Chapter 5

Knowledge in action. The informal statistical inference content knowledge of three pre-service teachers while teaching ISI in upper-primary school

This chapter is based on:

CHAPTER 5

Abstract
Teachers who introduce primary school students to informal statistical inference (ISI) must be able to use their own content knowledge of ISI (ISI-CK) in teaching. Research shows that teachers who teach ISI tend to focus on data analysis rather than on inferential reasoning. Our embedded multiple case study describes the ISI-CK of three participants during an ISI-related teacher college intervention as they were teaching their upper-primary students an empirical inquiry lesson containing affordances to discuss ISI. The qualitative analysis showed that while the sample size was discussed extensively, the sampling method was not. The pre-service teachers were able to make inferences from collected data, but they faced difficulties in explaining why making inferences is possible. Providing this explanation requires that the lesson contain tools that illustrate the law of large numbers. Future research should determine whether such tools indeed allow teachers to explain why making inferences is possible.
“Ultimately, teachers must be able to know and use mathematics in practice, not merely do well in courses or answer pedagogically contextualized questions in interviews.”

Ball, Lubienski, and Mewborn (2001, p. 451)

Introduction

The ability to make inferences from empirical data is of growing importance (Liu & Grusky, 2013). Given the societal importance of inferential reasoning, it may be beneficial to introduce inferential reasoning to students during their primary education to prepare them for their roles in future society. One way to introduce inferential reasoning to primary school students is to introduce them to informal statistical inference (ISI) (Groth & Meletiou-Mavrotheris, 2018), which can be defined as “a generalized conclusion expressed with uncertainty and evidenced by, yet extending beyond, available data” (Ben-Zvi, Bakker, & Makar, 2015, p. 293). Early introduction to ISI allows students to be exposed repeatedly to its statistical concepts over an extended period of time. This would gradually familiarize them with inferential reasoning and related statistical concepts (Watson & Moritz, 2000). Such an early introduction to inferential reasoning could prepare students for learning formal statistical inference and may help them develop a critical stance toward data and inference (Zieffler, Garfield, delMas, & Reading, 2008).

Whereas formal statistical inference usually relies on formal tests based on probability theory, the statistical reasoning involved in ISI is usually of lower complexity (Makar & Rubin, 2018). For example, ISI allows for qualitative expressions for uncertainty, such as “I am quite sure,” instead of quantitative expressions, such as confidence intervals. Furthermore, in ISI, inferences can be based on simulations instead of on closed-form formulas (Mills, 2002). In addition, a diversity of images of distributions can be used, such as dot plots that highlight individual samples in sampling distributions (Makar & Rubin, 2018). Research shows examples in which primary school students are indeed able to begin reasoning about inference (Ben-Zvi, 2006; Leavy, 2017; Makar, 2016; Meletiou-Mavrotheris & Paparistodemou, 2015; Paparistodemou & Meletiou-Mavrotheris, 2008).
Teachers are usually the ones to introduce ISI to primary school students. This requires that they possess appropriate content knowledge of ISI (ISI-CK) (Batanero & Díaz, 2010), as their content knowledge impacts both their pedagogical knowledge (Burgess, 2009; Shulman, 1986) and their students’ achievements (Fennema & Franke, 1992; Hill, Rowan, & Ball, 2005; Rivkin, Hanushek, & Kain, 2005). Although a clear link exists between the content knowledge a teacher shows in professional development settings and how she uses her content knowledge in teaching (Hill et al., 2008), this does not necessarily imply that she is able to apply her content knowledge in teaching (Ball et al., 2001; Rowland, Turner, Thwaites, & Hurst, 2009). A teacher also must be able to “unpack” the content knowledge for teaching purposes (Ma, 1999). The question thus arises to what extent pre-service teachers are able to unpack their ISI-CK in teaching. The research literature shows that (pre-service) teachers have problems teaching ISI. In particular, pre-service teachers teaching ISI show a tendency to focus on data collection and analysis at the expense of attention for inferential reasoning (Leavy, 2010). This tendency is ascribed to their inexperience with conducting statistical investigations and with managing the students’ inquiry process (Heaton & Mickelson, 2002; Leavy, 2010). Moreover, pre-service teachers often fail to design lessons that include opportunities to discuss ISI with students (Leavy, 2010; Makar & Rubin, 2009).

The aim of our study was to determine the extent to which pre-service teachers are able to put their ISI-CK in action when they are supplied with a lesson that provides affordances to discuss ISI with students. To this end, we employed an embedded case study of three pre-service teachers teaching an ISI lesson. This teaching was conducted in the setting of a short intervention consisting of five sessions at a teacher college for primary education (De Vetten, Schoonenboom, Keijzer, & Van Oers, 2018). During this intervention, the three pre-service primary school teachers showed average to good ISI-CK. In this paper, we describe the ISI-CK displayed by these participants during their teaching of an ISI lesson. We also describe how these expressions can possibly be understood. The lesson plan was provided by the first author, who was the pre-service teachers’ teacher educator. The lesson asked the students to answer the question “What is the most frequently used word in this pile of books?” Previous experiences with this lesson had shown it was potentially
useful to engage primary school students in ISI. The results of this study may inform teacher college education how the development of pre-service teachers’ ISI-CK can be supported.

Theoretical Background

For our study among pre-service teachers, we used the Makar and Rubin (2009) ISI framework and conceptualized ISI-CK, as follows:

1. Data as evidence: The inference is based on available data and not on tradition, personal beliefs, or personal experience.
2. Generalization beyond the data: The inference goes beyond a description of the sample data by making a probabilistic claim about a population or a mechanism that produced the sample data.
3. Probabilistic language: Due to sampling variability and the degree of sample representativeness, the inference is inherently uncertain and requires using probabilistic language. For the correct usage of probabilistic language, the origins of uncertainty in inferences must be understood. Therefore, we divided this component into four subcomponents:
   a. Sampling variability: The inference is based on an understanding of sampling variability; it is expressed from an understanding that the outcomes of representative samples are similar and that therefore, under certain circumstances, a sample can be used for an inference (De Vetten et al., 2018; Saldanha & Thompson, 2007).
   b. Sampling method: The inference includes a discussion of the sampling method and the implications of the sample representativeness.
   c. Sample size: The inference includes a discussion of the sample size and the implications of the sample representativeness.
   d. Uncertainty: The inference is expressed with uncertainty and includes a discussion of what the sample characteristics, such as the sampling method employed and the sample size, imply for the certainty of the inference.
Previous research suggests a need to develop (pre-service) primary school teachers’ ISI-CK, as many pre-service teachers have limited knowledge of sampling variability, sampling methods, sample size, and representativeness (Canada, 2006; De Vetten et al., 2018; De Vetten, Schoonenboom, Keijzer, & Van Oers, in press-a; Meletiou-Mavrotheris, Kleanthous, & Paparistodemou, 2014; Mooney, Duni, VanMeenen, & Langrall, 2014; Watson, 2001). Furthermore, they lack awareness that ISI tasks require an inference over and above a descriptive analysis of the data (De Vetten et al., 2018, in press-a). Mixed results were found regarding the extent to which pre-service teachers acknowledge the value of data as evidence and the possibility of using a sample to make (probabilistic) inferences (De Vetten et al., 2018, in press-a).

The need to develop pre-service teachers’ ISI-CK was addressed in one of our previous studies, where we conducted an intervention at a teacher college for primary education (De Vetten et al., 2018). This study showed that during the intervention sessions and in a post-test, key aspects of ISI-CK were probably understood by three quarters of the pre-service teachers. Although these were encouraging results, the purpose of fostering ISI-CK in pre-service teachers is to prepare them for introducing primary school students to ISI. However, it cannot be assumed that the pre-service teachers can use their knowledge in teaching as well (Ball et al., 2001; Putnam & Borko, 2000). For example, Thompson and Thompson (1994) show “that although a teacher had strong conceptual knowledge on a pencil-and-paper test and in a professional development setting, he had trouble talking conceptually about rates during a tutoring session” (quoted in Hill et al., 2008, p. 435). Teaching ISI requires teachers to react immediately to unexpected remarks, questions, and events, to deal with the outcomes of the ISI inquiry that cannot be planned beforehand, and to not only conduct an ISI inquiry themselves but also explain the rationale of making inferences. All of this requires a profound understanding of ISI. Because professional development settings usually do not set such demanding requirements on the knowledge shown, the depth of pre-service teachers’ content knowledge may not be shown in professional development settings (Ball et al., 2001; Mickelson & Heaton, 2004). This implies that teacher knowledge should also be investigated in the context of teaching (Fennema & Franke, 1992; Rowland et al., 2009). As part of the teacher college intervention, the participating pre-service teachers taught an ISI lesson in their placement.
classes in primary school. The current paper describes the expression of the ISI-CK of three participating pre-service teachers who showed average to good ISI-CK in the teacher college intervention while teaching an ISI lesson.

Previous research has shown the difficulties pre-service teachers face in designing a lesson that provides affordances to discuss ISI. It also has revealed that these lessons often focus excessively on data collection and analysis (Heaton & Mickelson, 2002; Leavy, 2010). Therefore, we did not have the pre-service teachers design their own lessons, but instead provided them with a lesson that had proven effective in fostering attention for ISI and that avoided a focus on descriptive analyses. Moreover, pre-service teachers who are also still in the process of learning ISI themselves are likely to experience feelings of fear and uncertainty regarding their ability to teach ISI (Heaton, 2000). Supplying them with a lesson plan could provide appropriate support for taking the risk of teaching new content.

To understand why the pre-service teachers showed particular ISI-CK, we took several context-related factors into account: (1) the pre-service teachers’ ISI-CK displayed during the teacher college intervention, (2) factors related to the design of the particular lesson, (3) pedagogical considerations of the pre-service teachers, such as decisions related to classroom management, and (4) pedagogical beliefs, such as considerations of what their students might or might not understand. The research question this paper addresses is: What ISI-CK do three pre-service primary school teachers express while teaching an ISI lesson in primary school?

Method

Context
This study reports on the ISI-CK of three pre-service primary education teachers – Celine, Demi and Alfred – as they taught an ISI lesson. These three pre-service teachers were among 21 participants of an intervention at a teacher college that aimed to foster their ISI-CK and the pedagogical content knowledge of ISI. Table 1 shows an overview of this intervention and the moments of teaching an ISI lesson in their placement schools.
Table 1

Overview of teacher college intervention

<table>
<thead>
<tr>
<th>Setting</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework Assignment</td>
<td>Homework assignment: Samples in the media</td>
</tr>
<tr>
<td>Session 1</td>
<td>Discussion of homework (60 minutes)</td>
</tr>
<tr>
<td></td>
<td>Real-time computer simulation, illustrating law of large numbers (20 minutes)</td>
</tr>
<tr>
<td>Session 2</td>
<td>Reiteration of learning points simulation (10 minutes)</td>
</tr>
<tr>
<td></td>
<td>Model lesson: “What is the most frequently used word?” (70 minutes)</td>
</tr>
<tr>
<td>Session 3</td>
<td>Car choice activity (20 minutes)</td>
</tr>
<tr>
<td></td>
<td>Discussion PCK of ISI (45 minutes)</td>
</tr>
<tr>
<td>Placement schools</td>
<td>First half of the participants (including Celine and Demi) teach ISI lesson</td>
</tr>
<tr>
<td>Session 4</td>
<td>Evaluation of ISI lessons in placement schools (30 minutes)</td>
</tr>
<tr>
<td>Placement schools</td>
<td>Second half of the participants (including Alfred) teach ISI lesson</td>
</tr>
<tr>
<td>Session 5</td>
<td>Evaluation of ISI lessons in placement schools (15 minutes)</td>
</tr>
</tbody>
</table>

Table 2

Overview characteristics of pre-service teachers

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Celine</th>
<th>Demi</th>
<th>Alfred</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Female</td>
<td>Female</td>
<td>Male</td>
</tr>
<tr>
<td>Age</td>
<td>18 years</td>
<td>19 years</td>
<td>20 years</td>
</tr>
<tr>
<td>Educational background</td>
<td>General secondary education(^a)</td>
<td>General secondary education</td>
<td>General secondary education</td>
</tr>
<tr>
<td>Mathematics in secondary education</td>
<td>Calculus-oriented mathematics</td>
<td>Descriptive statistics; introduction to formal statistical inference</td>
<td>Descriptive statistics; introduction to formal statistical inference</td>
</tr>
<tr>
<td>Statistics at teacher college</td>
<td>Descriptive statistics</td>
<td>Descriptive statistics</td>
<td>Descriptive statistics</td>
</tr>
<tr>
<td>Number of ISI learning objectives attained during intervention</td>
<td>9 out of 12</td>
<td>7 out of 12</td>
<td>9 out of 12</td>
</tr>
<tr>
<td>Average teacher college results</td>
<td>Best 20(^b)</td>
<td>Lowest 30(^b)</td>
<td>Best 20(^b)</td>
</tr>
<tr>
<td>Average results for three mathematics education courses</td>
<td>Best 20(^b)</td>
<td>Lowest 20(^b)</td>
<td>Best 20(^b)</td>
</tr>
<tr>
<td>Average results school placements</td>
<td>Lowest 40(^b)</td>
<td>Best 20(^b)</td>
<td>Best 40(^b)</td>
</tr>
</tbody>
</table>

\(^a\)General secondary education allows for entrance into professional universities, not into research-intensive universities. \(^b\)Relative to all 21 participants in the teacher college intervention.
Celine and Demi were among the first half of the participants who taught the ISI lesson between sessions 3 and 4; Alfred was among the second half who taught the lesson after the halfway evaluation in session 4. A detailed description of the intervention and how the activities fostered the development of the 21 participating pre-service teachers can be found in De Vetten et al. (2018).

**Participants**

The following criteria guided the selection of pre-service teachers from among the participants in the teacher college intervention. First, we hypothesized that appropriate ISI-CK demonstrated in a less demanding setting, such as the pretest or the teacher college sessions, would be necessary to apply this knowledge while teaching an ISI lesson. Therefore, we excluded pre-service teachers with low ISI-CK on the pretest. Second, we wanted to select pre-service teachers who were generally open in expressing their opinions and concerns. Third, the pre-service teachers needed to consent to participate. Fourth, practical planning issues played a role. The three selected pre-service teachers were Celine (high ISI-CK compared to the other participants), Demi (average ISI-CK), and Alfred (high ISI-CK). Table 2 provides an overview of the pre-service teachers’ relevant characteristics. The participants gave informed consent to be part of the research, and their placement schools gave informed consent for the observation and use of the data. The study design was approved by the ethical board of the Faculty of Behavioural and Movement Sciences of Vrije Universiteit, Amsterdam.

We expected that the ISI-CK demonstrated by the three pre-service teachers during the pretest and in the teacher college sessions would be among the main factors for understanding why they showed particular knowledge during teaching. Their ISI-CK shown during the pretest and in the sessions is therefore outlined below and summarized in Table 3. For each learning objective of the intervention, the pre-service teachers’ knowledge indicated by a plus sign (knowledge in line with the learning objective), a minus sign (knowledge not in line with the learning objective), or a plus/minus sign (mixed evidence about knowledge).
Table 3  
Pre-service teachers’ ISI-CK during pretest and sessions of teacher college intervention

<table>
<thead>
<tr>
<th>ISI component</th>
<th>Learning objective</th>
<th>Celine</th>
<th>Demi</th>
<th>Alfred</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data as evidence</td>
<td>Use the data as evidence for a conclusion, instead of other sources.</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Generalization beyond the data</td>
<td>Know that it is possible to use a sample to make general claims about the population.</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Are aware of the inferential nature of research questions.</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Realize sample statistics provide information about the likelihood of the values of the population parameter and know that not every value of the population parameter is evenly likely.</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sampling variability</td>
<td>Understand it is unlikely that another large random sample gives entirely different results.</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Sampling method</td>
<td>Know that random sampling is an appropriate method to obtain a representative sample.</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Prefer random sampling above distributed sampling.</td>
<td>-</td>
<td>-</td>
<td>+/-</td>
</tr>
<tr>
<td></td>
<td>Know that convenience sampling is an inappropriate sampling method to obtain a representative sample.</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Understand why an appropriate sampling method yields a sample in which aggregate characteristics are closely approximate to the population characteristics.</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Sample size</td>
<td>Understand why a sample size of 1,000 is a sufficient sample size for the entire Dutch population (17 million).</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>Acknowledge the uncertainty of inferences and the impossibility of making absolutely certain inferences.</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Know that larger samples yield more certainty about inferences.</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

**Total number of learning objectives understood**  
9    7    9

*Note:* Explanation of the symbols: +: Pre-service teacher’s knowledge was in line with learning objective; -: Knowledge was not in line with learning objective; +/-: Mixed evidence about the knowledge.

During the pretest and the teacher college sessions, all three pre-service teachers used the data as evidence when drawing conclusions from sample
data (objective 1) and acknowledged the possibility of making generalizations (2). During the pretest, they—as most other participants—were not aware that one of the tasks required an inference (3). However, during the teacher college sessions, they were all aware of the inferential reasoning involved in the tasks (3). Alfred stressed the importance of having a reliable sample a number of times. Celine was the only participant who recognized and understood the equiprobability bias in the students’ reasoning (4).

During the pretest, Alfred was the only participant who both acknowledged the appropriateness of random sampling (6) and preferred this random method (7) above distributed sampling (i.e., purposefully selecting individuals to obtain a distributed sample across critical population characteristics (see Watson & Moritz, 2000). However, though he stated at the start of the intervention that random sampling is a good sampling method, he later proposed (as did most participants) that distributed sampling is the preferred sampling method, both in general and in the actual context of selecting a sample of books. Although Demi agreed during the pretest that random sampling is appropriate, she preferred distributed sampling as well. Celine did not think random sampling would yield a representative sample; she also preferred distributed sampling.

Concerning sample size (10), none of the three pre-service teachers showed complete knowledge. All three acknowledged that a small sample might not be representative and that a sample of 10,000 would be unnecessarily large. However, Demi and Alfred also argued that the sample size is dependent on the population, up to a certain point. Moreover, none of the three objected during the first session when most participants denied that a sample of 1,000 is sufficient to represent the entire Dutch population. Celine doubted whether 1,553 is a sufficient sample size. In the pretest, Alfred argued that a sample of about 2,000 would be appropriate.

Celine, Demi, and Alfred all saw the impossibility of making absolute certain generalizations (11) and agreed that larger samples yield more certainty (12). Celine showed a profound understanding of sampling variability, while Alfred and Demi showed basic knowledge of sampling variability (5). Demi showed some inclination during the pretest to strive for complete certainty, as she argued that “precise” answers cannot be gotten from a sample.
Lesson: “What is the most frequently used word?”

The pre-service teachers taught a lesson called “What is the most frequently used word?” This lesson was laid out in a lesson plan that included learning objectives and teaching suggestions. The pre-service teachers were expected to prepare their lesson by adjusting the lesson plan to the specific situation in their placement schools and then document their preparation in their own lesson plan. It was stressed though that the main idea of engaging students in ISI needed to be upheld. Figure 1 shows an abridged version of the lesson plan. During the teacher college intervention, the teacher educator taught this lesson to the pre-service teachers to give them a good example of an ISI lesson.

The lesson involved a statistical investigation with hands-on activities. It centered on a large pile of Dutch children’s novels and the research question of which word would be observed most frequently in the pile. The enormity and visibility of the population were expected to elicit the need to draw a sample and to make inferences. The investigation was based on the five most likely words, which the students determined after discussion. Class discussion was used to reach consensus about the preferred sampling method so that separate groups’ sample data could later be pooled into one large sample. The groups were to conduct an investigation using the agreed-upon sampling method. The analysis of the sample data was kept simple, as only frequencies needed to be counted. In this way, ample time was left for discussing ISI. This discussion dealt with the question about the possibility and certainty of the generalization from the sample results, both from the individual groups’ sample data and from the pooled data. Dialogic classroom talk was used throughout the lesson to elicit the students’ own ideas and approaches, and the pre-service teachers were expected to steer the discussion into the desired direction and provide answers themselves when necessary (Wells, 1999).

The evaluation of the lessons taught by the first half of the intervention participants revealed that they desired more support to discuss the results. The conclusion phase was adapted. Rather than discussing the answers to the question on the handout, the discussion was more structured, with students indicating their answer to the research question before and after discussing a number of statements (see Figure 1, bottom part). As closure, the pre-service teachers could show the last three of these statements, which formed a complete reasoning underlying the conclusion.
Figure 1: Summary of the lesson plan “What is the most frequently used word?”
Table 4

Overview characteristics: lessons of the three pre-service teachers

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Celine</th>
<th>Demi</th>
<th>Alfred</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade level</td>
<td>3</td>
<td>5</td>
<td>3/4</td>
</tr>
<tr>
<td>Number of students</td>
<td>28</td>
<td>27</td>
<td>26</td>
</tr>
<tr>
<td>Time spent on class discussions and making pooled graph</td>
<td>25 minutes</td>
<td>44 minutes</td>
<td>54 minutes</td>
</tr>
<tr>
<td>Time spent on group work</td>
<td>16</td>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td>Total duration</td>
<td>41</td>
<td>59</td>
<td>68</td>
</tr>
</tbody>
</table>

Celine and Demi used the original lesson plan, while Alfred used the adapted version. In general, their lesson plans were closely aligned to the lesson plan provided by the teacher educator. Celine and Demi rewrote most parts in their words and occasionally copied sentences, while Alfred copied most parts and occasionally changed sentences. Table 4 provides an overview of relevant characteristics of the lessons taught by the pre-service teachers.

Data Collection

Two kinds of data were collected. Observational data formed the primary data source. For each lesson, the whole-class interactions were video recorded from two angles and audio recorded, while most of the group interactions were audio recorded. The latter were not used in data analysis, as these contained very little information about the pre-service teachers’ ISI-CK. The first author, who was also the teacher educator during the teacher college intervention, was present as observer. Other data sources were the lesson plans of the pre-service teachers, written work by the students, and the observer’s notes.

A second data source comprised the transcripts of the reflection interview between the pre-service teacher and the first author that were held shortly after the lessons. The discussions during these interviews included discussing notable moments, discussing explanations of the pre-service teachers’ and the students’ conduct during the lesson, discussing observations made by the teacher educator, checking interpretations made by the teacher educator, and discussing the extent to which the learning objectives were attained. These interviews, varying in duration between 20 and 45 minutes, were audio taped.
Data Analysis
The data analysis approach taken was a multiple case study analysis during which patterns in and explanations for the pre-service teachers’ ISI-CK were investigated. This analysis was conducted first for each pre-service teacher separately. Next, a cross-case analysis was performed to find themes within the ISI-CK of the three pre-service teachers.

All data were transcribed from the class discussion and the reflection interviews. The transcripts were coded in Atlas.ti using a process consisting of deductive and inductive elements. On the deductive side, the ISI framework was used to categorize the text data into one or more ISI components. On the inductive side, codes that were short summaries of the text were attached to the text to describe the exact meaning. These codes were subsequently combined into codes with similar meanings or on closely related issues. To decide whether a pre-service teacher’s response to a student’s remark constituted ISI-CK, we used the following criteria:

1. If the pre-service teacher merely repeated a student’s remark, this did not constitute ISI-CK.
2. If the pre-service teacher agreed with a student’s remark, this constituted ISI-CK. However, if the pre-service teacher also agreed with a conflicting remark, additional evidence was sought, either in other parts of the lesson and/or in the reflection interview, to determine whether this additional evidence provided clarity about the pre-service teacher’s ISI-CK. Without such additional evidence, we could not draw conclusions about this particular aspect of the pre-service teacher’s ISI-CK.
3. If the pre-service teacher built on a student’s suggestion and the lesson progressed in the suggested direction, then the student’s suggestion constituted ISI-CK of the pre-service teacher.
4. If the pre-service teacher wrongly summarized a student’s remark, for example by severely simplifying the remark, this could indicate a lack of a particular aspect of ISI-CK. Without further corroborating evidence, such an interpretation was made only tentatively.
5. The pre-service teacher’s intonation could signal whether she or he agreed with a student, and this could signal whether the pre-service teacher’s response constituted ISI-CK.

The first step of the analysis was using the codes to construct a tabulated overview for each pre-service teacher of the ISI-CK that they showed during their lesson. This overview was organized by phase of the lesson and by ISI component (see Table 4). The second step was to search for themes across the pre-service teachers’ ISI-CK. These were identified by jointly displaying the ISI-CK per component for the three pre-service teachers, as in Table 4. Three themes were identified that constituted major findings about the ISI-CK that were visible in all three lessons, although these themes played out differently in each lesson. Both the similarities and the differences between the pre-service teachers within these themes will be discussed in the Results section.

To understand why the pre-service teachers showed particular ISI-CK, the following context-related factors were taken into account: (1) The pre-service teachers’ ISI-CK displayed during the teacher college intervention, both during the pretest and during the sessions at the teacher college itself, (2) factors related to the design of the particular lesson, (3) pedagogical considerations of the pre-service teachers, such as decisions related to classroom management, and (4) pedagogical beliefs, such as beliefs what their students might or might not understand.

For each reflection interview, an overview was made of all statements made by the pre-service teacher during the interview that could help to interpret their behavior. These statements could provide evidence of whether a pre-service teacher’s remark reflected ISI-CK, pedagogical approval, or mere repetition of a student’s remark. The explanations given by the pre-service teachers regarding their conduct during the lesson were an important resource for interpreting their ISI-CK. These explanations were triangulated with the researcher observations, and the coding process and all results were discussed with an external researcher until consensus was reached about the results’ validity.
Table 5
Overview ISI-CK of the pre-service teachers while teaching an ISI lesson

<table>
<thead>
<tr>
<th>Component</th>
<th>ISI-CK of the pre-service teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Introduction phase of the lesson</strong></td>
<td></td>
</tr>
</tbody>
</table>
| Data as evidence | Celine: Made sure that an unlikely candidate, *who*, was not included. Outcome: sensible top 5, *(the, he, a, is, and).*  
Demi: Outcome: sensible top 5 *(the, “het”*, a, he, is).  
Alfred: Made sure that an unlikely candidate, *with*, was not included. Outcome: except *says*, reasonable top 5 *(and, I, the, “het”*, says).* |
| Sampling variability | Celine: Discussed the exact order for the first two words within the hypothetical top 5.  
Demi: Spent considerable time discussing the exact order within top 5.  
Alfred: No remarks. |
| **Design and data collection phase of the lesson** | |
| Data as evidence | Celine: Indicated that an answer to the research question can be found by reading books.  
Demi: Indicated that an answer to the research question can be found by conducting an investigation.  
Alfred: Indicated that an answer to the research question can be found by reading books. |
| Generalization | Celine: Asked whether all books needed to be read, but did not explicitly indicate that (and why) it was not necessary to read all books. Built on the answer that reading one or a few books can be read, without further discussion.  
Correctly identified what the population of books was (pile of books) and what it is not.  
Demi: Agreed that not all books needed to be read. Used a student’s suggestion to sample only a few books.  
Alfred: Told not all books needed to be read because that would take too long.  
At first, was unclear about what the population was (1 book, pile of books, school library); later on, more consistent. Regularly pointed at the books. |
| Sampling method | Celine: Asked whether it mattered which books are selected, but students did not provide satisfactory responses. Argued that one book is not representative for the pile of books. By having students take a book from the pile, the books were more or less randomly sampled.  
Asked whether a sufficient number of books was selected in order to balance the contents of the sampled books. Outcome: Each group took books from the pile.  
Demi: Did not discuss how the books would be sampled, but had the students take a book from the class’ bookcase. Argued that a book’s back cover is not indicative of its contents and argued that the number of pages of the book is not relevant.  
Outcome: Each student took a book from the bookcase.  
Alfred: Did not discuss how the books would be sampled, except argued that it is not practically feasible to have everyone reading the same book. Stated without further discussion that he wanted to investigate the inside, not the back cover, of books.  
Outcome: Each student took a book from the desk’s drawer. These books were likely less difficult than the books on the pile of books, which could lead to a biased sample. |
| Sample size | Celine: Her question whether a sufficient number of books was selected in order to balance (out) the contents of the (various) books might indicate she thought the sample size was sufficiently large.  
Argued the number of words per page needed to be comparable across books.  
Outcome: Sample of sufficient size (each group 10 lines from 1 book).  
Demi: Emphasis of discussion was on practical feasibility of sample.  
Outcome: Sample of sufficient size (each group 3 pages from 1 book).  
Alfred: Discussed extensively whether it is possible to compare chapters and pages of different lengths. Emphasized comparability across books. Argued that comparing lines was fair, as these would be all of the same length.  
Outcome: Sample of sufficient size (each student 20 lines from 1 book). |

---

1 The Dutch word ‘*het*’ can be translated as *it*, and it is a neutral article.
## Conclusion phase of the lesson

<table>
<thead>
<tr>
<th>Data as evidence</th>
<th>Celine: Results: <em>the</em> 94, <em>he</em> 89, and 82, <em>a</em> 71, is 48. Based her conclusions on the data.</th>
<th>Demi: Results: <em>the</em> 201, <em>a</em> 194, <em>he</em> 174, “<em>het</em>” 159, is 76. Based her conclusions on the data.</th>
<th>Alfred: Results: <em>the</em> 116, “<em>het</em>” 86, <em>I</em> 84, and <em>het</em> 75, says 16. Based his conclusions on the data.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalization</td>
<td>Celine: Drew a conclusion that pertained to the population. At the end of the lesson, she asked whether it is necessary to read all books, without giving an answer herself. There was consensus among students that all books needed to be read. She did not provide a response and closed the lesson.</td>
<td>Demi: Indicated a number of times that not all books needed to be read. Drew a conclusion that pertained to the population. Indicated that conclusions based on a specific part of the population do not necessarily hold for the entire population and indicated what the population is and what the population is not (series of Dutch children’s novels is only part of the population). During the evaluation of her lesson, explained clearly what a sample is (a small part of the population that is investigated) and claimed the is the most frequently used word in the population, without telling where she got this knowledge.</td>
<td>Alfred: Indicated that not all books needed to be read because that would take too long. Drew a conclusion that pertained to the population. During the evaluation, clearly explained what a sample is (part of the population that is investigated) and that generalization is possible and necessary to save time.</td>
</tr>
<tr>
<td>Sampling variability</td>
<td>Celine: Remarked she thought it “very funny” that the graph showed a descending pattern, meaning that the order in the actual top 5 was the same as in the hypothetical top 5, but she did not discuss this observation any further. Also expected to find the often in a larger sample. At the end of the lesson, she did not provide a reply to a student’s remark that a different sample would result in a different conclusion.</td>
<td>Demi: Asked to what extent the exact order of the hypothesized top 5 was found in the groups’ sample results. Initially argued that a larger or smaller sample would yield the same result because (1) the sample was sufficiently large, (2) the graph clearly indicated that the is the most frequently used word, and (3) different results between the books cancel out when merging samples. During the evaluation of the lesson, claimed she knew that the was the most frequently used word in the population, without making clear how she knew this, to explain that the most frequently used word in the population was the same as in the sample.</td>
<td>Alfred: His second argument why making inferences is possible was that during the process of constructing the pooled graph, the difference between the number 1, the, and the number 2 became larger.</td>
</tr>
<tr>
<td>Sample size</td>
<td>Celine: No remarks.</td>
<td>Demi: Argued implicitly that the used sample was sufficiently large. When a student argued that the sample was sufficiently large and that therefore generalization is possible, Demi only noticed that the class had investigated the question.</td>
<td>Alfred: His first argument for why making inferences is possible was that the collected data cannot be used to infer to a larger population than the school library, such as the municipal library or all books in the Netherlands.</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>Celine: Remarked that the small difference in frequency between numbers 1 and 2 in the dataset makes the conclusion more uncertain. Argued that more certainty can be gained by reading more books.</td>
<td>Demi: In the first instance, was firm in her statements that a larger or smaller sample would yield the same result. Only after a student pointed out the small difference between numbers 1 and 2 did she express herself less firmly and agree with 50% certainty. Her claim that she knew that the was the most frequently used word in the population suggested a belief in the possibility of complete certainty about the inference.</td>
<td>Alfred: Did not make statements related to uncertainty that evidently belonged to his ISI-CK. During extensive gathering of students’ opinions about the certainty of the inference, approved with contradictory statements. For example, he agreed with 50% certainty (because not all books had been read) and with 99% certainty (because other books would probably show the same result).</td>
</tr>
</tbody>
</table>
Results

We identified three themes that constituted major findings about the ISI-CK of Celine, Demi, and Alfred when teaching the ISI lesson. Although these themes were visible in all three lessons, they played out differently in each lesson. The first theme was that the three pre-service teachers showed the ability to use the data as evidence and to make generalizations beyond the data. Second, all three faced difficulties in explaining why making inferences is possible. Third, although the sample size to be used was discussed extensively, the sampling method to be employed remained largely undiscussed.

Using the data as evidence and generalizing to the population
The first theme was that the pre-service teachers used the data as evidence and drew conclusions that went beyond the sample data (see Table 5, conclusion phase, Data as evidence and Generalization). Any reliance on other sources for their conclusions, such as their own experiences or opinions, was absent. Moreover, they drew conclusions that pertained consistently to the population, rather than to the sample data only. Finally, all three were, in general, clear about what comprised the population and the sample; at the end of their lessons, Alfred and Demi provided clear definitions of what a sample is. For example, Demi indicated that a series of Dutch children’s novels was only part of the population of interest. Overall, they acted in accordance with the first two components of ISI, data as evidence and generalization.

The evidence suggests they did not merely follow the lesson plan but instead went beyond, in the sense that they had internalized the ideas behind it and were consciously engaged in making inferences based on sample data. For Celine, several instances during her lesson and during the reflection interview provided proof of this. She consciously used the data as evidence, explicitly indicating that an answer to the research question could be found by reading books. Regarding making generalizations beyond the data, she did not explicitly argue that not all books needed to be read, but various sources indicate that she was aware that not all books needed to be read. First, when asking whether all books needed to be read, she built on a student’s suggestion to read one or a couple of books. Second, it was clear in the reflection interview that she knew she was generalizing to the population and that this was
permissible, as she commented that she had the impression her students understood that not all books needed to be sampled:

...I had the feeling, when they were working with—and they had the idea and—that it really was like: “Yes, we are doing a sample now.”

Demi explicitly indicated that an answer to the research question could be found by conducting an investigation, and she argued that not all books needed to be read to find an answer. Moreover, her engagement in ISI went beyond what was described in the lesson plan. For example, she indicated that a particular series of Dutch children’s novels was only a small part of the population of interest. In addition, her explanation at the end of lesson of what constitutes a sample was not prescribed in the lesson plan:

What we did today was a sample. We took one part from the entire bookcase, and that part we investigated. This was the result, and from that we drew a conclusion. So we said: This is what we have read, and this holds for the entire bookcase.

Alfred’s summary at the end of the lesson provides evidence that he was fully aware that the data could be used as evidence and that inferences were made regarding a population of books:

Sample – that is how we call this. And that means that from this very large pile […] we can see […]: What word occurs most often? Then we are not going to read all books because that takes way too long. So we take small parts, and that is what we did today. We investigated this, we read books, and we saw how often they occurred, and then we saw in this file […] that the word the occurs most often. That is what you do with a sample. You take small parts from the whole, and on this part you let go a research question.

The explanation that they consciously used the data and made inferences is further corroborated by their behavior during the teacher college sessions, as throughout these sessions they showed awareness of the distinction between the sample and the population, and they all generalized to the population.

**Difficulties in explaining why making inferences is possible**
The second theme was that all three pre-service teachers faced difficulties in explaining why making inferences based on sample data is possible. In the context of the lesson “What is the most frequently used word?” such an
explanation requires coordinating arguments based on an understanding of sampling variability with statements related to uncertainty. It may look like the following: When one repeatedly draws random samples, the results will vary, which makes any inference uncertain. However, if the sample is sufficiently large, we may expect that most samples will have the same most frequently occurring word.

For all three pre-service teachers, insufficient mastery of ISI-CK and the translation of ISI-CK into teaching seemed to have played a major role. The way the difficulties played out and how these differences can be understood from context-related factors differed between the pre-service teachers. Therefore, we describe the difficulties for each pre-service teacher in turn, starting with a description of the difficulties and then putting it into context in an attempt to understand why these difficulties occurred.

*Celine: Making an inference when the difference between numbers 1 and 2 is small*

In the first instance, Celine used correct arguments related to the issue of whether the conclusion also pertained to the population (see Table 5, conclusion phase, Sampling variability and Uncertainty). For example, she agreed with a student who had remarked that the small difference between numbers 1 and 2 made the inference more uncertain:

*The* and *he* are very close, so that makes it a bit difficult because if maybe there was a larger difference, then we could have done, then maybe we could say more certainly, right? That is what you meant? Very good.

Building on this observation, Celine encountered problems when she asked for a show of hands to indicate whether all books needed to be read. To her apparent surprise, all students raised their hands to indicate that indeed all books needed to be read, while those initially intending to deny quickly pulled back. One student explained that all books needed to be read because many books that were not read could contain *the* less frequently than the books in the sample. Celine did not know how to reply; after some embarrassing hesitating remarks, she abruptly closed the lesson.

Her failure to respond to the class’s idea that all books need to be read may be called surprising, since her ISI-CK shown during the pretest indicated a thorough understanding of sampling variability in relation to making inferences. For instance, she remarked: “If your spread and your coverage is
sufficiently large of your sample [then] the chance [is] negligibly small that you have different results with a different group.” From the reflection interview, three initial reasons emerge to help understand why she failed to respond to the class’s idea that all books need to be read:

I found it very difficult though, when I was standing there like: “Well, done the entire lesson and now? Am I going to explain it another time that this is just, if you would read more books that kind of you get the same result?” And I also noticed that I had used quite difficult words now and then […] End result, result – those kinds of words. […] So that was a remarkable moment, when all hands went up with: “Read all books.” Then I thought: “No! What now?” But then: Time was already up, and we had already done everything, and it was already on the board, and I thought like: “Well, that was a goal we did not attain.”

The first reason is a pedagogical one: She did not have time left to explain why making inferences was possible. The second reason is that she was caught off guard and lost for words. During the reflection interview, she noted that she had not expected that students would almost in unison respond that all books needed to be read. Both Celine and the observer had the impression that the students understood that not all books needed to be read. A third reason is that she did not believe that repeating the same explanation would have been helpful. Also, as she indicated during the interview, the terminology may have hindered her ability to convince students.

Underlying this third reason may be a deeper reason: Celine did not have a tool at her disposal to explain why making inferences is possible if the difference between numbers 1 and 2 is small. Before a student pointed out this small difference, she had argued that making inferences is possible because a larger sample would probably yield a similar result. This argument would be convincing if the difference between numbers 1 and 2 is relatively large and if Celine had used this argument during the model lesson during the teacher college sessions. However, if the difference between numbers 1 and 2 is relatively small, another argument is needed—one that relates sampling variability to uncertainty. For example, an explanation shows that for sufficient sample sizes, most sample results yield the same number 1, even if the difference between 1 and 2 is small. The teacher educator demonstrated this line of reasoning during the simulation of the law of large numbers (see Table
1, session 1), but in the context of the lesson “What is the most frequently used word?” such an explanation was not offered to her. The lack of an explanatory tool in the context of the lesson she was presenting may have been the ultimate reason Celine had difficulty explaining why making inferences is possible.

Demi: Neglecting the uncertainty inherent to inference

Although Demi used correct arguments based on knowledge about sampling variability to explain why making inferences is possible (see Table 5, conclusion phase, Sampling variability), she also showed a strong tendency to neglect the uncertainty involved in her arguments (see Table 5, conclusion phase, Uncertainty). Her language and tone were characterized by a decisiveness that was not appropriate given the uncertainty involved in making inferences. One example of her negligence of uncertainty is her unwarranted emphasis on the exact order within the top 5 (see Table 5, introduction phase, Sampling variability). Another example is her disapproval of a student’s request to sample another book in order to increase the certainty:

Take another book? But then at a certain point you have had all books, and just at the beginning we decided we don’t need to read all books to find it out.

Furthermore, Demi did not notice the small difference between numbers 1 and 2, which would have made her less certain about her inference (see Table 5, introduction phase, Uncertainty). Only after a student pointed out that this small difference made the inference more uncertain did Demi express herself less confidently, but she did not discuss the student’s reasoning:

Student: But miss, but [...] a is almost on top. So if you read more books and the goes badly and a goes all the way ahead, then a is on the top position. It’s approximate, right?

Demi: All right, so how certain are we the is number 1?

A similar tendency to express herself overconfidently and too decisively was observed at the close of her lesson, when she claimed that she knew that the was the most frequently used word in the population, the class’s bookcase, without making clear how she knew this (see Table 5, introduction phase, Generalization, Sampling variability and Uncertainty):

Demi: OK, if I tell you the is the most frequently used word. So I say, the is the most frequently used word and I tell that if we
would read more books, *the* is still in position 1. [...] But when you know this and think like, if I make the number of books larger, will the number 1 remain the same?

Student: No.

Student: I don’t understand.

Demi: I tell you that if we would read the entire bookcase [...] , *the* will still be in position 1.

Student: No.

Student: Yes.

Demi: It is. [...] 

Student: Yes, but I don’t get it. Yes, I don’t get it.

Demi: OK, one last time, I will try once more: We have a bookcase, this is the bookcase. We tested this small part of the bookcase, so we looked into this part to see which words occur most often in this part. So far we are all right? Here, this is the result: *The* is in position 1, and *a* is 2. If we would read the entire bookcase—of course, we won’t—then we get the same result.

Student: Huh?

Demi: Then we get the same result, then *the* is still in position 1.

During the reflection interview, Demi explained that her claim that *the* was the most frequently used word in the population was just a slip of the tongue: She intended to argue that, if in the sample *the* was the most frequently used word, one could assume that *the* was also the most frequently used word in the population. Nonetheless, this episode shows how she used non-probabilistic language ("...then we get the same result," "It is," "I tell you ...") and a decisive tone that to want to press an explanation onto the students. It seemed that she wanted to convey the message that making inferences is possible because another sample *will*—rather than *might*—yield a similar outcome.

One possible reason for why she neglected the role of uncertainty is that Demi, like Celine, lacked the arguments to answer the question of why making inferences is possible. During the reflection interview, she mentioned her insufficient mastery of ISI-CK. During the model lesson at the teacher college,
she missed an answer to the question “what is it in the end?” (i.e., why making inferences is possible). She wanted to provide an answer for her students, and therefore, in preparing the lesson, she had tried to get a firmer grasp of this issue: “Yesterday, I’ve been reading the lesson plan over and over again so I would have certain answers, so I would know what it is.” Apparently, she did not find a satisfactory answer to the question of why making inferences is possible.

In the absence of a suitable explanation for why making inferences is possible, Demi resorted to evading altogether any discussion about the uncertainty of inference. This might be related to a tendency to have precise answers, as the above quote suggests, and some support for this is found in the pretest of the intervention. At that time, she still agreed that no claims at all can be made about the population: “By samples, you never know precisely because you could precisely miss the group whose opinions are totally different from other groups.” The word “precise” illustrates her tendency to look for certain answers. The combination of a lack of tools and a tendency to aim for precise answers may have made her neglect the role of uncertainty and made her claim that another sample will yield a similar outcome.

Although Demi’s lesson appeared to be very different from Celine’s lesson, a common factor might have been at play: Both might have lacked the tools to convincingly explain “how it works” (i.e., why making inferences is possible). While Celine ultimately fell silent, Demi, in her attempt to explain to her students how inference works “in the end,” neglected the inherent uncertainty in inferences.

*Alfred: Incomplete and wrong explanations and gathering responses without purpose*

Alfred encountered two problems in the discussions of why making inferences is possible. First, he used both correct and incorrect arguments, and his overall argumentation was incomplete. Second, the discussions related about the certainty of inferences lacked purpose.

While discussing the first statement (“I can’t say anything at all about which word is most frequently used in the pile of books because we haven’t checked all books”), Alfred supported his explanation of why not all books needed to be read with both an incorrect and a correct argument. His first argument was both wrong and contra-productive (see Table 5, conclusion
phase, Sample size): He claimed that the used sample could not be used to
generalize to a larger population than the one used, such as the books in the
municipal library or all children’s novels in the Netherlands:

Alfred: But suppose, we have incredibly many books because […]
when you go to the town library, it is many times larger. And
there are also many more books in the whole of the
Netherlands. […] Can we then ask which word […] occurs
the most often, if we do it like this?

Student: No.

Alfred: No.

He thus showed the misconception that the sample size depends on the
size of the population size—something he had argued during the teacher
college intervention several times as well. Worse still, even if this idea were
correct, it was an argument against why all books did not need to be read, as
the idea might suggest that even more books need to be read than the ones
sampled.

His second argument was correct (see Table 5, conclusion phase, Sampling
variability): He observed that when he was entering the sample results of the
various groups into the pooled graph, it became increasingly clear that *the*
was the most frequently used word in the sample.

…still I want an answer to my question. […] So do we have to read all books to
see which word occurs most often? We just saw that *the* becomes higher and
higher, more and more, and that it protruded above the rest. […] Do we really
need to read all books?

Alfred acknowledged that the size of the difference between the first and
second most frequently use words in the sample made it more likely that *the*
would also be the most frequently used word in the population. An argument
missing in his explanations was the idea that if the sample is sufficiently large,
one can expect that most samples will have the same most frequently occurring
word.

Second, the lack of purpose was prominently visible in the discussions
about the certainty of the inference based on a sample (see Table 5, conclusion
phase, Uncertainty). After having made the graph of the pooled data and
reading from the graph, he discussed the certainty of the conclusions. None of
his statements related to uncertainty evidently belonging to his ISI. Rather, his approval of contradictory statements related to uncertainty seemed to have a pedagogical goal. He abstained from any value judgment when one student argued that another sample of sufficient size would likely yield a similar result, or when another argued 50% certainty because not all books were read and each book contained different words, while these are conflicting views. In the same vein, he even affirmed contradictory answers when discussing the certainty of the generalization. For example, he agreed with 50% certainty because not all books have been read, with 99% certainty because other books would probably show the same result, and with 100% certainty because the students had just conducted the research themselves.

During the reflection interview, Alfred himself gave a reason why this discussion, but also the discussion of the sample size (see next subsection), went on so long: “I think that has to do with the fact that I find this topic so difficult myself, that actually I cannot explain it well, that I cannot approach from multiple angles, respond quickly to what students say.” Thus, his lack of ISI-CK hindered his ability to respond appropriately to the students. He indicated that he was able to apply a strategy of eliciting useful approaches from the students for content he knew well, such as grade 3 arithmetic, although he also said he was still learning when to steer and when to let them go.

The finding that he did not know the aim of the discussion is further supported by the fact that during this part of the lesson he was holding the lesson plan in his hands but seemed to look for what was coming next. His problems with the aims of discussing the certainty of the inferences is also in line with the comments of several pre-service teachers during the teacher college intervention who indicated having problems with expressing certainty.

Alfred’s explanation that he did not master ISI-CK sufficiently enough to teach about ISI lends further support to his misconception that the sample size depends on the population size, though he did not mention the idea that another sample would probably yield a similar result. He did, however, point out the large difference between numbers 1 and 2. During the model lesson, he had observed this large difference as well. At that time, he remarked that this large difference made him more certain about the inference.
Alfred’s lack of ISI-CK is a plausible explanation for why he had trouble steering the discussion. Demi showed herself to be an average pre-service teacher in terms of ISI-CK, but in preparing the lesson, she tried to find answers to issues she did not know. In contrast, had almost completely copied the teacher educator’s lesson plan into his own lesson plan. His effort in internalizing the lesson’s flow and underlying ideas may have been insufficient, and thus it was unclear where the discussion should head.

Conclusion of the second theme

All three pre-service teachers had problems explaining why making inference is possible. Celine fell silent, Demi neglected the inherent role of uncertainty in inferences, and Alfred’s lesson lacked purpose. These problems could be attributed in part to their insufficient mastery of ISI-CK. This was one reason given by the pre-service teachers themselves. Their ISI-CK lacked the tools to show the workings of sampling variability by means of combining growing and repeated samples in the context of this particular lesson. Such a tool could had been used to show that large samples often yield similar results that allow for making inferences with quite some certainty. However, although the lesson plan suggested discussing whether another sample would yield a different result, it did not explain why and when it would yield a similar result and how this made making inferences possible. This lack of tools was most prominently visible in Celine’s lesson, but it may also have been the reason why Demi neglected the role of uncertainty, since she had searched the lesson plan for such a tool while preparing her lesson.

Much attention for sample size, little attention for sampling methods

The third theme that we identified was that during the design phase of the lesson, when the class had to decide on a method to draw a sample from the pile of books, discussions focused on how to collect a sample within limited time, while the sampling method to be used was left largely undiscussed.

In all three lessons, the discussion about how many books and pages to survey resulted in a sample of sufficient size that could be collected within limited time (see Table 5, design phase, Sample size). This raises the question, however, of whether the pre-service teachers also thought the sample was sufficiently large for allowing inferences. For Celine, the evidence seems to be
strongest that she thought the sample to be sufficiently large. Her question of whether they had sampled enough books to balance out the differences between books suggests that she understood how to obtain a representative sample and that she thought the number of books was sufficiently large. The latter argument was also used by Demi. However, during the teacher college sessions and the pretest, neither Celine nor Demi showed evidence that they knew when a sample is sufficiently large.

Alfred did not make any comments related to the sufficiency of the sample size used. In his lesson, the discussion about how many books and pages to survey was characterized by lengthy sessions of gathering suggestions. After some of these gatherings, he even made clear that he did not think the approach to be fair after all. For example, he asked numerous students how many pages to sample and eventually concluded that using pages was not a fair strategy. Finally, he suggested himself that each group could sample 20 lines, arguing that this was fair since each line contains about the same number of words.

During the reflection interview, Alfred said that he hoped the students would come up with useful suggestions concerning the sample size, but because he had not mastered ISI, he had problems valuing their suggestions on the spot and effectively steering the discussions in the direction he wanted.

Celine was the only one who tried to discuss the sampling method by asking whether it mattered which books were to be selected (see Table 5, design phase, Sampling method). Just before her students collected the data, she raised the issue of which books to sample. However, her students did not provide satisfactory responses to her question. Therefore, she did not pursue this direction any further and decided to move on. In addition, the actual way the books were sampled resulted in de facto random sampling; as she remarked during the reflection interview: “I had tried to discuss it by asking, but they did not really go into it, and then I thought like: ‘Well, just pick a random book, then we just leave it like that. It has been enough.’” The reason she gave thus signals a PCK consideration not to discuss the sampling method any further.

Likewise, in Demi’s lesson the sampling method resulted in de facto random sampling, as she had students take a book from the class’s bookcase. The only remark related to representativeness was her rhetorical question of
whether the words used on the back of a book are representative of those used within. Alfred did not discuss the sampling method to be used. Rather, he had students use the books in their desks drawers. As the books on the pile were likely more difficult than these books, the resulting sample could be biased. He indicated that he did not want to use the back covers of the books, but he did not provide a reason for this.

The emphasis on the sample size and the lack of attention for sampling methods may be called surprising, given that during the model lesson at the teacher college the sampling method was discussed at length and all three pre-service teachers stressed the importance of a reliable sample during the teacher college sessions. We think the most reasonable explanation for the lack of attention to the sampling method is that discussing the sampling method was not indispensable for the progression of the lesson, while discussing the sample size was. The choice of a particular sampling method was of no obvious importance: In whatever way books were selected, most strategies were practically feasible and potential bias would remain hidden. In contrast, a discussion of the sample size to be used was necessary for the progression of the lesson, as the data needed to be collected within about ten minutes. The pre-service teachers taught ISI for the first time—a topic they did not sufficiently master themselves. Therefore, they may have been focused on coming through the lesson, which made them less able to reflect on the importance of the sampling method. A second explanation might be that the lesson plan did not support the teachers in discussing the sampling method; it also focused more on the sample size than on the sampling method. Finally, the discussion of the sampling method did not get off the ground in Celine’s lesson because this issue might be beyond the reach of grade 3 students.

Ultimately, the way the sample size was discussed and the absence of discussions about the sampling method suggest that the primary concern of the pre-service teachers was to agree on a sampling design that would facilitate the smooth operation of the lesson, instead of purposely steering the discussion toward a sampling design that could yield a representative sample.
Discussion and Conclusion

Our study of three pre-service primary school teachers’ expressions of ISI-CK while teaching an ISI lesson in upper-primary school showed their ability to make inferences from collected data when introducing primary school students to ISI. This result is encouraging, since it showed that pre-service teachers who have limited experience with ISI are able to express their newly acquired ISI-CK in teaching. However, all pre-service teachers faced difficulties in explaining why making inferences is possible. In addition, although the sample size to be used was discussed extensively, the sampling method to be employed remained largely undiscussed.

Discussion of main findings

In contrast to previous research, a major finding was that the pre-service teachers were actually engaged in ISI. This use of data as evidence is found in other studies as well (De Vetten et al., 2018, in press-a; De Vetten, Schoonenboom, Keijzer, & Van Oers, in press-b; Leavy, 2010), but the literature shows that it is difficult to engage pre-service teachers (Leavy, 2010) and students in general (Makar & Rubin, 2009) in ISI. We argued that the reason why the pre-service teachers in our study were engaged in inferential reasoning was probably due to the design of the lesson provided them. The lesson included a research question that was closely linked to a visible population and required little descriptive analysis. Thus, we ensured that the lesson contained sufficient affordances to engage in inferential reasoning, in contrast to when pre-service teachers design lessons themselves (Chick & Pierce, 2008; Leavy, 2010). We also showed that the pre-service teachers not only stuck closely to the lesson plan but went beyond it, in the sense that they internalized the ideas behind it and were consciously engaged in making inferences based on sample data.

The difficulty in explaining why making inferences is possible could be partly due to the lack of tools available in the lesson “What is the most frequently used word?” Part of this explanation could be due to using the lesson’s element of having small groups compare their sample results, but this step was omitted both during the modeled lesson at teacher college and in the pre-service teachers’ lessons. This element could be extended to include a
combination of the growing samples heuristic (Bakker, 2004) and repeated sampling (Saldanha & Thompson, 2002). The results of multiple smaller samples would be compared, and likewise for larger samples. For small samples, the proportion of samples that yield the same most frequently used word would be smaller than for large samples. This activity would illustrate the law of large numbers and could thus provide a tool to explain why making inferences is possible.

Demi’s case points at the role of agency when she, as a teacher, had to use her own content knowledge in preparing and teaching a lesson on new subject matter (Engeström, 2011). Agency can be defined as “the actual ways situated persons willfully master their own life” (van Oers, 2015, p. 19). Her ISI-CK during the teacher college intervention was average, but the reflection interview sketched a picture of a teacher who took the responsibility for the personal version of the lesson plan. She seriously reflected on her own (lack of) ISI-CK and on the PCK of ISI. Moreover, she seemed motivated to use the support offered to her to find an explanation for why making inferences is possible. Although she ultimately failed to find such an explanation, her case is an example of how her agency helped her to prepare an ISI lesson, to the end of giving her students good explanations. Future research may, therefore, consider ways how the agency of (pre-service) teachers who are required to teach new curriculum matter should be fostered.

Our study suggests that the complexity of teaching an ISI lesson may be partly due to the inquiry-based teaching approach. In guiding the co-construction of knowledge, the teacher needs to respond adequately to students’ suggestions and questions, to steer the discussion, and to provide convincing explanations to unforeseen questions (Dobber, Zwart, Tanis, & van Oers, 2017). The problems encountered by the three pre-service teachers are therefore not surprising, since such teaching skills are among the most difficult to conduct well (Rowland & Turner, 2007), and even an experienced and very competent teacher can encounter many problems in teaching statistical investigations (Mickelson & Heaton, 2004). One possible remedy could be to make the lesson less open—for example, by discussing the sampling method using a menu of options students can choose from. However, this might come at the cost of losing some experience of being involved in the real process of conducting an ISI investigation (Mickelson & Heaton, 2004).
Regarding the finding that the sampling method and the sufficiency of the sample size remained largely undiscussed, it can be questioned whether this constitutes a serious problem during an introduction to ISI. For this introduction, two elements may be essential. First, students need to be engaged in the act of collecting sample data and making generalized claims about the population. Second, the activities need to foster an understanding of why making uncertain inferences is possible. As long as the students trust the representativeness of the sample, it might not be necessary to discuss extensively the representativeness of the sampling method and the sample size (Cobb & Tzou, 2009). Future research could investigate the effects of ISI lessons that pay less attention to sampling issues on the students’ trust in the sample representativeness and on their willingness to make inferences based on the sample data.

The preceding discussion of the results allows us to compare how the pre-service teachers’ ISI-CK was expressed during teaching with how this knowledge was expressed in the context of the teacher knowledge intervention. First, their behavior in terms of using the data to make inferences was similar to their behavior during the teacher college sessions. Second, for some concepts (the sampling method, for example), the pre-service teachers’ ISI-CK was less visible than in the teacher college context, partly because it appeared not to be necessary to discuss these concepts. Third, teaching ISI revealed the pre-service teachers’ lack of depth regarding some ISI-CK components. For example, although they were able to tell during the teacher college intervention what they had learned from a computer demonstration of a simulation of the law of large numbers, they were not able to transfer what they had learned from this demonstration about sampling variability to the ISI lesson they taught themselves. In addition, the problems the pre-service teachers faced with expressing uncertainty, which were already expressed by some participants during the teacher college sessions, now became urgent, as an understanding of uncertainty is necessary to accept that uncertain inferences can be made.

**Limitations**

In interpreting these findings, several limitations must be kept in mind. First, we investigated the expression of ISI-CK during teaching ISI in a context where...
the pre-service teachers had limited time to learn ISI and to practice teaching ISI (De Vetten et al., 2018), while teaching a subject well requires a thorough understanding of the subject and, thus, much experience (Hill et al., 2008). In the case of ISI, Leavy (2010) pointed out that the teachers’ unfamiliarity with conducting statistical investigations themselves hindered their teaching of ISI. Also, even though the three pre-service teachers in our study had average to good ISI-CK, their experience with ISI remained limited, and they themselves expressed concerns about their mastery of the subject. Future research could investigate whether similar challenges are faced by teachers with more ISI experience. Such research could use designs that allow for multiple teaching opportunities and experiences, such as Lesson Study designs (Groth, 2017; Leavy, 2010).

A second limitation is that the ISI-CK expressed by the pre-service teachers depends on their pre-existing ISI-CK and the specific support offered them during the intervention. These factors limit the generalizability of the findings. However, our descriptions of the support of the lesson plan and of the ISI-CK the pre-service teachers expressed during the teacher college intervention may help other researchers to estimate the extent of congruence between our context and their context (Lewis & Ritchie, 2003).

A third limitation might be that the pre-service teachers’ behavior could have been influenced by their observer’s dual role as their teacher educator. This could have influenced their ISI-CK positively if the pre-service teachers put in greater effort than without their teacher educator being present, or negatively if it resulted in more stress. However, no indications were found—for example, from the reflection interview—that either of these influences played a large role.

Conclusion
In conclusion, our approach to investigating the ISI-CK of three pre-service teachers while teaching an ISI lesson proved to be a useful one, as this approach clarified the elements of ISI-CK they understood well and those they did not. In particular, it became clear that they encountered problems in explaining clearly why making inferences is possible. This issue thus deserves attention in ISI education for (pre-service) teachers. They must be offered multiple opportunities to be engaged in activities that illustrate the law of large
numbers, while these activities should also include ones that resemble lessons they can use to introduce primary school students to ISI. Our findings suggest that these lessons should contain tools that (pre-service) teachers can use to explain why making inferences is possible. More research conducted in primary classrooms is needed to determine whether these tools indeed allow (pre-service) teachers to provide such explanations. The ultimate interest would be in research that investigates whether these explanations indeed help primary school students to understand that making inferences based on sample data is possible and help them to find the balance between knowing nothing and knowing everything (Rubin, Bruce, & Tenney, 1990).
References


CHAPTER 5


