PRE-SERVICE PRIMARY SCHOOL TEACHERS' KNOWLEDGE DURING TEACHING INFORMAL STATISTICAL INFERENCE

ARJEN DE VETTEN
Leiden University, the Netherlands
a.j.de.vetten@iclon.leidenuniv.nl

RONALD KEIJZER
iPabo University of Applied Sciences, the Netherlands
r.keijzer@ipabo.nl

JUDITH SCHOONENBOOM
University of Vienna, Austria
judith.schoonenboom@univie.ac.at

BERT VAN OERS
Vrije Universiteit Amsterdam, the Netherlands
bert.van.oers@vu.nl

ABSTRACT

The study reported in this article investigated the appropriateness of Mathematical Knowledge in Teaching of three pre-service primary school teachers (PSTs), teaching an informal statistical inference (ISI) lesson to primary school students. Using an ISI framework and the Knowledge Quartet framework, the presence and appropriateness of the PSTs' teaching actions were coded and categorized. The results showed that PSTs were consciously engaged in making inferences based on sample data. The PSTs struggled to correctly interpret students' conceptual input and to explain ISI, in particular, how generalizing from a sample is possible. Teacher education should focus on how PSTs can foster students' understanding of the logic of drawing conclusions about a population based on a sample.

Keywords: *Statistics education research; informal statistical inference; primary education; initial teacher education; teacher knowledge*

1. INTRODUCTION

Data and statistics have become pervasive in our society (Gravemeijer et al., 2017) and individual citizens are required to engage in data-based decision making. Often, this decision-making involves inferential reasoning and critically evaluating inferences from empirical data (Liu & Grusky, 2013). To prepare children for their roles in society, primary school students need to learn to reason inferentially. One way to introduce primary school students to inferential reasoning is by means of informal statistical inference ([ISI]; Groth & Meletiou-Mavrotheris, 2018), which is defined as constructing "a generalized conclusion expressed with uncertainty and evidenced by, yet extending beyond, available data" (Ben-Zvi et al., 2015, p. 293). Previous studies have found that primary school students are able to begin reasoning about inferences (Leavy & Sloane, 2017; Makar, 2016). In some countries, such as Australia and New Zealand, introductions to ISI are part of primary education, while other countries, such as the Netherlands, have plans to add ISI to the curriculum (curriculum.nu, 2019). To introduce primary school students to ISI, their teachers need to learn how to conduct this introduction. Therefore, it is important to understand how primary school teachers can be supported to introduce their students to ISI.

Statistics Education Research Journal, 22(2), Article 12. https://doi.org/10.52041/serj.v22i2.424

In a previous study, we examined how pre-service primary school teachers' ISI knowledge for teaching developed during a short teacher college intervention (De Vetten et al., 2018). Since the transfer of knowledge acquired in a teacher college setting to an actual teaching situation is not straightforward (Ball et al., 2001; Fennema & Franke, 1992; Rowland et al., 2009), we also need to evaluate teachers' ability to mobilize their ISI knowledge during teaching. To date, there has been no detailed investigation of teachers' ISI knowledge during teaching. The aim of the present article is to investigate the extent to which pre-service teachers (PSTs) are able to show appropriate ISI knowledge during teaching an ISI lesson in their primary internship classrooms.

2. THEORETICAL BACKGROUND

It is generally held that teaching mathematics, of which ISI is a subdomain, requires general pedagogical knowledge and Mathematical Knowledge for Teaching ([MKT]; Ball et al., 2008; Blömeke et al., 2011; Schmidt et al., 2008; Shulman, 1987). Pedagogical knowledge, as conceived in this article, is general knowledge about teaching strategies that is independent of the subject matter. Our study focuses on MKT, which consists of content knowledge (CK) and pedagogical content knowledge (PCK). CK is knowledge of the mathematics itself; PCK is the knowledge of how to teach mathematics, such as appropriate examples, representations and explanations to the topic at hand, and knowledge of the difficulties students might encounter and the (mis)conceptions they may hold (Shulman, 1987). Following Charalambous et al. (2020), we investigated PSTs' CK and PCK synergistically.

While usually teachers acquire their MKT during initial teacher education, it cannot be assumed that teachers automatically know how to use their MKT in actual teaching (Ball et al., 2001; Putnam & Borko, 2000; Rowland et al., 2009). Teachers also need to be able to "unpack" their MKT for teaching purposes (Ma, 1999). Unpacking MKT in a teaching setting appears not to be straightforward, as "Thompson and Thompson (1994) show[ed] 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" (Hill et al., 2008, p. 435). The complexity of using one's MKT in teaching stems from the more demanding requirements of teaching compared to professional development settings (Ball et al., 2001; Mickelson & Heaton, 2004). For instance, teaching requires teachers to react immediately to unexpected remarks, questions, and events. Therefore, studying teachers' MKT in the context of teaching can reveal what MKT they are able to mobilize during teaching (Ball et al., 2001; Fennema & Franke, 1992; Rowland et al., 2009). We label this as Mathematical Knowledge *in* Teaching ([MKiT]; Weston, 2013). MKiT is informed by teachers' MKT and is displayed in observable ways teachers use their MKT (Auletto & Stein, 2020; Rowland et al., 2005).

To study PSTs' MKiT, we used the Knowledge Quartet framework ([KQ]; Rowland et al., 2005). The KQ is comprised of four dimensions of teachers' MKiT: foundation, transformation, connection, and contingency; each dimension consists of four to seven aspects (see Appendix 1). Using the coding protocol developed by Hill and colleagues (2008), we coded both the presence and appropriateness of teaching actions, where appropriate teaching actions are those that help the lesson move towards attainment of the learning objectives, while inappropriate teaching actions are those that significantly hinder the attainment of the learning objectives (Learning Mathematics for Teaching, 2006; Weston, 2013).

In order to describe the PSTs' MKiT of ISI (ISI-MKiT), we used the ISI framework by Makar and Rubin (2009). As in our previous studies (De Vetten et al., 2018; De Vetten et al., 2019a), we conceptualized the three components of this framework 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 thus, under certain circumstances, a sample can be used for an inference (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 suggested a need to develop (pre-service) primary school teachers' CK of ISI (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., 2019a; De Vetten et al., 2019b; Meletiou-Mavrotheris et al., 2014; Mooney et al., 2014; Watson, 2001). Furthermore, evidence showed PSTs lack awareness that ISI tasks require an inference over and above a descriptive analysis of the data, while 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., 2019a, 2019b).

Previous studies, including two of our own, focused on the development of pre-service primary school teachers' ISI-CK by employing teacher college interventions (De Vetten et al., 2018, 2019b; Leavy, 2010). The only study to date that investigated ISI-MKiT is Leavy (2010). The PSTs in that study designed ISI lessons themselves, and even though the final-year PSTs specialized in mathematics education and had strong ISI-CK, they faced difficulties in designing a lesson that contained sufficient affordance to discuss ISI. The resulting lessons, therefore, provided few opportunities to study PSTs' ISI-MKiT. There has thus been no detailed investigation of ISI-MKiT.

As part of our second teacher college intervention (De Vetten et al., 2018), the participating PSTs taught an ISI lesson in their placement schools. As previous research has revealed the difficulties PSTs face in designing a lesson that contains sufficient affordances to discuss ISI (Chick & Pierce, 2012; Leavy, 2010; Makar & Rubin, 2009), and since Dutch primary teacher education usually devotes little time to statistics and none to ISI, we provided the PSTs with a lesson plan containing sufficient affordances to teach ISI. This article describes the ISI-MKiT of three PSTs, who participated in the teacher college intervention, while teaching ISI to primary school students. The results provide initial evidence of how teacher education can support PSTs to introduce primary school students to ISI. The research question is:

> What ISI-MKiT is present in the teaching of pre-service teachers' ISI lessons in primary school and what is the appropriateness of their ISI-MKiT?

3. METHOD

3.1. CONTEXT AND PARTICIPANTS

This study reports on a multiple case study of three second-year PSTs—Celine, Demi, and Alfred (pseudonyms)—as they taught an ISI lesson to primary school students. These PSTs were among 21 participants of an intervention at a teacher college in a large Dutch city. These three PSTs were selected based on their sufficient to strong ISI-CK during the intervention's pre-test (De Vetten et al., 2018), their ability to openly express their opinions and concerns, and their willingness to participate. Table 1 shows an overview of this intervention; a detailed description of the intervention and the results can be found in De Vetten et al. (2018). Celine taught her ISI lesson between the third and fourth meeting of the intervention; Alfred and Demi between the fourth and fifth meeting. Celine (female; 18 years; showing high ISI-CK as compared to the other participants during the teacher college intervention) taught in third grade, Demi (female; 19 years; average ISI-CK) in fifth grade, and Alfred (male; 20 years; high ISI-CK) in a combined third/fourth grade class. During the teacher college intervention, their ISI-CK was relatively strong for the components Data as evidence, Generalization beyond the data, Sampling variability, and Uncertainty. It was relatively weak for Sampling method and Sample size. All parties gave informed consent. The study design was approved by the ethical board of the Faculty of Movement and Behavorial Sciences of Vrije Universiteit Amsterdam.

Table 1. Overview of Teacher College Intervention

Week	Setting	Activity
1	Homework	Homework assignment: Samples in the media
3	Session 1	Discussion of homework (60 minutes) Real-time computer simulation illustrating law of large numbers (20 minutes)
5	Session 2	Reiteration of learning points simulation (10 minutes) Teacher educators models lesson: “What is the most frequently used word in a stack of children novels?” (70 minutes)
5	Session 3	Car choice activity (20 minutes) Discussion of ISI-PCK (45 minutes)
6–11	Placement Schools	First half of the participants (including Celine and Demi) teach ISI lesson
12	Session 4	Evaluation of ISI lessons in placement schools (30 minutes)
13–15	Placement Schools	Second half of the participants (including Alfred) teach ISI lesson
16	Session 5	Evaluation of ISI lessons in placement schools (15 minutes)

3.2. LESSON “WHAT IS THE MOST FREQUENTLY USED WORD IN A STACK OF CHILDREN NOVELS?”

The PSTs all taught the lesson “What is the most frequently used word in a stack of children novels?” (see Appendix 2 for the learning objectives and the lesson plan). The lesson was modelled during the teacher college intervention (see Table 1, Session 2), after we had successfully piloted the lesson in primary classrooms. The lesson centered on a large collection of Dutch children’s novels. The driving question was which word would be the most frequently used in the collection of books. The enormity and visibility of the population was expected to elicit the need to draw a sample and make inferences, and, since schools usually have a school library, all PSTs could easily implement the lesson. The investigation was based on five words that the class expected to find most frequently. Class discussion was used to reach consensus about the preferred sampling method, usually scanning in small groups a limited number of lines or pages for the five words, 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 the frequency needed to be determined. The subsequent discussion dealt with the possibility and certainty of generalization from sample results, both from the individual groups’ sample data and from the pooled data.

The lesson plan included a suggested sequence of activities and questions to ask the students. Since the lesson used the concept of dialogic classroom talk to elicit and discuss students’ suggestions (Wells, 1999), the lesson provided considerable leeway for PSTs to steer the direction of the lesson. Table 2 provides an overview of relevant characteristics of the PSTs’ lessons.

Table 2. Overview characteristics of PSTs’ lessons

Characteristic	Alfred	Celine	Demi
Grade Level	3/4	3	5
Number of Students	26	28	27
Time Spent on Class Discussions and Making a Pooled Graph (min)	54	25	44
Time Spent on Group Work (min)	14	16	15
Total Duration (min)	68	41	59
Number of fragments	13	13	22
Number of fragments including ISI discussion	11	12	18

3.3. DATA COLLECTION AND ANALYSIS

Two kinds of data were collected. The first data source were video recordings of the classroom interactions. The first author, who was also the teacher educator during the teacher college intervention, was present as an observer. The second data source was comprised of transcripts of audio-taped reflection interviews between the PST and the first author, which were conducted shortly after the lessons. The discussions during these interviews included evaluations of the extent the learning objectives were attained by the students, explanations of the PST's and students' conduct during the lesson, and checking interpretations made by the observer. Additionally, PSTs' lesson plans, students' written work, and the observer's notes were used as data. All data from the class discussions and the reflection interviews were transcribed.

As the unit of analysis, we chose a lesson fragment, consisting of a whole class discussion concerning one substantive topic, because such meaningful units allow for descriptions of the processes implemented in class (Schlesinger & Jentsch, 2016). Table 2 shows the number of fragments for each lesson. Each fragment was coded using both the ISI and the KQ frameworks. The coding process from the ISI framework used a process consisting of deductive and inductive elements. On the inductive side, short summaries of the text were attached as codes to the text to describe the content of the PSTs' actions. These codes were subsequently combined into groups with similar meanings or issues. On the deductive side, the ISI framework was used to categorize the codes into the ISI components.

The coding process from the KQ framework followed the approach of Learning Mathematics for Teaching (2006) and focused on the presence and appropriateness of teaching actions. First, for each fragment the presence of each of the 20 KQ codes was coded (present versus non-present). Second, the appropriateness of the (non-)presence of teaching actions was evaluated (appropriate versus inappropriate), given the classroom setting and the context of the entire lesson. The reflection interviews were crucial for coding the appropriateness of teaching actions, as the interviews revealed the considerations of the PST to act in a particular way, and helped to understand how during the specific teaching action the PST interpreted the situation and the students' understanding at that point of time. Present teaching actions were coded as appropriate when these teaching actions helped the lesson move towards attainment of the learning objectives. Absent teaching actions were coded as appropriate when this absence did not hinder the attainment of the learning objectives. Present teaching actions were coded as inappropriate when these teaching actions hindered the attainment of the learning objectives. Absent teaching actions were coded as inappropriate when these teaching actions were essential to move the lesson towards attainment of the learning objectives. Six KQ codes were excluded from analyses, as these codes were only present in at most one fragment. The analyses are based on the remaining 14 KQ codes. As the proportions of presence and appropriateness of the various KQ codes within the dimensions differed, we report the results only at the KQ level, and do not report the aggregated results at the KQ dimension level.

The analyses were conducted in *atlas.ti*, *Excel*, and *R*. They involved (cross-)tabulating the codes from the KQ and ISI frameworks and searching for patterns and notable results. The explanations given by the PSTs regarding their conduct during the lesson were triangulated with the researcher's observations. To understand why the teachers demonstrated particular ISI-MKiT, evidence was interpreted from the following context-related factors: (1) ISI-CK displayed during the teacher college intervention, both during pretest and sessions; (2) lesson design factors; and (3) pedagogical considerations, such as decisions related to classroom management. To validate the coding process, the first and second author independently coded about 10% of the fragments and discussed the results until consensus was reached about the coding results and the results' validity.

4. RESULTS

Table 3 provides an overview of the presence and appropriateness of the PSTs' teaching actions. The overall appropriateness of teaching actions was 81% for Alfred, 93% for Celine, and 84% for Demi.

Table 3. Proportion of appropriate present and not present teaching actions[1]

KQ code	Appr total	Present-appr.			Not present-appr.			Present-inappr.			Not present-inappr.		
		Alf	Cel	Dem	Alf	Cel	Dem	Alf	Cel	Dem	Alf	Cel	Dem
Use of mathematical terminology (ut)	1	.09	[2]	.22	.91	1	.78[3]						
Use of instructional materials (uim)	.98	.64	.83	.22	.36	.17	.72			.06			
Decisions about sequencing (ds)	.95	.09		.06	.91	1	.83			.11			
Choice of representations (cur)	.95	.27	.17	.11	.64	.83	.83			.06	.09		
Deviation from agenda (da)	.90	.09			.91	.92	.83		.08				.17
Adherence to textbook (atb)	.85	.09		.06	.82	.92	.72	.09	.08	.22			
Responding to students' ideas (rsi)	.76	.64	.75	.78		.08		.36	.17	.22			
Identifying pupils' errors (ie)	.90	.45	.50	.72	.27	.42	.28				.27	.08	
Teacher demonstration (dt)	.80	.36	.25	.06	.36	.58	.78			.11	.27	.17	.06
Recognition of conceptual appropriateness (rca)	.83	.45	.67	.50	.27	.25	.33				.27	.08	.17
Awareness of purpose (ap)	.78	.73	.92	.72							.27	.08	.28
Anticipation of complexity (ac)	.78	.18	.42	.33	.45	.58	.39				.36		.28
Overt display of subject knowledge (osk)	.78		.08		.73	.83	.72				.27	.08	.28
Making connections between concepts (mcc)	.73	.18	.08	.17	.45	.75	.56				.36	.17	.28

[1]Alfred 11 fragments, Celine 12 fragments; Demi 18 fragments. [2]Cells with 0% are left empty. [3]Grey-shaded blocks of cells highlight trends discussed in the text.

Closer inspection of Table 3 reveals some trends (shaded in Table 3). First, mathematical language (KQ code *ut*) was largely absent from the lessons, apart from Alfred and Demi defining a sample. This is in line with the informal nature of ISI, which stimulates the use of context language rather than formal statistical language. Second, Alfred and Celine used instructional materials (*uim,* i.e., pointing at the stack of books as the population, holding one book as a sample) in the majority of the fragments (64% and 83% respectively), and Demi only in a minority (23%) of the fragments. This difference may be explained by the fact that Alfred and Celine actually had piled up books in front of their classrooms, while Demi had left the books in the class' bookcase. Third, in most fragments the PSTs did not deviate from the provided lesson plan and the suggested representations, either on own their initiative or in response to children's input (*atb, da, cur, ds*), apart from Demi. Fourth, in 98% of all fragments, the PSTs responded to students' ideas, underlining the constructivist nature of the lesson. In the majority of the fragments, these responses were appropriate, although to a lesser extent for Alfred's lesson. Finally, inappropriate teaching actions often concern the failure to correctly interpret students' conceptual input (*ie, ac, rca*), to provide correct conceptual explanations of the content (*osk, dt, mcc*), and, for Alfred and Demi, a lack of purpose (*ap*). Although explanations were correct in the majority of the fragments, in crucial fragments PSTs failed to provide appropriate explanations and to make connections between concepts. These fragments will be discussed below for each PST separately.

Table 4 relates the appropriateness of teaching actions to ISI. It shows for each PST the number of fragments in which a particular ISI component was discussed, and the appropriateness of the teaching actions, both present and absent, in these fragments. Generalization beyond the data was most often discussed (20 fragments); Sampling methods least (five fragments). The overall appropriateness was relatively high for teaching actions related to Data as evidence and Generalization beyond the data. In particular, the transcript showed the PSTs used the data as evidence for their conclusions and drew conclusions that pertained consistently to the population, rather than to the sample data only. Also, all three were generally clear about what comprised the population and the sample, and Alfred and Demi provided clear definitions of a sample. Demi's relatively low score on Data as evidence is deflated by inappropriate teaching actions that were related to other ISI components. Concerning Sample size, 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 a limited time. Sampling method received limited attention: Celine briefly discussed whether it mattered which books would be selected (see below); and Alfred and Demi dismissed the suggestion to check back covers as unrepresentative for the content of a book. Alfred did not address sampling variability at all, and Uncertainty in one fragment only, with a low level of appropriateness.

Table 4. Proportion of appropriate (absence of) teaching actions for fragments where an ISI component was discussed[1]

	Alfred		Celine		Demi	
ISI comp.	Fragments	Appr.	Fragments	Appr.	Fragments	Appr.
Data as evidence	3	.976	6	.905	3	.690
Generalization beyond the data	8	.893	5	.823	7	.878
Sampling variability	0	na	3	.786	5	.829
Sampling methods	3	.786	1	1	3	.952
Sample size	2	.893	4	.911	5	.871
Uncertainty	1	.429	2	1	3	.714

[1]Alfred 11 fragments, Celine 12 fragments; Demi 18 fragments.

4.1. ALFRED

Apart from the trends discussed above, two issues stand out in Alfred's lesson. The first issue relates to the question whether it is necessary to sample all books from the stack. He correctly emphasized the impossibility of sampling all books and the necessity of drawing a sample. He also correctly remarked that the large difference between the first (*the*) and second most frequently used words in the sample made it more likely that *the* would also be the most frequently used word in the population:

> Alfred: 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?

Despite the use of the data as evidence for making generalizations, Alfred did not address the idea that if the sample is sufficiently large, one can expect that most samples will have the same most frequently used word. He thus did not relate the possibility of making generalizations to sampling variability (*dt*, *mcc*).

A second issue is that whole-class discussions concerning Sample size and Uncertainty lacked purpose and did not help to attain of the learning objectives (*ap*, *ac*). This was evidenced by lengthy series of gathering responses without discussing the conceptual appropriateness of students' suggestions (*ie*, *rca*, *rsi*, *dt*). For example, when eliciting initial likely top five words in the introduction phase of the lesson, he did not discuss which words are more likely than others. Also, when discussing the sample size, he asked numerous students how many pages to sample, without discussing the appropriateness of their suggestions. After several rounds of gathering suggestions, he concluded that the suggested sampling strategies were not fair after all and suggested that each group could sample 20 lines. The lack of purpose and failure to discuss students' input came most problematically to the fore when he discussed the certainty of the inference (see Table 4; 42.9% appropriateness for Uncertainty). Again, Alfred gathered numerous responses without discussing whether these responses made sense, even when students made contradictory claims. Students expressed 50% certainty, because not all books had been read; 99% certainty, because other books would probably show the same result; and 100% certainty, because the students had conducted the research themselves—Alfred agreed with them all.

During the reflection interview, Alfred attributed the lengthiness of the discussions to his lack of ISI-CK. Indeed, in 27% of the fragments, Alfred made obvious mistakes (*osk*: np-i). He indicated that he was able to elicit useful ideas from students for mathematical topics he knew well, but for ISI his lack of CK hindered his ability to respond appropriately to students: “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 it from multiple angles, respond quickly to what students say.” Although Alfred demonstrated comparatively high ISI-CK during the pre-test and the teacher college sessions, he claimed that these problems were caused by his lack of ISI-CK.

This episode provides an example of the difficulty PSTs can have using their CK in teaching. Another explanation could be that the amount of time Alfred had put in preparing the lesson was too little to internalize the lesson’s flow and underlying ideas and allow him to guide the discussions. Alfred had almost completely copied the teacher educator’s lesson plan into his own. This finding is further supported by the fact that while discussing the certainty of the conclusions, he held the lesson plan in his hands and seemed to look for the next step.

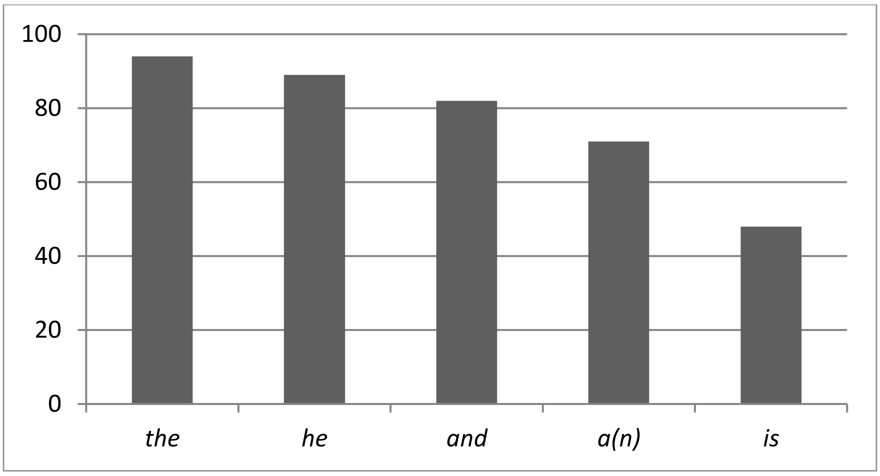

Figure 1. Bar graph of the pooled data for all groups in Celine’s lesson. A *and* an *refer to the same Dutch word (een)*

4.2. CELINE

During the introduction and research design phases, all Celine’s teaching actions were appropriate. She skillfully and efficiently steered the lesson towards the obtainment of the learning objectives. For example, when Celine discussed the interpretation of the pooled data, she used correct sampling variability arguments to discuss whether the conclusion also pertained to the population of books (*dt*, *mcc*). She agreed with a student who remarked that the small difference between the two most frequently used words increased the inference’s certainty (see Figure 1):

Celine: *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? […] Very good.

However, after this statement, Celine ran into problems for the first time. 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 this was necessary. 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 respond to the students’ unanimous opinion, and, after some hesitation, Celine closed the lesson abruptly. Like Alfred, she thus failed to use sampling variability arguments to convincingly demonstrate how it is possible to use a sample to make general claims about the population (*dt*, *mcc*).

The reflection interview revealed three reasons why she failed to respond to the class’s idea that all books needed 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?" […] 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 given was a practical one: there was no time left to explain how making inferences was possible. The second reason was that she was caught off guard and lost for words. Both Celine and the observer had the impression that before this incident the students understood that not all books needed to be read. During the reflection interview, Celine noted that she had not expected that students would respond almost unanimously that all books needed to be read. A third reason was that she did not believe that repeating the same explanation would have been helpful. Underlying this third reason could be another reason; Celine did not know how to explain how making inferences is possible if the difference between the two most frequently used words 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. Celine had used this argument when the teacher educator modelled the lesson in teacher college. However, it only works for relatively large differences. For small differences, another argument—one that relates sampling variability to uncertainty—is required. For example, one could demonstrate that for a sufficient sample size, most sample results yield the same most frequently used word, even for small differences between the two most frequently used words. The teacher educator demonstrated this line of reasoning during the simulation of the law of large numbers (see Table 1, session 1). However, Celine did not know how to use this reasoning in the context of the lesson she was teaching.

4.3. DEMI

Two issues stand out in Demi's lesson. The first issue was that Demi was the only PST to deviate from the lesson plan regularly, and most of her planned and unplanned deviations did not contribute to the attainment of the learning objectives (*atb*: 22%; da; 17%). These inappropriate deviations may be related to a lack of awareness of the purpose of the particular lesson fragment (*ap*: np-i 28%). Also, Demi addressed issues irrelevant for ISI; for example, while the discussion of words the class expected to find most frequently was only meant to facilitate the investigation by limiting it to five words, Demi put an unwarranted emphasis on the exact order within these hypothesized top five words. This may be based on an incorrect understanding of ISI.

The second issue was that in several fragments, Demi combined overt shortcomings in ISI-CK, shown in obvious mistakes (*osk*: np-i 28%) and incorrect responses to students' ideas (*rsi*: p-i 22%) with a tendency to neglect uncertainty (Uncertainty: 71.4% appropriate). Often, her language and tone were characterized by a decisiveness that was inappropriate given the uncertainty involved in making inferences. An example was her disapproval of a student's suggestion 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." Moreover, Demi did not notice the small difference between the two most frequently used words. Only after a student pointed out that this small difference made the inference more uncertain, Demi expressed herself less confidently, without discussing the student's reasoning. A second example of her tendency to express herself overconfidently and too decisively was observed at the close of her lesson, at which point she claimed that she knew that *the* was the most frequently used word in the population, the class bookcase, without clarifying how she knew this. This led to general confusion among the students.

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 [you] 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 1:	No.
	[…]
Student 2:	I don't understand.

Demi:	I tell you that if we would read the entire bookcase… *the* will still be in position 1.
Student 3:	No.
Student 4:	Yes
Demi:	It is.
	[…]
Student 5:	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.
Several students:	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 this would be the case in the population as well. Nonetheless, this episode showed how Demi used non-probabilistic language ("then we get the same result," "It is," "I tell you") and a decisive tone 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.

Demi attributed her negligence of uncertainty to her limited ISI-CK. During the reflection interview, she indicated that during the model lesson, she had missed an answer to the question, "What is it in the end?" (i.e., how making inferences is possible). In preparing the lesson, she had tried to get a firmer grasp of this issue in order to give her students a clear answer: "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. In the absence of a suitable explanation for how making inferences is possible, Demi appeared to resort to evading any discussion about the uncertainty of the inference. This might be related to a tendency to have precise answers, as the above quotation 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 could be made about the population. The lack of an explanation and a tendency to aim for certainty may have made her neglect the role of uncertainty, resulting in her claim that another sample "will" yield a similar outcome.

5. DISCUSSION

This study reports on the results of the analysis of the MKiT of three pre-service primary school teachers teaching ISI, who participated in a teacher college intervention and who had little exposure to ISI. The results show that for most aspects of ISI, the PSTs were able to mobilize their ISI-CK gained in the teacher college intervention when teaching ISI, as most of their teaching actions were appropriate. In most fragments, they acted appropriately in relation to the first two components of ISI, Data as evidence and Generalization beyond the data. Coding the appropriateness of both present and absent teaching actions revealed that inappropriate teaching actions often concerned the failure to correctly interpret students' conceptual input, to provide correct explanations, and, for Alfred and Demi, a lack of purpose. In relation to Sampling variability and Uncertainty, the PSTs struggled to provide correct and complete explanations.

A major finding was that the PSTs were consciously engaged in making inferences based on sample data. This engagement extended beyond simply following the lesson plan. Even though we have previously shown that, within a limited time frame, teacher college activities can be used to foster PSTs' ISI-CK (De Vetten et al., 2018), it is not evident that they were able to use their newly acquired knowledge in teaching (Lobato, 2006). In particular, the modelling of the ISI lesson in teacher college appeared valuable to the participants and helped them to prepare and teach their ISI lessons. The conscious engagement with ISI stands in contrast to the PSTs in Leavy (2010), who excessively focused on descriptive procedures. This may be due to providing the PSTs with a lesson plan that contained sufficient affordances for inferential reasoning, while in Leavy' study the PSTs designed ISI lessons themselves. In this study, the lesson plan appeared to have supported the PSTs to use instructional materials, to build on students' ideas, and to use informal language, appropriate to the students' level of understanding. Thus, we created the conditions for the conscious engagement in inferential reasoning. For teacher education, this finding may implicate that if PSTs have limited experience with ISI and no experience in designing such lessons, it may be more effective to provide them with a lesson plan rather than have them design lessons themselves (Chick & Pierce, 2012; Sullivan et al., 2009).

Coding the appropriateness of absent teachings actions revealed that the main shortcomings in the PSTs' MKiT were all related to insufficient ISI-CK. Insufficient ISI-CK was manifested in problems with explaining issues of sampling variability and uncertainty, with interpreting students' conceptual input, and in a lack of purpose. This shows, first, that CK is foundational for teaching, as hypothesized by Rowland et al. (2009), and in line with previous results in mathematics education research (e.g., Blömeke et al., 2011; Getenet & Callingham, 2021; Schmidt et al., 2008). Second, it shows that the PSTs found it difficult to "unpack" their ISI-CK related to sampling variability and uncertainty (Ma, 1999). In particular when unexpected remarks by students confronted the PSTs with issues not discussed at teacher college, they struggled to provide conceptually appropriate explanations. To facilitate the transfer from what PSTs learn in teacher college to primary classrooms and enable them to deal with unexpected events, teacher education may need to pay more attention to recontextualizing (Van Oers, 1998) ISI-CK, and helping PSTs to adapt the meaning of a concept learned in teacher college to a teaching setting. Furthermore, PSTs could be given more opportunities to deepen their ISI-CK (Haskell, 2001), in particular regarding their understanding of Sampling variability and the logic of inference (Lobato, 2006). Alternatively, when only limited time is available, lesson plans could include a ready-to-use, context-specific explanations, which allow PSTs to demonstrate sampling variability and how making inferences is possible.

From a data-analytic point of view, this study illustrates that using the KQ framework to also code the appropriateness of *absent* teaching actions, light can be shed on aspects of teacher knowledge that otherwise may remain hidden (Weston, 2013). Coding the appropriateness of both present and absent teaching actions thus helps to gain a more complete picture of teachers' MKiT.

5.1. LIMITATIONS

This study has a number of limitations. First, the PSTs' experience with ISI was limited. Teaching a subject requires a thorough MKT and solid teaching experience (Hill et al., 2008). As in Leavy (2010), PSTs' unfamiliarity with the subject hindered their teaching. Although the aim of our investigation was to study the expression of ISI-CK in a context where the PSTs had limited time to learn and practice teaching ISI, future research could investigate whether similar challenges are faced by teachers with more ISI experience and with multiple opportunities to teach ISI (Groth, 2017; Leavy, 2010).

Second, the PSTs' MKiT was dependent on participant selection, their pre-existing ISI-CK, the lesson taught, and the specific support offered to them during the intervention. These factors limit the generalizability of the findings, which necessitates research in other contexts. However, our finding of the necessity of sufficient MKT to show appropriate MKiT is consistent with studies in other domains of mathematics (Blömeke et al., 2011; Getenet & Callingham, 2021; Schmidt et al., 2008). Moreover, our finding of the importance that PSTs are able to provide convincing explanations and demonstrations related to sampling variability may inform teacher educators in general about the relative emphasis to put on sampling variability when supporting teachers to introduce primary school students to ISI.

Third, the PSTs' behavior may have been influenced by the observer's dual role as their teacher educator. This could have influenced their ISI-MKiT positively, if the PSTs put in greater effort than they would have otherwise, or negatively, if it resulted in more stress. However, no indications were found—for example, from the reflection interviews—that either of these influences played a vital role.

5.2. CONCLUSION

The aim of this study was to demonstrate what happens when PSTs with limited ISI experience teach an ISI lesson. The result that PSTs were consciously engaged in making inferences based on sample data is promising in light of the trend to have primary school students be introduced to ISI and the subsequent need to prepare primary school teachers to conduct this introduction. To conduct this introduction well, it is crucially important that teachers understand how making inferences is possible. Therefore, teacher education should include a variety of activities that focus on the core logic of inference, using tools to explain how making inferences is possible. These activities should resemble those that teachers can use to introduce primary school students to ISI. Providing a lesson plan that requires little descriptive analyses and that contains sufficient affordances to discuss ISI may help teachers to focus on inferential reasoning. Future research conducted in primary classrooms could determine whether these tools indeed allow teachers to provide such explanations. The ultimate interest would be in research that investigates whether these explanations help primary school students to understand that making inferences based on sample data is possible and how to find the balance between knowing nothing and knowing everything (Rubin et al., 1990).

REFERENCES

Auletto, A., & Stein, K. C. (2020). Observable mathematical teaching expertise among upper elementary teachers: Connections to student experiences and professional learning. *Journal of Mathematics Teacher Education, 23*(5), 433–461. https://doi.org/10.1007/s10857-019-09433-4

Ball, D. L., Lubienski, S. T., & Mewborn, D. S. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook of research on teaching (4th edition)* (pp. 433–456). MacMillan.

Ball, D. L., Phelps, G., & Thames, M. H. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*(5), 389–407. https://doi.org/10.1177%2F0022487108324554

Ben-Zvi, D., Bakker, A., & Makar, K. (2015). Learning to reason from samples. *Educational Studies in Mathematics, 88*(3), 291–303. https://doi.org/10.1007/s10649-015-9593-3

Blömeke, S., Kaiser, G., & Suhl, U. (2011). Teacher education effectiveness: Quality and equity of future primary teachers' mathematics and mathematics pedagogical content knowledge. *Journal of Teacher Education, 62*(2), 154–171. https://doi.org/10.1177%2F0022487110386798

Canada, D. (2006). Elementary pre-service teachers' conceptions of variation in a probability context. *Statistics Education Research Journal, 5*(1), 36–64. https://doi.org/10.52041/serj.v5i1.508

Charalambous, C. Y., Chin, M. J., Hill, H. C., & McGinn, D. (2020). Mathematical content knowledge and knowledge for teaching: Exploring their distinguishability and contribution to student learning. *Journal of Mathematics Teacher Education, 23*(6), 579–613. https://doi.org/10.1007/s10857-019-09443-2

Chick, H. L., & Pierce, R. (2012). Teaching for statistical literacy: Utilising affordances in real-world data. *International Journal of Science and Mathematics Education, 10*(2), 339–362. https://doi.org/10.1007/s10763-011-9303-2

curriculum.nu. (2019). *Voorstel voor de basis van de herziening van de kerndoelen en eindtermen van de leraren en schoolleiders uit het ontwikkelteam Rekenen & Wiskunde*. *[Proposal for the basis of the redesign of the core objectives of the end goals for teachers and school-leaders from the Arithmetic and Mathematics development team]*. Amersfoort, The Netherlands: SLO.

De Vetten, A., Schoonenboom, J., Keijzer, R., & Van Oers, B. (2018). The development of informal statistical inference content knowledge of pre-service primary school teachers during a teacher college intervention. *Educational Studies in Mathematics, 99*(2), 217–234. https://doi.org/10.1007/s10649-018-9823-6

De Vetten, A., Schoonenboom, J., Keijzer, R., & Van Oers, B. (2019a). Pre-service primary school teachers' knowledge of informal statistical inference. *Journal of Mathematics Teacher Education, 22*, 639–661. https://doi.org/10.1007/s10857-018-9403-9

De Vetten, A., Schoonenboom, J., Keijzer, R., & Van Oers, B. (2019b). Pre-service teachers and informal statistical inference: Exploring their reasoning during a growing samples activity. In G. Burrill & D. Ben-Zvi (Eds.), *Topics and trends in current statistics education research: International perspectives* (pp. 199–224). Springer.

Fennema, E., & Franke, L. M. (1992). Teachers' knowledge and its impact. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 147–164). Macmillan.

Getenet, S., & Callingham, R. (2021). Teaching interrelated concepts of fraction for understanding and teacher's pedagogical content knowledge. *Mathematics Education Research Journal, 33*(2), 201–221. https://doi.org/10.1007/s13394-019-00275-0

Gravemeijer, K., Stephan, M., Julie, C., Lin, F.-L., & Ohtani, M. (2017). What mathematics education may prepare students for the society of the future? *International Journal of Science and Mathematics Education, 15*(1), 105–123. https://doi.org/10.1007/s10763-017-9814-6

Groth, R. E. (2017). Developing statistical knowledge for teaching during design-based research. *Statistics Education Research Journal*, *16*(2), 376–396. https://doi.org/10.52041/serj.v16i2.197

Groth, R. E., & Meletiou-Mavrotheris, M. (2018). Research on statistics teachers' cognitive and affective characteristics. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 327–355). Springer.

Haskell, R. E. (2001). *Transfer of learning. Cognition, instruction and reasoning*. Academic Press.

Hill, H. C., Ball, D. L., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., & Sleep, L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction, 26*(4), 430–511. https://doi.org/10.1080/07370000802177235

Learning Mathematics for Teaching. (2006). *A coding rubric for measuring the quality of mathematics in instruction (Technical Report LMTI 1.06)*. University of Michigan, School of Education.

Leavy, A. M. (2010). The challenge of preparing preservice teachers to teach informal inferential reasoning. *Statistics Education Research Journal, 9*(1), 46–67. https://doi.org/10.52041/serj.v9i1.387

Leavy, A. M., & Sloane, F. (2017). Insights into the approaches of young children when making informal inferences about data. In T. Dooley & G. Gueudet (Eds.), *Proceedings of the tenth Congress of the European Society for Research in Mathematics Education* (pp. 788–795). DCU Institute of Education and ERME.

Liu, Y., & Grusky, D. B. (2013). The payoff to skill in the third industrial revolution. *American Journal of Sociology, 118*(5), 1330–1374. https://doi.org/10.1086/669498

Lobato, J. (2006). Alternative perspectives on the transfer of learning: History, issues, and challenges for future research. *The Journal of the Learning Sciences, 15*(4), 431–449. https://doi.org/10.1207/s15327809jls1504_1

Ma, L. (1999). *Knowing and teaching elementary mathematics. Teachers' understanding of fundamental mathematics in China and the United States*. Erlbaum.

Makar, K. (2016). Developing young children's emergent inferential practices in statistics. *Mathematical Thinking and Learning, 18*(1), 1–24. https://doi.org/10.1080/10986065.2016.1107820

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, *8*(1), 82–105. https://doi.org/10.52041/serj.v8i1.457

Meletiou-Mavrotheris, M., Kleanthous, I., & Paparistodemou, E. (2014). Developing pre-service teachers' technological pedagogical content knowledge (TPACK) of sampling. In K. Makar, B. De Sousa & R. Gould (Eds.), *Sustainability in statistics education*. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9), Flagstaff, AZ. https://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_3A1_MELETIOUMAVROTHERIS.pdf?1405041596

Mickelson, W. T., & Heaton, R. M. (2004). Primary teachers' statistical reasoning about data. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 327–352). Kluwer.

Mooney, E., Duni, D., , VanMeenen, E. & Langrall, C. (2014). Preservice teachers' awareness of variability. In K. Makar, B. De Sousa & R. Gould (Eds.), *Sustainability in statistics education*. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9), Flagstaff, AZ. https://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_C149_MOONEY.pdf?1405041831

Putnam, R. T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher, 29*(1), 4–15. https://doi.org/10.3102%2F0013189X029001004

Rowland, T., Huckstep, P., & Thwaites, A. (2005). Elementary teachers' mathematics subject knowledge: The knowledge quartet and the case of Naomi. *Journal of Mathematics Teacher Education, 8*(3), 255–281. https://doi.org/10.1007/s10857-005-0853-5

Rowland, T., Hurst, C., Thwaites, A., & Turner, F. (2009). *Developing primary mathematics teaching: Reflecting on practice with the knowledge quartet*. SAGE Publications.

Rubin, A., Bruce, B., & Tenney, Y. (1990). Learning about sampling: Trouble at the core of statistics. *Proceedings of the Third International Conference on Teaching Statistics* (ICOTS3), Dunedin, New Zealand (pp. 314–319). https://iase-web.org/documents/papers/icots3/BOOK1/A9-4.pdf

Saldanha, L. A., & Thompson, P. W. (2007). Exploring connections between sampling distributions and statistical inference: An analysis of students' engagement and thinking in the context of instruction involving repeated sampling. *International Electronic Journal of Mathematics Education*, *2*(3), 270–297.

Schlesinger, L., & Jentsch, A. (2016). Theoretical and methodological challenges in measuring instructional quality in mathematics education using classroom observations. *ZDM Mathematics Education*, *48*(1–2), 29–40. https://doi.org/10.1007/s11858-016-0765-0

Schmidt, W. H., Houang, R. T., Cogan, L., Blömeke, S., Tatto, M. T. Hsieh, F. J., Santillan, M., Bankov, K., Han, S. I., Cedillo, T., Schwille, J., & Paine, L. (2008). Opportunity to learn in the preparation of mathematics teachers: Its structure and how it varies across six countries. *ZDM Mathematics Education*, *40*(5), 735–747. https://doi.org/10.1007/s11858-008-0115-y

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*(1), 1–23. https://doi.org/10.17763/haer.57.1.j463w79r56455411

Sullivan, P., Clarke, B., & Clarke, D. (2009). Converting mathematics tasks to learning opportunities: An important aspect of knowledge for mathematics teaching. *Mathematics Education Research Journal*, *21*(1), 85–105. https://doi.org/10.1007/BF03217539

Van Oers, B. (1998). From context to contextualizing. *Learning and Instruction*, *8*(6), 473–488. https://doi.org/10.1016/S0959-4752(98)00031-0

Watson, J. M. (2001). Profiling teachers' competence and confidence to teach particular mathematics topics: The case of chance and data. *Journal of Mathematics Teacher Education*, *4*(4), 305–337. https://doi.org/10.1023/A:1013383110860

Wells, G. (1999). *Dialogic inquiry: Towards a socio-cultural practice and theory of education*. Cambridge University Press.

Weston, T. L. (2013). Using the knowledge quartet to quantify mathematical knowledge in teaching: The development of a protocol for initial teacher education. *Research in Mathematics Education*, *15*(3), 286–302. https://doi.org/10.1080/14794802.2013.849865

ARJEN DE VETTEN (corresponding author)
ICLON, Leiden University
P.O. Box 905
2300 AX Leiden
The Netherlands

APPENDIX 1

DEFINITION OF KNOWLEDGE QUARTET DIMENSIONS AND CODES

KQ dimensions and codes	Definition
Foundation	**PST’s knowledge, beliefs and understanding acquired in the academy, in preparation for their role in the classroom**
Adherence to textbook	PST deviates from provided lesson plan, in planning or delivery of lesson
Awareness of purpose	PST is aware of the purpose of the lesson
Concentration on procedures	PST concentrates on applying or teaching procedures without an attempt to discuss the reasoning behind the procedure
Identifying pupils’ errors	PST identifies students’ mistake(s)
Overt display of subject knowledge	PST displays subject knowledge that clearly extends beyond or falls below (i.e., makes obvious errors) what can be expected from a teacher with the same background
Theoretical underpinning of pedagogy	PST shows deliberate use of theoretical insights regarding the teaching of the subject in planning or delivery of lesson
Use of mathematical terminology	PST uses mathematical terminology at a for the students appropriate level
Transformation	**PST’s capacity to transform content knowledge into pedagogically powerful forms**
Choice of examples	PST selects and uses examples, other than suggested in the lesson plan
Choice of representations	PST selects and uses mathematical representations that fit the content to be explained
Use of instructional materials	PST uses instructional materials, either concrete or symbolic, to explain the content
Teacher demonstration	PST explains or demonstrates content
Connection	**PST’s ability to bind together choices and decisions that are made for more or less discrete parts of mathematical content**
Anticipation of complexity	PST anticipates the complexity of the content by assessing whether students will be able to understand the content and by assessing possible misconceptions
Decisions about sequencing	PST introduces content, ideas and strategies in an appropriately progressive order (i.e., in an order that makes the content understandable to students)
Making connections between concepts	PST explicitly discusses (with the students) the connections between various mathematical concepts
Making connections between procedures	PST explicitly discusses (with the students) the connections between various mathematical procedures
Recognition of conceptual appropriateness	PST recognizes the conceptual appropriateness of students’ remarks concerning mathematical concepts
Contingency	**PST’s ability to act upon unplanned classroom events (i.e., to ‘think on one’s feet’)**
Deviation from agenda	PST deviates from the lesson plan, as a result of what happens in class
Responding to students’ ideas	PST responds to students’ ideas or suggestions related to the content of the lesson
Responding to (un)availability of tools and resources	PST responds to unplanned and/or unexpected (un)availability of tools and resources
Teacher insight	PST reflects on his or her teaching during teaching (‘reflection in action’)

APPENDIX 2

SUMMARY OF THE LESSON PLAN, "WHAT IS THE MOST FREQUENTLY USED WORD IN A STACK OF CHILDREN'S NOVELS?"

Introduction phase

1. Introduce the topic, the most frequently used words in Dutch children's novels, by discussing whether all words need to be known to understand a book.
2. Let students formulate in small groups the hypothetical top three likely candidates for the most frequently used word in Dutch children's novels. Use a class discussion to reach consensus about the top five, and steer toward a sensible top five.
3. Tell the students they will conduct an investigation. Show the driving question: "What is the most frequently used word in Dutch children's novels in the school library?"

Research design and data collection phase

1. Ask the class: "How can we find an answer to the driving question?" Try to elicit the response: by taking a sample. Steer class discussion toward a good sampling method, such as sampling from the pile a number of (random) books and pages or lines. Agree on a sampling method.
2. Provide the students with the form and instruct them how to collect the data: "Read the agreed-upon number of pages/lines and check a box each time one of the top five occurs. Count totals." Let students conduct investigation in groups of two or three. Students answer the questions on the handout.
3. Let each group compare their results with those of another group and let them formulate one answer. Emphasize that only one word can be the most frequently used. If possible, let groups with different answers compare their results.

Conclusion phase (original version)

1. Hang forms on the wall. Ask for the students' answers and organize cognitive conflict by discussing diverging results. Steer toward the solution that pools the groups' sample results.
2. Use the template Excel file to make a bar graph of the pooled data for all groups.
3. Show the graph based on pooled results to the class. Discuss what the final answer to the driving question is. Discuss the certainty of the conclusion. Eventually discuss whether another sample would yield a different result. Discuss how to increase the certainty of the conclusion. Evaluate the lesson by asking what the students have learned.

Conclusion phase (adapted version)

1. Use the template Excel file to make a bar graph of the pooled data for all groups.
2. Hang sheets of paper with the top five words on the wall, including a sheet stating: "I cannot say anything about which word is most frequently used in the pile of books." Have students stand at one of the sheets to indicate their answers to the driving question.
3. One-by-one, show four fictitious statements to students. Have students indicate their approval, and let them explain why:
   1) "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."
   2) "As long as we check enough books and pages, we don't have to read all books."
   3) "We should select different kinds of books from the pile because the books we select should resemble all the books in the pile."
   4) "I am quite sure that the most frequently used word in the graph is also the most frequently used word in the entire pile of books." (If the two most frequently used words are close, use: "I am not entirely certain which word is the most frequently used in the entire pile of books, but I am still quite certain that number 1 or 2 in the graph is also the most frequently used word in the entire pile of books."
4. Show an overview of the correct statements, which form a complete reasoning underlying the conclusion. Have students again stand at one of the sheets to indicate what their answer to the driving question is. Discuss any changes. Evaluate the lesson.