Automating the Detection of Reflection-on-Action

Jenny Saucerman
Web Courseworks
Madison, Wisconsin

A. R. Ruis
Wisconsin Center for Education Research
University of Wisconsin–Madison

David Williamson Shaffer
Department of Educational Psychology
University of Wisconsin–Madison
dws@education.wisc.edu

ABSTRACT: Learning to solve *complex problems* — problems whose solutions require the application of more than basic facts and skills — is critical to meaningful participation in the economic, social, and cultural life of the digital age. In this paper, we use a theoretical understanding of how professionals use reflection-in-action to solve complex problems to investigate how students learn this critical 21st-century skill and how we can develop and automate learning analytic techniques to assess that learning. We present a preliminary study examining the automated detection of reflective discourse during collaborative, complex problem solving. We analyze student reflection-on-action in a virtual learning environment, focusing on both reflection in individual discourse and collaborative reflection among students. Our results suggest that it is possible to detect student reflection on complex problems in virtual learning environments, but that different models may be appropriate depending on students' prior domain experience.

Keywords: Student reflection, reflection-on-action, epistemic frame theory, virtual internship

In the last three decades, the economies of many developed countries have shifted from the production of goods to investment in human knowledge (Powell & Snellman, 2004). Jobs used to require what Murnane and Levy (1993) refer to as *basic skills*, those needed to produce commodities. Because most production is now outsourced to temporary workers or to workers in less-developed countries (Friedman, 2006), mastery of basic skills is no longer sufficient to obtain high-quality employment (Ruckelshaus & Leberstein, 2014). Despite this shift in economic and social priorities, the standard school curriculum continues to emphasize acquisition of basic knowledge and skills.

To succeed in the knowledge economy, young people need to learn how to frame, investigate, and solve problems that require more than just basic facts and skills. Learning to solve *complex problems* — ill-

formed problems whose solutions require more than the application of basic knowledge and skills or routine procedures — is a critical component of equipping young people with the ability to participate meaningfully in the economic, social, and cultural life of the digital age (Autor, Levy, & Murnane, 2003a, 2003b; Levy & Murnane, 2004).

In this paper, we adopt Schön's (1983) theoretical perspective on how professionals solve complex problems — a process that he describes as *reflection-in-action* — to explore (a) how students can learn this critical 21st-century skill and (b) how we can assess that learning. Specifically, we focus on how to support this kind of learning in immersive virtual learning environments by exploring a learning analytic technique for automating the assessment of reflection-on-action during collaborative problem-solving activities.

Reflection-in-action is the ability to adapt the solutions developed for past problems to some current problem. In other words, it is an element of mastery that enables experts to draw on their experience to analyze and solve new problems. Schön argues that novices, who do not yet have the experience necessary for reflection-*in*-action, learn this skill through reflection-*on*-action: discussing their attempts to investigate and solve complex problems with each other and with mentors, or more knowledgeable others who help them understand how to analyze and interpret their actions in the domain.

In what follows, we present a preliminary study examining the automated detection of reflective discourse during complex problem solving. We begin by examining the conceptual underpinnings of reflection, and specifically of reflection-on-action. We then apply the resulting framework to analyze student reflection in an immersive virtual learning environment, focusing on both individual and collaborative reflection. Our results suggest that it is possible to detect student reflection on complex problems in virtual learning environments, but that different models may be appropriate depending on students' prior domain experience.

1   THEORY

Complex problems can be distinguished from non-complex problems by the fact that they do not have *well-formed solutions*. For example, as Schön (1992) argues, when a civil engineer considers a road construction problem, he or she cannot solve it by applying "locational techniques or decision theory"; rather, "he [or she] confronts a complex and ill-defined situation in which geographic, financial, economic, and political factors are usually mixed up together" (p. 6). In other words, there is no single solution for the civil engineer's problem because the implementation of a typical solution from the civil engineering domain ("locational techniques or decision theory") would not sufficiently account for the ways in which geography, finances, the economy, the environment, or politics might affect the problem's solution. This complex problem requires a different problem-solving technique.

In many cases, real-world problems are not well formed and instead appear messy and indeterminate. Howard (1983) argues that ill-defined problems have vague goals and that the kind of information

relevant to the problem is often unclear. Wood (1983) characterizes ill-defined problems as having components that are either unknown or not known with any degree of confidence; Kitchener (1983) suggests that such problems either have multiple solution paths or none at all. Building on this foundational work, Spector, Merrill, Elen, and Bishop (2013) argue that complex problems are characterized by a lack of consensual agreement on the appropriate solution. Graesser et al. (in press) further argue that complex problems can have families of solutions. Thus, complex problems cannot be solved algorithmically.

Practitioners who work in complex domains, then, cannot solve problems either by referring to some pre-existing procedure or by directly applying a method used in some previous problem. Instead, solutions are found through an iterative process of trial and observation. But these trials are not simply random guesswork. Schön (1995) argues that when professionals encounter novel problems, they attempt to solve them by running *informed experiments* performed and evaluated in real time as the problem is addressed (Schön, 1984).

This ability to perform and evaluate informed experiments in real time is a critical — perhaps *the* critical — feature of work in a complex domain. Schön calls this process *reflection-in-action*: the skill that "permits experimenters to carry out on-the-spot experiments that generate new data in the field while the intervention is still underway" (p. 26). Reflection-in-action takes place as experts in a domain (a) identify similarities between novel problems and past problems, (b) adapt the solutions from those past problems based on their understanding of the current problem, and then (c) evaluate the results of applying the adapted solution to the problem at hand, repeating these steps as needed until the problem is solved (Schön, 1983).

Although past solutions cannot be *directly* applied to a new complex problem, Schön argues that the process of reflection-in-action depends on having a professional repertoire of experiences gained by having previously solved similar problems. Professionals are able to identify the attributes of the novel problem that are both similar and dissimilar to problems that already exist in their professional repertoires. They can then make informed decisions regarding potential solutions and adjust those solutions as necessary.

This poses an issue for people who are learning to solve complex problems, including apprentices, interns, and students. Novices have few of the domain-relevant experiences necessary for reflection-in-action and little understanding of how to interpret the experiences they do have in domain-relevant terms. If reflection-in-action is a series of informed experiments performed and evaluated in real time as the problem is being addressed, then novices cannot reflect-in-action because they do not have the necessary experience in the domain to make informed decisions. They are capable only of relatively *uninformed experiments* that must be interpreted *outside* of the problem-solving process itself with the help of more experienced mentors.

Schön refers to this process of relatively uninformed experimentation and interpretation outside the problem-solving process as *reflection-on-action*. Although novices in a domain do not have an extensive professional repertoire from which to draw potential solutions, they can learn to address complex issues in a domain by solving problems and then talking about their solutions — what worked, what did not work, and why — with mentors, who are more knowledgeable others in the domain (Schön, 1987).[1]

Reflection-on-action is thus a critical process for learning to solve complex problems through reflection-in-action.

1.1 Development of Reflection-on-Action

To understand what constitutes reflection-on-action and how to assess it, we first need to understand what it means to take action in the context of complex problem solving. Brown, Collins, and Duguid (1989) argue that in any domain there are routine behaviours that practitioners use to solve problems. Lave (1988), in turn, argues that one way novices develop expertise is by participating in some of the behaviours of the domain that practitioners use. Critically, however, Lave argues that these individual behaviours cannot be performed without what she calls a *conceptual model*. That is, novices cannot solve problems in a domain without knowing how to interpret their actions in the context of the domain.

Professionals have a particular way of looking at their actions: domain-specific interpretations of the actions performed in the context of the practice. For example, Goodwin (1994) argues that even though a farmer and an archeologist may look at the same mound of soil, they will notice different phenomena occurring within it: that is, they pay attention to the attributes of the soil that their respective domains consider to be important. Goodwin refers to this domain-specific interpretation as *professional vision*. Although the farmer and the archeologist are both engaging in the same action — evaluating soil — they have very different interpretations of that action: where the farmer sees the potential for nourishing crops, the archeologist sees the impact of structural decay.

These actions and their corresponding interpretations are what Novak and Cañas (2006) call a *concept*. They argue that when members of a practice believe that a particular concept occurs frequently, it is associated with a specific word or words. These labels for concepts, in turn, form a shorthand in discourse that allows members of the community to quickly reference them. In other words, practitioners use labels (specific words or phrases) to refer to significant concepts (actions and their associated interpretations) when they discuss problem solving in a domain.

---

[1] Some scholars have raised concerns that there may be little distinction between reflection-on-action and reflection-in-action (e.g., Eraut, 1994; Usher & Bryant, 1997). Reflection-on-action differs from reflection-in-action in that it only occurs *after* uninformed experimentation. According to Schön (1984), when novices practice reflection-on-action consistently, they become more and more able to practice reflection-in-action.

Thus, novices need to learn how to interpret action in terms of domain-relevant labels. When novices first enter a domain, they do not know these labels. The only way novices begin to understand these domain-relevant labels is by having experiences in the domain and, with the help of a more knowledgeable other, learning how to interpret those experiences using the language of the domain. As novices practice interpreting their experiences in the domain, more knowledgeable mentors provide feedback by labelling those experiences in terms of domain-relevant concepts. Over time, novices begin to identify concepts (actions and their associated interpretations) and refer to them by the appropriate labels (see Figure 1).

However, learning isolated concepts (and their associated labels) may only be the first step. Weeber, Klein, de Jong-van den Berg, and Vos (2001) suggest that new knowledge in a domain is created only when a domain-relevant connection is made between two pieces of indirectly related information. Novak and Cañas (2006) similarly suggest that expertise in a domain requires not only awareness of concepts and labels, but also understanding the relationships between concepts. They refer to these relationships as propositions: "statements about some object or event in the universe, either naturally occurring or constructed . . . [that] contain two or more concepts connected with other words to form a meaningful statement" (p. 1). It is therefore important for novices to understand not only the concepts that are important in a domain, but also the relevant connections between those important concepts.

Shaffer (2006; 2012) similarly argues that propositions (i.e., connections between pairs of concepts) are a critical element of complex thinking. He suggests that professionals see the world in an *epistemic frame*: a domain-specific configuration of connections among concepts that systematically links (a) skills (the things that a person does); (b) knowledge (the understandings that a person has); (c) values (the beliefs that a person has); (d) identity (the way a person sees him or herself); and (e) epistemology (the warrants that a person uses to justify decisions and actions). From this perspective, farmers do not only notice different phenomena of interest in soil than archaeologists do; they have different epistemic frames, different ways of thinking, acting, and making and justifying decisions.

In order to develop the ability to solve complex problems through reflection-on-action, novices need to learn not only how actions are interpreted and discussed in the domain, but also how these key concepts are systematically related to one another. The goal of reflection-on-action — the process that creates reflection-in-action — is to help novices make two different but related kinds of connections. The first kind involves *action-to-interpretation connections*, or concepts, which link actions performed in a domain to the interpretation of those particular actions in the domain. The second kind involves *concept-to-concept connections*, which build on the first and link one action-to-interpretation pairing to another (see Figure 1). Learning to make the first kind of connection, action-to-interpretation, develops a novice's professional vision and establishes a novice's ability to talk about the domain in the way experts do. The second kind, concept-to-concept, develops a novice's epistemic frame and establishes a novice's ability to think, act, and make and justify their decisions appropriately in the domain.
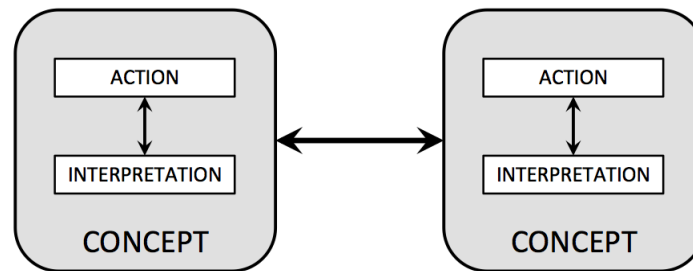
Figure 1. Reflection-on-action as a process involving relevant action-to-interpretation and concept-to-concept connections.

1.2   Assessment of Reflection-on-Action

In order to assess reflection-on-action, we need to detect both the novice's development of professional vision (action-to-interpretation connections) and her/his development of an epistemic frame (concept-to-concept connections). Identifying connections indicative of professional vision requires determining whether novices are using appropriate labels for the domain interpretations of their specific actions. One way to accomplish this is to identify the domain concepts by their labels (the name given to specific actions and their corresponding interpretations) in the novices' discourse. Previous studies of domain-specific discourse have operationalized important domain labels by identifying simple keywords and complex character string matching (Arastoopour, Chesler, & Shaffer, 2014; Califf, & Mooney, 2003). In the context of reflection-on-action, we need to identify the labels relevant only to the specific action to be reflected upon. For example, the appropriate labels during reflective discourse in the farming domain will differ depending on whether the novice farmers are examining soil or tending to a sick animal.

Velardi, Fabriani, and Missikoff (2001) argue that domain-specific ontologies, or ways of thinking, can be captured by identifying and defining the concepts and relationships that characterize a domain. To do this, they use text-mining tools on documents created by domain experts to discover labels that are potentially useful identifiers for important domain concepts. Thus, complex character string matching can be used to identify domain-relevant labels (and hence, action-to-interpretation connections) in discourse.

The second key component of assessing good reflection-on-action is the identification of connections between those concepts.) A number of researchers (e.g., Chesler et al., 2015; Dorogovtsev & Mendes, 2003; Landauer, McNamara, Dennis, & Kintsch, 2007; Lund & Burgess, 1996; Siebert-Evenstone, Arastoopour, Collier, Swiecki, Ruis, & Shaffer, 2016) argue that connections between domain concepts can be detected when they are all present within a given segment of data, or through *co-occurrences*. I Cancho & Solé (2001) argue that co-occurrences between domain concepts in discourse are significant because they are not simply the result of a known frequency of word distribution: they do, in fact, have meaning, especially when they co-occur frequently (Newman, 2004). These co-occurrences between

domain concepts are therefore likely to occur more frequently than chance co-occurrence would explain.

Shaffer et al. (2009) argue that the associative structure of conceptual connections can be modelled by identifying co-occurrences of relevant concepts in close temporal proximity. One possible operationalization of relevant concept-to-concept connections, then, is to identify relevant co-occurrences between concepts. This method has been used in previous research to describe the way professionals interpret their own actions within their respective domains and how that interpretation develops from reflection-on-action in professional practice (e.g., Hatfield, Shaffer, Bagley, Nulty, & Nash, 2008; Svarovsky & Shaffer, 2006). This work suggests that novices who practice reflection-on-action with more knowledgeable others start to exhibit the same co-occurrences — the same relevant connections between domain concepts — in discourse as experts (Nash & Shaffer, 2013). This imitation of relevant co-occurrences between domain concepts, they argue, is indicative of increased expertise in the domain.

The identification of the domain-relevant labels in discourse through the use of regular expression matching may therefore provide a good operationalization of the action-to-interpretation connections novices need to learn in order to develop their professional visions. Additionally, relevant co-occurrences between these labels may provide a good model of the relevant concept-to-concept connections novices need to make in order to develop their epistemic frames, a process that can be fully automated.

Automating the detection of reflection-on-action is thus a simpler problem than automating the detection of reflection more broadly in learning analytics. First, although students are learning to solve complex problems, reflection-on-action typically occurs in well-defined contexts, in which students reflect on one part of the problem-solving process with the guidance of a more knowledgeable other. Because the goal of the reflective activity is to help students interpret actions and make the connections indicative of professional practice in a specific domain, the detection problem is significantly constrained by the context. Because of this constraint, automated detection algorithms need not be as complex as those designed for more general contexts, such as the technique developed by Ullmann, Wild, and Scott (2012) to detect reflective writing in blogs.

Second, the nuance and contextualization critical to many learning analytic techniques designed for reflective language, such as those developed to detect student attitudes, are not as important in contexts where the goal of the reflective activity is to learn to frame, investigate, and solve problems in a domain the way professionals do. For example, Gibson and Kitto (2015) argue that fully automated coding processes are ineffective for identifying complex linguistic devices, such as sarcasm or personal satisfaction with progress toward a goal. However, detection of reflection-on-action, as we have defined it here, does not require sensitivity to such linguistic elements.

In what follows, we explore a method for automating the detection of reflection-on-action in the context of a virtual internship.

1.3   The Unit of Analysis for Assessing Reflection-on-Action

To accurately assess novices' ability to reflect-on-action, we must understand the process by which reflection-on-action develops. Much of the research on how novices learn how to reflect on their actions has been conceptualized in terms of *metacognition*: the process with which people monitor, assess, and modify their own thoughts and behaviours (Kim, Park, Moore, & Varma, 2013). In this sense, reflection-on-action is a critical form of metacognition (e.g., Hacker, Dunlosky, & Graesser, 2009; Desautel, 2009; Grant, 2001; Fogarty, 1994). Although scholars theorize different relationships between reflection and metacognition (e.g., Gama, 2004; McAleese, 1998), the literature on the development of metacognition provides a useful frame for understanding reflection, and thus the development of reflection-on-action.

One key finding from this literature is that interaction with others plays a key role in the development of metacognition (e.g., Miller & Geraci, 2011; Coutinho, Wiemer-Hastings, Skowronski, & Britt, 2005; Veenman, Van Hout-Wolters, & Afflerbach, 2006). Kim and colleagues (2013), for example, argue that novices in a domain are faced with a metacognitive paradox: metacognition is only possible when people know that they need to be conscious of their thoughts and behaviours, and the only way to learn to be aware of thoughts and behaviours is to first recognize that you are not aware of them, which cannot happen in isolation. This can only be resolved, Kim and colleagues suggest, by having someone else draw attention to the metacognitive process, hence the critical role of interaction with others in the development of metacognitive ability. In particular, Kim and colleagues argue that peers are important sources of interaction in the development of an individual's ability to think metacognitively. They argue that metacognition (and the development of metacognition) occurs not only at the level of the individual, but also at the level of the group. Because of this, peers are able to *scaffold* each others' capabilities within a domain, even when they are still in the process of development (see also Xun & Land, 2004).

According to Wood, Bruner, and Ross (1976), scaffolds reduce the complexity of a task that is initially beyond what a novice can accomplish in a domain. Peers are able to act as metacognitive scaffolds by expressing other perspectives on the same complex problem. Campione, Shapiro, and Brown (1995), for example, argue that effective learning environments can facilitate development of expertise by encouraging novices to explore different aspects of a topic that interests them. In this scenario, no single student has complete expertise. Rather, they specialize in one part of the content. Learners in such an environment engage in what Palinscar and Brown (1984) call reciprocal teaching. When novices engage in reciprocal teaching, they share the knowledge they gained about the different aspects of the topic they studied and learn about a different aspect of that topic from other students. In this form of scaffolding, Campione and colleagues (1995) argue, teamwork arises from the pooling of varieties of expertise.

1.4 Data Segmentation

This sharing of differential expertise may play an important role in the assessment of reflection-on-action. In particular, a critical component of good reflection-on-action is the ability to identify relevant concept-to-concept connections — that is, connections relevant to the action being discussed. Because peers are often able to fill in missing knowledge gaps for other members of their peer groups, reflection-on-action may need to be assessed both at the level of the individual — the connections among concepts novices make themselves — and at the level of the group. Or, it may be that assessing individual students' ability to reflect-on-action is more appropriate for more advanced students in the domain and assessing a group's ability to reflect-on-action may be more appropriate for novice students. The possibility that sharing of expertise plays a role in reflection-on-action suggests that data segmentation is a critical concern in modelling the presence of these connections — that is, we need to specify the range in the data over which we will measure co-occurrences of action-relevant labels.

In the context of discourse analysis, data segmentation refers to the identification of units of analysis within a given set of discourse data. For example, analysis might take place either at the level of the word, the sentence, or the paragraph. An analysis might also consider an utterance to be the unit of analysis (individual students) or perhaps an entire class discussion (groups of students). Rupp, Gushta, Mislevy, and Shaffer (2010) argue that these segmentation boundaries have significant consequences for results of discourse analyses. Consider the following example:

> **Student 1**: My stakeholder cared about **birds**.
> **Student 2**: They also cared about **housing**.

In this example, we might care about, say, the co-occurrence of "birds" and "housing." If we define the unit of analysis to be at the level of an utterance (one single turn of talk in discourse), we might argue that the co-occurrence is not present, as Student 1 only talked about birds, and Student 2 only talked about housing. But if we were to define the unit of analysis to be at the level of the entire discussion, we might argue that the co-occurrence *is* present, as Student 1 talked about birds, and Student 2 talked about housing. The level at which data is segmented may therefore greatly affect the assessment of novices' reflection-on-action, so this must be considered when developing an automated assessment.

1.5 Assessing Reflection-on-Action in a Virtual Internship

In this paper, we examine how to assess students' ability to reflect-on-action from two perspectives. We argue that appropriate identification of novices' reflection-on-action will potentially depend on the extent to which:

1. Students can construct concepts — that is, action-to-interpretation connections — and also connect concepts to one another. In other words, to what extent have students developed their professional visions and epistemic frames, respectively?

2. Students rely on their peers to make relevant action-to-interpretation connections and relevant concept-to-concept connections rather than making these connections on their own.

If novices in a domain begin to learn how to reflect-on-action by first developing a professional vision and then by developing an epistemic frame, it would be helpful to develop an assessment that uses evidence for the development of a professional vision *or* an epistemic frame depending on the novices' prior expertise. An assessment more appropriate for very novice students might detect the presence of action-to-interpretation connections that are central to the development of a professional vision. Similarly, an assessment more appropriate for more expert students might detect the presence of the relevant concept-to-concept connections that are key in the development of an epistemic frame. Moreover, novices may be able to start making relevant concept-to-concept connections but only by relying on other members of the group to fill in gaps in knowledge or skill. In that case, it may be more appropriate to assess the novice group as a whole rather than assess each individual.

We address these issues in the context of *Land Science*, a virtual internship in urban planning. Virtual internships are digital simulations of real-world internships and thus are modelled after the culture of a particular professional domain. To create an accurate model of a domain's culture, researchers conduct an ethnographic study that examines the ways in which novices learn within a particular domain (Bagley, 2010; Hatfield & Shaffer, 2010; Nash & Shaffer, 2013). Researchers can then identify activities, reflective practices, and pedagogical techniques within ill-structured professional domains that should be accounted for in the design of the internship (Arastoopour & Shaffer, 2015). The development of *Land Science*, for example, was informed by the results of an ethnographic study of the urban planning domain (Bagley, 2010).

In their ethnography, Bagley (2010) describe the ways in which an urban planning professor taught reflection-on-action to his students. He began by modelling reflection-on-action for the students and then proceeded to provide the students with feedback regarding their own reflection-on-action while problem solving. After several sessions in which the professor modelled reflection-on-action, the urban planning students began to solve problems in urban planning by reflecting on their actions using language similar to their professor's. Urban planning is thus a domain in which reflection-on-action is a critical component of learning how to solve complex problems. As such, a simulation of that domain needs to contain not only domain-relevant activities, but also reflection-on-action with a more knowledgeable other.

1.6   Research Questions

In what follows, we investigate three critical issues in the assessment of reflection-on-action. First, we ask whether identifying relevant co-occurrences of codes is an appropriate model for detecting whether students have made relevant concept-to-concept connections within an utterance — and thus whether they are engaging in reflection-on-action. Second, if this model is appropriate, does the discourse of students with and without prior domain experience (relative domain experts and novices, respectively)

differ based on the relevant co-occurrences of codes in the discourse of individuals? As discussed above, experts are more likely to make relevant concept-to-concept connections, so we hypothesize that relative domain experts will exhibit more co-occurrences of codes than novices. Lastly, if novices are less able than relative experts to make connections individually, are novices able to make relevant connections collaboratively? In other words, are novices able to provide metacognitive scaffolding for one another?

We address these issues with the following research questions:

RQ1: Do relevant co-occurrences of codes model relevant concept-to-concept connections in single utterances?

RQ2: Are relative domain experts more likely to exhibit relevant co-occurrences of codes in single utterances than novices?

RQ3: Do novice groups exhibit relevant co-occurrences of codes differently than relative domain experts?

2   METHODS

2.1   The *Land Science* Virtual Internship

The data analyzed in this study were collected from the virtual internship *Land Science* (Bagley, 2010). In *Land Science*, students take on the role of interns at a fictitious urban and regional design firm. The objective of the internship is to present a land-use plan in response to a fictitious request for proposals from the mayor of Lowell, MA. Students work together in groups with adult mentors through an online platform that includes email, chat interface, and various tools and resources. They try to balance the demands of various stakeholder groups, which may be in conflict, and weigh the trade-offs of making various land-use decisions for Lowell.

One key element in the practice of urban planning is a site visit. On a site visit, urban planners physically visit an area of interest in order to observe geographic and demographic features that may affect an eventual land use plan (White & Feiner, 2009). For example, they may look at residents' habits, the presence and behaviour of animals in the area, and whether there are important characteristics or natural features not accounted for in maps or other existing documents. Urban planners conduct these site visits to determine the ways in which city plans can meet the needs of the people who live and work in that city. One critical component of this process is identifying and meeting with *stakeholders*: groups of people who have shared interests and desires for the site. Representing stakeholder interests is therefore an important concept in the domain of urban planning.

One key activity in the virtual internship is thus a *Virtual Site Visit* (VSV). During the VSV, students gather information regarding one particular stakeholder group's needs. They do so by reading documents that, within the fiction of the internship, were written by members of the stakeholder group. The students

take notes on their stakeholders' concerns based on the results of their research, which inform students' proposed land-use plans later in the internship. In other words, the VSV simulates one of the complex problems that real urban planners solve, namely, identifying stakeholder concerns. But of course, novices can't interpret their actions in the domain or make connections between relevant domain concepts without the guidance of a more knowledgeable other through reflective discussion. Therefore, in order to accurately simulate professional practice, the VSV must be followed by a reflective discussion about the site visit with a more knowledgeable other. To prompt this reflective discussion within the *Land Science* simulation, the adult mentor asked the students the following question:

> So, planners, you just conducted a virtual site visit. What did you find out about your stakeholders?

In *Land Science*, the stakeholders' concerns are categorized in terms of social issues (e.g., housing for low-income residents, job creation, or the local economy) and environmental issues (e.g., the amount of runoff in the water, carbon monoxide levels, or bird populations). These three urban planning concepts — knowledge of social issues, knowledge of environmental issues, and knowledge of representing stakeholders — are both (a) relevant to the activity just completed (the VSV) and (b) important in the context of the urban planning domain (Dodman, McGranahan, & Dalal-Clayton, 2013). The internship therefore prompts students to make relevant concept-to-concept connections between the concepts of representing stakeholder concerns and both the social and environmental issues that stakeholders care about.

2.2 Participants

Data were collected from 186 students at schools in the United States. Of these, 69 were high school students with no prior experience of urban planning before participating in the *Land Science* virtual internship. The remaining 117 were college students enrolled in an introductory urban planning course at a large public university. Of the participants, 91 were male and 95 were female. No other demographic data were collected.

2.3 Data Collection

All participants used the same version of *Land Science*, and all activities occur within the online interface. The conditions of each implementation were standardized to the extent possible in educational settings. *Land Science* consists of a set of discreet activities, which take approximately 10 hours of contact time to complete. High school students completed *Land Science* either as a stand-alone activity (out-of-school time) or as part of a class (e.g., a science or civics class); none of the high school students had learned about urban planning prior to using the simulation. College students completed *Land Science* as part of an introductory urban planning course; thus, the college students learned about urban planning theories and practices immediately before using the simulation.

Because the college students in the study had been introduced to urban planning concepts in their coursework and had more educational experience in general, they were categorized as having *prior domain experience* in the analysis. They were categorized as *relative domain experts*, while the high school students were categorized as *novices*.

The data used in this study were collected from one activity, the VSV reflection, which occurs in a single session early in the simulation, after a review of the request for land-use proposals and the VSV itself. Students were randomly assigned to project teams of 4–5 individuals, and they remained in the same teams from the beginning of the simulation through the VSV reflection activity.

The *Land Science* virtual internship automatically records all messages sent by students and mentors using the built-in online chat interface, which is how students communicate with their groups and their mentors during the simulation. These messages were segmented by utterance, where an *utterance* is defined as a single instant message in the chat program: any text typed into the chat interface and shared with the group using the *Send* button.

There were a total of 693 utterances in response to this reflective question. College students averaged 4.34 utterances in response to the reflection prompt (SD = 3.58), and high school students averaged 2.68 utterances in response to the prompt (SD = 1.77).

2.4 Human Evaluation of Relevant Concept-to-Concept Connections

To determine whether relevant co-occurrences of codes can model human-evaluated relevant concept-to-concept connections, two trained humans evaluated the data for the following:

- relevant concept-to-concept connections within utterances;
- relevant concept-to-concept connections within group discussions.

To identify relevant concept-to-concept connections, the human raters used a coding rubric (see Table 1) to determine the presence or absence of the three key concepts: knowledge of social issues, knowledge of environmental issues, and knowledge of representing stakeholders. The raters then used their judgment to assess whether students made connections among the concepts.

Table 1. Concept Codes

Code	Code Description	Example
Knowledge of Social Issues	Utterance referring to social issues (e.g., jobs, crime, housing)	I worked with a group that cared about nests, housing, phosphorous, and runoffs.
Knowledge of Environmental Issues	Utterance referring to environmental issues (e.g., runoff, pollution, animal habitats)	I worked with the Connecticut River Water council and they cared about the environment.
Knowledge of Representing Stakeholders	Utterance referring to representing stakeholders (e.g., referring to a specific stakeholders' needs by name, referring to the needs of the stakeholder group)	You may have to make compromises, because the stakeholder groups sometimes disagree.

Relevant concept-to-concept connections within utterances. Because the purpose of the VSV was to collect information about stakeholder concerns for the community, we defined utterances as exhibiting *relevant concept-to-concept connections* whenever a student identified a relationship between the urban planning concepts of stakeholders and their social and environmental issues. Here is one such example from the dataset:

> they would like to boost the economy without harming any natural habitats

This utterance was considered to have made the relevant concept-to-concept connections because this student recognized that the stakeholders ("they") understood that boosting "the economy" might cause some harm to the environment ("natural habitats"). The following utterance also draws this connection:

> They each have something they want to address and some of them intertwine as well. Their
> primary mission is to alleviate the poverty gap by having lower income people get involved in the
> housing market so they want a lot of houses. They also care a lot about the water quality.

This utterance also draws an explicit connection ("some of them intertwine as well") with the stakeholders ("they"; "their primary mission") and their environmental ("water quality") and social ("alleviate the poverty gap") concerns.

In contrast, this utterance was not considered to have made the relevant concept-to-concept connections:

> I definitely would prefer to preserve wildlife over increasing housing for the community. I feel
> that species that are struggling to survive deserve protection from people who want expansion. I
> feel that their ecosystem should not be destroyed in order to provide more houses for the town.

Although this utterance referred to environmental issues ("wildlife") and social issues ("housing"), it was not considered to have made the relevant concept-to-concept connections. The focus of this particular domain activity was on the stakeholders and *their* desires for the city plan. This speaker instead focused on her/his own desires. This utterance is therefore irrelevant to the actions performed in the VSV and was not coded as having made the relevant concept-to-concept connections.

Two trained human raters independently evaluated 40 randomly selected utterances for each of the three codes and indicated whether they contained the relevant concept-to-concept connection. Their inter-rater reliability was calculated using Cohen's kappa (κ), and a high level of agreement was found: κ > 0.82 for all codes (see Table 2, below).

Relevant concept-to-concept connections within group discussions. We then segmented the data by group discussion to determine whether groups were able to make relevant concept-to-concept connections, even if no individual in the group was able to do so alone. Two human raters analyzed a random sample of group discussions to determine whether the group made the relevant concept-to-

concept connections. This coding included connections made across utterances. In other words, the human raters evaluated whether the group as a *whole* made those connections, even if no single individual made them.

For example, the following conversation was coded as having made the relevant concept-to-concept connections:

> **Student 1**: I found that they are mostly concerned with the environment.
> **Student 2**: A lot of their concerns are a direct effect of the local industries and manufacturers.
> **Student 3**: I found out that that as like Student 1 they are concerned for the environment.
> **Student 1**: The population of wildlife, water quality, and air quality are the three biggest concerns I saw.

Students 1 and 3 both made connections between stakeholders and their environmental concerns, but they didn't make any connections to the stakeholders' social concerns. Student 2 filled in their knowledge gap by providing them with the information about the stakeholders' social concerns. Although none of these students made the concept-to-concept connections between stakeholders and their environmental *and* social concerns by themselves, they were able to make those connections collectively.

Two trained human raters independently evaluated 54 randomly selected group conversations and indicated whether they contained the relevant concept-to-concept connections. Their inter-rater reliability was calculated, and a high level of agreement was found: κ = 0.81.

2.5 Automated Coding

Because learning how to make action-to-interpretation connections (and thus learning how to talk using labels in the domain) is a necessary step in the development of reflection-on-action, we developed an automated coding model for the codes in Table 1 to detect the relevant concept labels for this particular activity. To do so, we used regular expression-matching to code relevant concepts in the student discourse.

For example, to automate the code *knowledge of environmental issues*, we developed an algorithm that identifies text relevant to the environmental issues in *Land Science*, such as "carbon monoxide," "phosphorous," and "air quality." Regular expressions ensure accurate string matching. For instance, the regular expression /bCO/b identifies instances of "CO" (carbon monoxide) but not words containing "co," such as "council" or "economy."

All three automated coding algorithms were validated by two trained human raters, as shown in Table 2. For each code, two trained human raters and the coding algorithm independently rated a random sample of 40 chat utterances. Cohen's kappa was calculated between the two human raters and between each human rater and the coding algorithm. To determine whether the kappa values obtained for these samples can be reasonably generalized to the whole dataset, we calculated a rho (ρ) value for

each kappa using the *rhoR* package (Shaffer, Rogers, Eagan, & Marquart, 2016) for the *R* statistical computing software platform. Rho uses an empirical sampling process that produces, for any inter-rater reliability statistic, an estimate of its expected Type I error rate for a given sample. Because κ > 0.65 and ρ < 0.05 for all codes and all combinations of raters, we used the automated coding algorithms to code all the utterances in the dataset.

Table 2. Code Validation

Code	R1 vs. R2		R1 vs. CA		R2 vs. CA	
	Kappa	**Rho**	**Kappa**	**Rho**	**Kappa**	**Rho**
Knowledge of Social Issues	0.86	0.01	1.00	< 0.01	1.00	0.01
Knowledge of Environmental Issues	1.00	< 0.01	1.00	0.01	0.93	0.01
Knowledge of Representing Stakeholders	0.82	0.01	0.88	< 0.01	0.94	0.01

R1 = Human Rater One; R2 = Human Rater Two; CA = Coding Algorithm

Using these automated coding algorithms, we were able to detect:

1. Relevant co-occurrences of codes within utterances
2. Relevant co-occurrences of codes within group discussions

We then tested whether *co-occurrence* of the three codes indicated *connections* among them. While co-occurrence is necessary for connection, it may not be sufficient. However, research (i Cancho & Solé, 2001; Lund & Burgess, 1996) suggests that co-occurrence is a good proxy for connection, and we tested that hypothesis here.

Relevant co-occurrences of codes as a model of relevant concept-to-concept connections within utterances. Because the purpose of the VSV is to collect information about stakeholder concerns, we considered co-occurrences among the urban planning codes pertaining to representing stakeholders and their social and environmental issues to be a model of a connection made between those three concepts. All three of these concept codes were required to be present in the utterance in order to qualify as having made the *relevant co-occurrences of codes*. In other words, an utterance that coded positively for the codes *knowledge of social issues* and *knowledge of environmental issues* but negatively *knowledge of representing stakeholders* would not be considered to have the relevant co-occurrences of codes. The following utterance exhibits the relevant co-occurrences of codes:

> there are several concerns among the **stakeholders**. mainly **environment** concerns like the **water** quality and the **runoff**, and the **economic** concerns as the development of the community

This utterance coded positively for *knowledge of representing stakeholders* ("stakeholders"), *knowledge of environmental issues* ("environment," "water," and "runoff"), and *knowledge of social issues* ("economic"). Because all three of the relevant codes appeared within the same utterance, this utterance was considered to have made the relevant co-occurrences of codes.

Relevant co-occurrences of codes within group discussions. The relevant co-occurrences of codes were detected both at the level of individual utterances and at the level of the group discussion to determine whether relevant co-occurrences of codes indicates relevant connections made via peer scaffolding.

Chance levels of co-occurrences of codes based on prior domain experience. To account for the possibility that co-occurrences of codes may be random, and thus not indicative of connections, the likelihood that novices and relative domain experts would exhibit the relevant co-occurrences of codes by chance was calculated by computing the base rate of each code within the dataset and then calculating the product of those base rates.

3 RESULTS

RQ1: Do relevant co-occurrences of codes model relevant concept-to-concept connections in single utterances?

We used a logistic regression model to predict relevant concept-to-concept connections as a function of the presence of relevant co-occurrences of codes within an utterance:

$$P(CC = 1|RCC) = \frac{1}{(1 + e^{-(\beta + \beta RCC)})}$$

Where CC = the presence of the relevant concept-to-concept connections
RCC = the presence of the relevant co-occurrences of codes

There was a mean of 0.11 relevant concept-to-concept connections made throughout the dataset (SD = 0.31), while there was a mean of 0.05 relevant co-occurrences of codes made throughout the dataset (SD = 0.22).

Using a logistic regression (see Table 3), the presence of the relevant co-occurrences of codes in an utterance was found to be a significant predictor of relevant concept-to-concept connections. We calculated the model's goodness of fit using Nagelkerke/Cragg & Uhler's pseudo R^2, which was found to be 0.57. When the relevant co-occurrences of codes were not present, the chance that the utterance was coded as containing relevant concept-to-concept connections was only 0.02%. However, the presence of the relevant co-occurrences of codes increased the odds (or relative chance) that the utterance was coded as containing relevant concept-to-concept connections by a multiplicative factor of 87 (which corresponds to an 8700% increase).

Table 3. Logistic Regression Analysis of Relevant Concept-to-Concept Connections

Independent Variable	B	SE B	Wald	Sig.	Exp(B)
Intercept	−4.01	0.58	−6.88	0.00	0.07
Relevant Co-occurrences of Codes	4.47	0.69	6.48	0.00	87.08

These results indicate that the presence of the relevant co-occurrences of codes is a strong predictor of relevant concept-to-concept connections, which in turn implies that the relevant co-occurrences of codes can be used to automatically detect the relevant concept-to-concept connections necessary for reflection-on-action.

RQ2: Are relative domain experts more likely to exhibit relevant co-occurrences of codes in single utterances than novices?

Differences in the likelihood of exhibiting the relevant co-occurrences of codes within a single utterance between novices (high school students) and relative domain experts (college students) were assessed using an independent samples *t*-test. Novices exhibited a slightly higher base rate for each individual code than the relative domain experts (see Table 4). They were thus slightly more likely than the relative domain experts to exhibit the relevant co-occurrences of codes in their utterances by chance.

Table 4. Likelihood of Exhibiting Relevant Co-Occurrences of Codes by Chance

	Base rate			
	Knowledge of Social Issues	Knowledge of Environmental Issues	Knowledge of Representing Stakeholders	Random Co-Occurrences of Codes
Novices	0.35	0.32	0.36	0.04
Relative Domain Experts	0.34	0.28	0.31	0.03

Although novices were more likely to exhibit the relevant co-occurrences of codes by chance, relative domain experts (M = 0.81, SD = 0.40) were significantly more likely than novices (M = 0.48, SD = 0.51) to have at least one utterance in the conversation that contained the relevant co-occurrences of codes: $t(49.02) = 2.387$, $p < 0.05$ (see Figure 2).
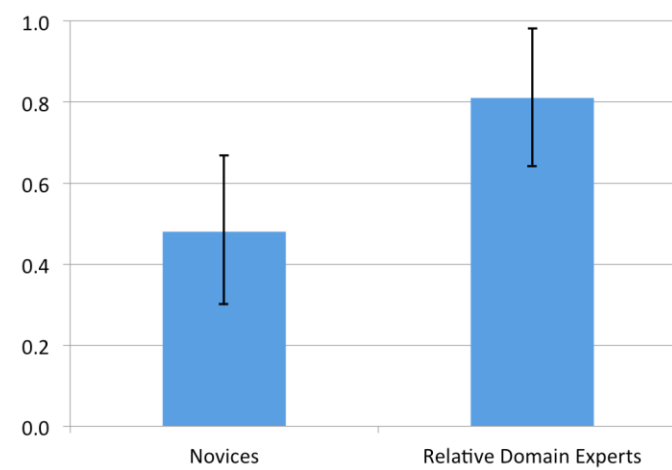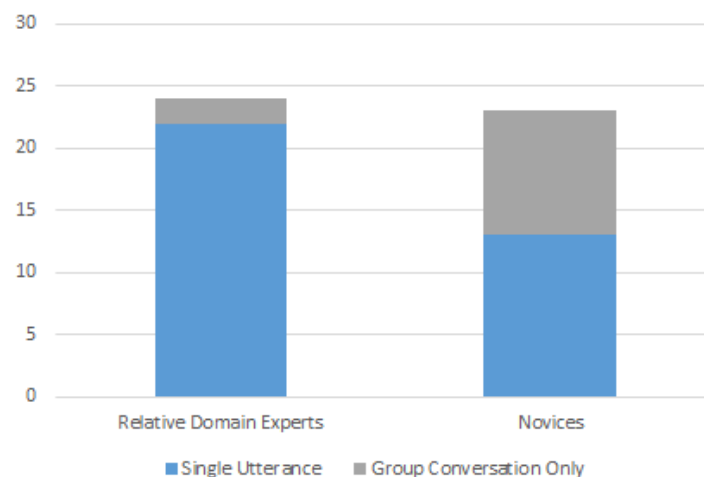

Figure 2. Mean relevant co-occurrences of codes in the discourse of novices and relative domain experts.

This result indicates that prior domain experience is a significant factor in the ability of individual students to identify relevant concepts and make connections among them.

RQ3: Do novice groups exhibit relevant co-occurrences of codes differently than relative domain experts?

To answer this question, we examined whether novices (high school students) and relative domain experts (college students) had a different likelihood of having relevant co-occurrences of codes within a group conversation *in which no single utterance contained the relevant co-occurrences of codes*. We assessed this using an independent samples *t*-test.

Novices (M = 0.37, SD = 0.36) who had no prior domain experience were significantly more likely than relative domain experts (M = 0.07, SD = 0.32) to have conversations that contained the relevant co-occurrences of codes even though no single utterance contained relevant co-occurrences: $t(40.10) = 2.75$, $p < 0.01$ (see Figure 3).

Figure 3. Number of conversations in which at least one utterance contained the relevant co-occurrences of codes (blue) versus the number of conversations in which the conversation contained them but no single utterance did (gray) by prior domain experience.

This result suggests that while college students were generally able to make the relevant connections individually, high school students tended to do so only collaboratively. For example, one student might make a connection between stakeholders and their social concerns, and a second student made the connection between stakeholders and their environmental concerns. In other words, individual college students were more likely to understand the land-use problem in the virtual internship as a complex eco-social problem, whereas high school students tended to see different parts of the problem in isolation. Thus, the group discussion was more critical for the high school students to make the relevant concept-to-concept connections.

4   DISCUSSION

These results suggest that relevant co-occurrences of concepts in student discourse are a good proxy for relevant concept-to-concept connections, which in turn can indicate when students are reflecting-on-action. The results further indicate that automated coding algorithms based on regular expression matching can reliably identify relevant co-occurrences of concepts, but that prior domain experience may affect how and to what extent students are able to make those connections.

Prior research has shown that co-occurrence of concepts in natural language discourse is indicative of genuine concept-to-concept connections (see, e.g., Dorogovtsev & Mendes, 2003; i Cancho & Solé, 2001; Landauer, McNamara, Dennis, & Kintsch, 2007; Lund & Burgess, 1996). This study confirms these findings but also demonstrates that identification of domain-specific concepts in a constrained context can be automated using simple regular expression matching. Of course, defining the relevant concepts still requires domain expertise. Although there are a constellation of natural language processing techniques often described as *topic modelling* (Blei & Lafferty, 2009) that can find latent concepts based on correlations of word usage, these statistical trends do not necessarily correspond to concepts of interest in learning analytics contexts (see, e.g., Andrzejewski, Zhu, & Craven, 2009; Southavilay, Yacef, Reimann, & Calvo, 2013; Tang, Meng, Nguyen, Mei, & Zhang, 2014). However, our findings suggest that automated coding algorithms can reliably identify relevant concepts using little more than keyword matching and simple regular expressions. This is possible in part because the context in which reflection-on-action takes place in many educational settings is highly constrained. In *Land Science*, for example, students are responding to a specific question that prompts them to reflect on their actions in a specific activity (the virtual site visit) in the context of a specific land-use problem set in a specific location.

Additional research is needed to characterize the extent to which, and under what circumstances, co-occurrence of concepts is equivalent to concept-to-concept connection. However, the approach presented here has considerable potential for work in learning analytics, as reflection-on-action is only one area where connections among concepts are theorized to be important. For example, DiSessa (1988) describes learning as a process whereby *phenomenological primitives* — isolated elements of experiential knowledge — are connected through theoretical frameworks to develop not just new knowledge but deep, systematic understanding. Similarly, Linn, Eylon, and Davis (2004) argue that students develop expertise by constructing a *knowledge web*: a repertoire of ideas and the connections among them. Such theories have also been developed in specific domains. Madani and colleagues (2017), for instance, argue that surgical expertise is characterized by connections among core concepts and principles that guide decision-making in unique and diverse scenarios. Thus, the approach described here for identifying concept-to-concept connections may have broad applicability in learning analytics.

Automated assessment of reflection-on-action in particular may improve learning analytics in virtual learning environments. In a virtual learning environment, novices can learn to address complex issues in a domain by solving problems and then talking about their solutions — what worked, what didn't and why — with one another and with more knowledgeable others. This process can be scaffolded by

allowing novices to work on problems with their peers, who might be able to bring different information and perspectives to the discussion. The automated assessment of reflection-on-action thus may enable virtual learning environments to scaffold the problem solving process, helping students to make connections by analyzing conversations in real time. Doing so may be useful for novices who do not have an extensive professional repertoire from which to draw potential solutions.

However, although relevant co-occurrences of codes in student discourse can be used to assess relevant concept-to-concept connections, relevant co-occurrences of codes might need to be measured differently depending on students' level of experience in a domain. Relative domain experts were more likely to make relevant co-occurrences of codes in a single turn of talk than novices. That is, the relative domain experts were able to make relevant concept-to-concept connections independently based on their previous experience, while the novices did not have previous domain experiences on which to base such connections. In contrast, when we excluded single utterances that made the relevant co-occurrences of codes in order to examine the role of peer scaffolding and collaboration on reflection-on-action, novices were significantly more likely to make the relevant connections over several turns in discourse. In other words, the novices in this study were able to identify the domain concepts relevant to the activity by using domain-relevant labels, but they were significantly less likely to make the relevant connections by themselves.

These results have implications for the assessment of reflection-on-action in virtual learning environments, and perhaps more broadly as well. For students who have prior experience in the domain, using the utterance as the unit of analysis may be an appropriate model of good reflection-on-action. However, for students with no prior experience in the domain, it appears that using the group discussion as the unit of analysis may be more appropriate. These results suggest that novices first learn how to link their actions to the interpretation of those actions in the domain before they are able to create the relevant concept-to-concept connections necessary for good reflective discourse. In other words, novices may first learn how to talk about the domain in the way experts do, and only then develop their ability to think, act, identify, and justify their decisions appropriately in the domain. However, development of expertise may not proceed so linearly (see, e.g., Arts, Gijselaers, and Boshuizen, 2006). This study includes only two groups of students, and thus only two levels of expertise, so it is difficult to draw larger conclusions about the development of reflective ability over a longer period of education or training.

The identification of concept-to-concept connections through the use of co-occurrences appears to be one valid method for automatically assessing reflection-on-action. Reflection-on-action is a critical learning process for 21st-century thinking, as it is the means through which novices develop the ability to solve complex problems through reflection-in-action. Future studies will need to examine this approach in other domains and with larger and more diverse levels of student expertise to characterize the circumstances under which the approach remains valid.

4.1   Limitations

This study argued that the presence of relevant concept-to-concept connections in reflective discourse is indicative of the development of an epistemic frame within the domain. However, according to epistemic frame theory, an epistemic frame is more than a set of concept-to-concept connections relevant to an action taken in discourse, but is instead a coherent structure of appropriate and appropriately weighted connections (Shaffer, 2012). In this study, we focused solely on a small number of concept-to-concept connection rather than the connections among those concept-to-concept connections. The presence of relevant concept-to-concept connections may therefore be a necessary, but not sufficient, form of evidence to warrant the claim of the existence of an epistemic frame.

Even in the more limited context of detecting concept-to-concept connections, further work is needed. This study suggests that co-occurrences of relevant codes can serve as a proxy for concept-to-concept connections, but this should be tested against other techniques commonly used to classify text, such as *latent semantic analysis* (Dumais, Furnas, Landauer, Deerwester, & Harshman, 1988) and other natural language processing techniques. In future research, we will conduct studies to compare the approach described here with other text classification processes. In particular, we will do so with larger numbers of students, representing a range of levels of expertise, in various virtual learning environments that incorporate reflection-on-action. This will allow for better characterization of the strengths and limitations of the approach developed in this study.

Another area where further study is needed involves connections across utterances. In cases where concepts co-occur across multiple utterances, research has shown that such co-occurrences are likely to be meaningful only in recent temporal context (Siebert-Evenstone, Arastoopour, Collier, Swiecki, Ruis, & Shaffer, 2016). Additional research is needed to characterize the size of the *window* (i.e., the number of utterances or length of time) in which the co-occurrence of concepts is a meaningful measure of concept-to-concept connection.

5   CONCLUSION

In this paper, we presented a learning analytic technique for the automated detection of reflection-on-action in discourse during complex problem-solving activities. We focused on both reflection in individual discourse and collaborative reflection among student groups. Our results suggest that it is possible to detect student reflection-on-action in virtual learning environments by identifying co-occurrences of complex character string matches, but that different models may be appropriate depending on students' prior domain experience.

ACKNOWLEDGEMENTS

MacArthur Foundation, the Spencer Foundation, the Wisconsin Alumni Research Foundation, and the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin–Madison. The opinions, findings, and conclusions do not reflect the views of the funding agencies, co-operating institutions, or other individuals.

REFERENCES

Andrzejewski, D., Zhu, X., & Craven, M. (2009). Incorporating domain knowledge into topic modeling via Dirichlet forest priors. *Proceedings of the 26th Annual International Conference on Machine Learning* (ICML '09), 14–18 June 2009, Montreal, Quebec, Canada (pp. 25–32). New York: ACM. http://dx.doi.org/10.1145/1553374.1553378

Arastoopour, G., Chesler, N. C., & Shaffer, D. W. (2014). Epistemic persistence: A simulation-based approach to increasing participation of women in engineering. *Journal of Women and Minorities in Science and Engineering, 20*(3), 211–234. http://dx.doi.org/10.1615/JWomenMinorScienEng.2014007317

Arastoopour, G., & Shaffer, D. W. (2015). Epistemography and professional CSCL environment design. *Exploring the material conditions of learning: The Computer Supported Collaborative Learning (CSCL) Conference 2015*, 7–11 June 2015, Gothenberg, Sweden (Vol. 1, pp. 2014–211). International Society of the Learning Sciences.

Arts, J. A. R., Gijselaers, W. H., & Boshuizen, H. P. A. (2006). Understanding managerial problem-solving, knowledge use and information processing: Investigating stages from school to the workplace. *Contemporary Educational Psychology, 31*(4), 387–410. https://doi.org/10.1016/j.cedpsych.2006.05.005

Autor, D. H., Levy, F., & Murnane, R. J. (2003a). Computer-based technological change and skill demands: Reconciling the perspectives of economists and sociologists. In E. Applebaum, A. D. Bernhardt, & R. J. Murnane, *Low-wage America: How employers are reshaping opportunity in the workplace* (pp. 121–154). New York: Russell Sage Foundation.

Autor, D. H., Levy, F., & Murnane, R. J. (2003b). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics, 118*(4), 1279–1333. https://doi.org/10.1162/003355303322552801

Bagley, E. (2010). The epistemography of an urban and regional planning practicum: Appropriation in the face of resistance. WCER Working Paper 2010-8. Wisconsin Center for Education Research, University of Wisconsin–Madison.

Blei, D. M., & Lafferty, J. D. (2009). Topic models. In A. N. Srivastava & M. Sahami (Eds.), *Text mining: Classification, clustering, and applications* (pp. 71–94). Boca Raton, FL: CRC Press.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32–42. https://doi.org/10.3102/0013189X018001032

Campione, J. C., Shapiro, A. M., & Brown, A. L. (1995). Forms of transfer in a community of learners: Flexible learning and understanding. In A. McKeough, J. Lupart, & A. Marini (Eds.), *Teaching for transfer: Fostering generalization in learning* (pp. 35–69). Mahwah, NJ: Erlbaum

Califf, M. E., & Mooney, R. J. (2003). Bottom-up relational learning of pattern matching rules for information extraction. *The Journal of Machine Learning Research, 4,* 177–210.

Chesler, N. C., Ruis, A. R., Collier, W., Swiecki, Z., Arastoopour, G., & Shaffer, D. W. (2015). A novel paradigm for engineering education: Virtual internships with individualized mentoring and assessment of engineering thinking. *Journal of Biomechanical Engineering, 137*(2), 024701:1-8. http://dx.doi.org/10.1115/1.4029235

Coutinho, S., Wiemer-Hastings, K., Skowronski, J. J., & Britt, M. A. (2005). Metacognition, need for cognition and use of explanations during ongoing learning and problem solving. *Learning and Individual Differences, 15*(4), 321–337. https://doi.org/10.1016/j.lindif.2005.06.001

Desautel, D. (2009). Becoming a thinking thinker: Metacognition, self-reflection, and classroom practice. *The Teachers College Record, 111*(8), 1997–2020.

DiSessa, A. A. (1988). Knowledge in pieces. In G. Forman & P. Pufall (Eds.), *Constructivism in the computer age* (pp. 47–70). Hillsdale, NJ: Erlbaum.

Dodman, D., McGranahan, G., & Dalal-Clayton, B. (2013). *Integrating the environment in urban planning and management: Key principles and approaches for cities in the 21st century*. Geneva, Switzerland: United Nations Environmental Programme.

Dorogovtsev, S. N., & Mendes, J. F. (2003). *Evolution of networks: From biological nets to the Internet and WWW*. Oxford, UK: Oxford University Press.

Dumais, S., Furnas, G., Landauer, T., Deerwester, S., & Harshman, R. (1988). Using latent semantic analysis to improve access to textual information. In J. J. O'Hare (Ed.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '98), 4–9 April 2009, Boston, MA, USA (pp. 281–285). Washington, DC: Human Factors in Computing Systems. http://dx.doi.org/10.1145/57167.57214

Eraut, M. (1994). *Developing professional knowledge and competence*. London: Falmer.

Fogarty, R. (1994). The mindful school: How to teach for metacognitive reflection. Palatine, IL: Skylight Publishing.

Friedman, T. L. (2006). *The world is flat [updated and expanded]: A brief history of the twenty-first century*. New York: Macmillan.

Gama, C. (2004). Metacognition in interactive learning environments: The reflection assistant model. In J. C. Lester, R. M. Vicari, & F. Paraguaçu (Eds.), *Proceedings of the 7th International Conference on Intelligent Tutoring Systems* (ITS 2004), 30 August–3 September 2004, Maceió, Alagoas, Brazil (pp. 668–677). Springer: Berlin Heidelberg. http://dx.doi.org/10.1007/b100137

Gibson, A., & Kitto, K. (2015). Analysing reflective text for learning analytics: An approach using anomaly recontextualisation. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 275–279). New York: ACM. http://dx.doi.org/10.1145/2723576.2723635

Goodwin, C. (1994). Professional vision. *American Anthropologist, 96*(3), 606–633.

Graesser, A. C., Foltz, P. W., Rosen, Y., Shaffer, D. W., Forsyth, C., & Germany, M. (in press). Challenges of assessing collaborative problem solving. In E. Care, P. Griffin, & M. Wilson (Eds.), *Assessment and Teaching of 21st Century Skills*. Springer.

Grant, A. M. (2001). Rethinking psychological mindedness: Metacognition, self-reflection, and insight. *Behaviour Change, 18*(1), 8–17.

Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.). (2009). *Handbook of metacognition in education*. Mahwah, NJ: Erlbaum/Taylor & Francis.

Hatfield, D. L., & Shaffer, D. W. (2010). The epistemography of journalism 335: Complexity in developing journalistic expertise. *Proceedings of the 9th International Conference of the Learning Sciences* (ICLS 2010), 29 June–2 July 2010, Chicago, IL, USA (Vol. 1, pp. 628–635). International Society of the Learning Sciences.

Hatfield, D., Shaffer, D. W., Bagley, E. S., Nulty, A., & Nash, P. (2008). Reflection in professional play. *Proceedings of the 8th International Conference of the Learning Sciences* (ICLS '08), 24–28 June 2008, Utrecht, Netherlands (Vol. 3, pp. 245–252). International Society of the Learning Sciences.

Howard, D. V. (1983). *Cognitive psychology: Memory, language, and thought.* New York: Macmillan.

i Cancho, R. F., & Solé, R. V. (2001). The small world of human language. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 268*(1482), 2261–2265. http://dx.doi.org/10.1098/rspb.2001.1800

Kim, Y. R., Park, M. S., Moore, T. J., & Varma, S. (2013). Multiple levels of metacognition and their elicitation through complex problem-solving tasks. *The Journal of Mathematical Behavior, 32*(3), 377–396. https://doi.org/10.1016/j.jmathb.2013.04.002

Kitchener, K. S. (1983). Cognition, metacognition, and epistemic cognition: A three level model of cognitive processing. *Human Development, 26*, 222–232. https://doi.org/10.1159/000272885

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Erlbaum.

Lave, J. (1988). *The culture of acquisition and the practice of understanding*. (IRL report 88-00087). Palo Alto, CA: Institute for Research on Learning.

Levy, F., & Murnane, R. J. (2004). *The new division of labor*. Princeton, NJ: Princeton University Press.

Linn, M. C., Eylon, B.-S., & Davis, E. A. (2004). The knowledge integration perspective on learning. In M. C. Linn, E. A. Davis, & P. Bell (Eds.), *Internet environments for science education* (pp. 29–46). Mahwah, NJ: Erlbaum.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers, 28*(2), 203–208. http://dx.doi.org/10.3758/BF03204766

McAleese, R. (1998). The knowledge arena as an extension to the concept map: Reflection in action. *Interactive Learning Environments, 6*(3), 251–272.

Madani, A., Vassiliou, M. C., Watanabe, Y., Al-Halabi, B., Al-Rowais, M. S., Deckelbaum, D. L., … Feldman, L. S. (2017). What are the principles that guide behaviors in the operating room? Creating a framework to define and measure performance. *Annals of Surgery, 265*(2), 255–267. http://dx.doi.org/10.1097/SLA.0000000000001962

Miller, T. M., & Geraci, L. (2011). Training metacognition in the classroom: The influence of incentives and feedback on exam predictions. *Metacognition and Learning, 6*(3), 303–314. http://dx.doi.org/10.1007/s11409-011-9083-7

Murnane, R. J., & Levy, F. (1993). Why today's high-school-educated males earn less than their fathers did: The problem and an assessment of responses. *Harvard Educational Review, 63*(1), 1–20. https://doi.org/10.17763/haer.63.1.7585v420548725x0

Nash, P., & Shaffer, D. W. (2013). Epistemic trajectories: Mentoring in a game design practicum. *Instructional Science, 41*(4), 745–771. http:/dx.doi.org/10.1007/s11251-012-9255-0

Newman, M. E. (2004). Analysis of weighted networks. *Physical Review E, 70*(5), 056131. https://doi.org/10.1103/PhysRevE.70.056131

Novak, J. D., & Cañas, A. J. (2006). *The theory underlying concept maps and how to construct them*. Pensacola, FL: Florida Institute for Human and Machine Cognition.

Palinscar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*(2), 117–175. http://dx.doi.org/10.1207/s1532690xci0102_1

Powell, W. W., & Snellman, K. (2004). The knowledge economy. *Annual Review of Sociology*, 199–220.

Ruckelshaus, C., & Leberstein, S. (2014). *Manufacturing low pay: Declining wages in the jobs that built America's middle class.* http://www.nelp.org/page/-/Justice/2014/Manufacturing-Low-Pay-Declining-Wages-Jobs-Built-Middle-Class.pdf?nocdn=1

Rupp, A. A., Gushta, M., Mislevy, R. J., & Shaffer, D. W. (2010). Evidence-centered design of epistemic games: Measurement principles for complex learning environments. *The Journal of Technology, Learning and Assessment, 8*(4). https://ejournals.bc.edu/ojs/index.php/jtla/article/view/1623

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. London: Temple Smith.

Schön, D. A. (1984). The architectural studio as an exemplar of education for reflection-in-action. *Journal of Architectural Education, 38*(1), 2–9. http://dx.doi.org/10.1080/10464883.1984.10758345

Schön, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. San Francisco, CA: Jossey-Bass.

Schön, D. A. (1992). The crisis of professional knowledge and the pursuit of an epistemology of practice. *Journal of Interprofessional Care, 6*(1), 49–63. http://dx.doi.org/10.3109/13561829209049595

Schön, D. A. (1995). Knowing-in-action: The new scholarship requires a new epistemology. *Change,* November/December, 27–34. http://dx.doi.org/10.1080/00091383.1995.10544673

Shaffer, D. W. (2006). Epistemic frames for epistemic games. *Computers and Education, 46*(3), 223–234. http://dx.doi.org/10.1016/j.compedu.2005.11.003

Shaffer, D. W. (2012). Models of situated action: Computer games and the problem of transfer. In C. Steinkuehler, K. Squire, & S. Barab (Eds.), *Games, learning, and society: Learning and meaning in the digital age* (pp. 403–433). Cambridge, UK: Cambridge University Press

Shaffer, D. W., Hatfield, D., Svarovsky, G., Nash, P. Nulty, A., Bagley, E., Franke, K., Rupp, A. A., & Mislevy, J. R. (2009). Epistemic network analysis: A prototype for 21st century assessment of learning. *The International Journal of Learning and Media*, *1*(2), 33–53. http://dx.doi.org/10.1162/ijlm.2009.0013

Shaffer, D. W., Rogers, B., Eagan, B. R., & Marquart, C. (2016). rhoR: Rho for inter-rater reliability [Computer software]. https://CRAN.R-project.org/package=rhoR.

Siebert-Evenstone, A. L., Arastoopour, G., Collier, W., Swiecki, Z., Ruis, A. R., & Shaffer, D. W. (2016). In search of conversational grain size: Modeling semantic structure using moving stanza windows. In C.‑K. Looi, J. Polman, U. Cress, & P. Reimann (Eds.), *Transforming learning, empowering learners: Proceedings of the 12th International Conference of the Learning Sciences* (ICLS '16), 20–24 June 2016, Singapore (Vol. 1, pp. 631–638).

Southavilay, V., Yacef, K., Reimann, P., & Calvo, R. A. (2013). Analysis of collaborative writing processes using revision maps and probabilistic topic models. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium (pp. 38–47). New York: ACM. http://dx.doi.org/10.1145/2460296.2460307

Spector, J. M., Merrill, M. D., Elen, J., & Bishop, M. J. (2013). *Handbook of research on educational communications and technology*. Springer Science & Business.

Svarovsky, G. N., & Shaffer, D. W. (2006). Design meetings and design notebooks as tools for reflection in the engineering design course. *Proceedings of the 36th Annual Frontiers in Education Conference* (FIE 2006), 27 October–1 November 2006, San Diego, CA, USA (pp. 7–12). IEEE.

Tang, J., Meng, Z., Nguyen, X., Mei, Q., & Zhang, M. (2014). Understanding the limiting factors of topic modeling via posterior contraction analysis. *Proceedings of the 31st International Conference on Machine Learning, PMLR, 32*(1), 190–198.

Ullmann, T. D., Wild, F., & Scott, P. (2012). Comparing automatically detected reflective texts with human judgements. *Proceedings of the 2nd Workshop on Awareness and Reflection in Technology-Enhanced Learning* (AR-TEL '12), 18 September 2013, Saarbrucken, Germany (pp. 101–116).

Usher, R., & Bryant, I. (1997). *Adult education and the postmodern challenge*. London: Routledge.

Velardi, P., Fabriani, P., & Missikoff, M. (2001). Using text processing techniques to automatically enrich a domain ontology. *Proceedings of the International Conference on Formal Ontology in Information Systems* (FOIS '01), 17–19 October 2001, Ogunquit, Maine, USA (pp. 270–284). New York: ACM. http://dx.doi.org/10.1145/505168.505194

Veenman, M. V., Van Hout-Wolters, B. H., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning, 1*(1), 3–14. http://dx.doi.org/10.1007/s11409-006-6893-0

Weeber, M., Klein, H., de Jong-van den Berg, L., & Vos, R. (2001). Using concepts in literature-based discovery: Simulating Swanson's Raynaud–fish oil and migraine–magnesium discoveries. *Journal of the American Society for Information Science and Technology, 52*(7), 548–557. http://dx.doi.org/10.1002/asi.1104

White, S., & Feiner, S. (2009). SiteLens: Situated visualization techniques for urban site visits. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '09), 4–9 April 2009, Boston, MA, USA (pp. 1117–1120). New York: ACM. http://dx.doi.org/10.1145/1518701.1518871

Wood, D., Bruner, J., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Child Psychiatry, 17*, 89–100. http://dx.doi.org/10.1111/j.1469-7610.1976.tb00381.x

Wood, P. K. (1983). Inquiring systems and problem structures: Implications for cognitive development. *Human Development, 26*, 249–265. http://dx.doi.org/10.1159/000137808

Xun, G. E., & Land, S. M. (2004). A conceptual framework for scaffolding Ill-structured problem-solving processes using question prompts and peer interactions. *Educational Technology Research and Development, 52*(2), 5–22. http://dx.doi.org/10.1007/BF02504836