Education*, *88*(1), 4–27. https://doi.org/10.1002/sce.10101

Sadler, T. D., & Zeidler, D. L. (2005). Patterns of informal reasoning in the context of socioscientific decision making. *Journal of Research in Science Teaching*, *42*(1), 112–138. https://doi.org/10.1002/tea.20042

Simonneaux, L., & Simonneaux, J. (2009). Students' socio-scientific reasoning on controversies from the viewpoint of education for sustainable development. *Cultural Studies of Science Education*, *4*(3), 657–687. https://doi.org/10.1007/s11422-008-9141-x

Sinatra, G. M., Broughton, S. H., & Lombardi, D. (2014). Emotions in science education. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 415–436). Routledge.

Sinatra, G. M., & Seyranian, V. (2015). Warm change about hot topics: The role of motivation and emotion in attitude and conceptual change about controversial science topics. In P. Alexander & P. Winne (Eds.), *Handbook of educational psychology* (pp. 259–270). Routledge.

Sinatra, G. M., & Taasoobshirazi, G. (2018). The self-regulation of learning and conceptual change in science: Research, theory, and educational applications. In D. H. Schunk & J. A. Greene (Eds.), *Educational psychology handbook series. Handbook of self-regulation of learning and performance* (pp. 153–165). Routledge/Taylor & Francis Group.

Sjöström, J., & Eilks, I. (2018). Reconsidering different visions of scientific literacy and science education based on the concept of bildung. In Y. Dori, Z. Mevarech, D. Baker (Eds.), *Cognition, metacognition, and culture in STEM education – learning, teaching and assessment* (Part of the Innovations in Science Education and Technology book series, Vol. 24, pp. 65–88). Springer. https://doi.org/10.1007/978-3-319-66659-4_4

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, *15*(1), 72–101. https://doi.org/10.2307/1412159

Todd, A., & Romine, W. L. (2017). Empirical validation of a modern genetics progression web for college biology students. *International Journal of Science Education*, *39*(4), 488–505. https://doi.org/10.1080/09500693.2017.1296207

Wilson, M. (2009a). Measuring progressions: Assessment structures underlying a learning progression. *Journal of Research in Science Teaching*, *46*(6), 716–730. https://doi.org/10.1002/tea.20318

Wilson, M. (2009b, June 24–26). *Structured constructs models (SCM): A family of statistical models related to learning progressions*. Learning Progressions in Science (LeaPS) conference, Iowa City, IA.

Womack, A. J., Sadler, T. D., & Oertli, T. (2019, March 31–April 3). *Automated scoring of a constructed response vision II scientific literacy assessment*. Paper presented at the NARST Annual Meeting, Baltimore, MD.
Wright, B. D., Linacre, J. M., Gustafson, J. E., & Martin-Loff, P. (1994). Reasonable mean square fit values. *Rasch Measurement Transactions*, *8*(3), 370.
Wu, Y. T., & Tsai, C. C. (2007). High school students' informal reasoning on a socio-scientific issue: Qualitative and quantitative analyses. *International Journal of Science Education*, *29*(9), 1163–1187. https://doi.org/10.1080/09500690601083375
Yang, F.-Y., & Anderson, O. R. (2003). Senior high school students' preference and reasoning modes about nuclear energy use. *International Journal of Science Education*, *25*(2), 221–244. https://doi.org/10.1080/09500690210126739
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*(1), 74–101. doi:10.1002/tea.20281