HOW CHILDREN WITH DIFFERENT DIALECTS NAVIGATED UNCERTAIN LANGUAGE IN A STATISTICS INVESTIGATION

MITALI THATTE
Quality Education Support Trust
mitali.thatte@quest.org.in

KATIE MAKAR
The University of Queensland
k.makar@uq.edu.au

NILESH NIMKAR
Quality Education Support Trust
nilesh.nimkar@quest.org.in

ABSTRACT

Language, culture, and conceptions of uncertainty can impact the way that students respond to a statistical investigation. The aim of this small exploratory study was to gain insight into how two groups of children (aged 9-12) speaking different home dialects adopted expressions of uncertainty in a standard dialect used in school, specifically the use of the Marathi word for "about." One group of children went to a government school where the home dialect differed from the standard dialect, while a second group attended a relatively elite private school where children spoke the standard dialect at home. Our findings suggest that children from both groups did not spontaneously use the word "about" while describing data even when nudged by the researcher to do so, though the children from the second group were more quickly able to adopt the word and use it in the way the researchers expected. The findings have the potential of exploring and impacting the influences of language on the learning of statistics in a non-Western culture.

Keywords: *Statistics Education Research, Uncertain Language, Statistical Language*

1. INTRODUCTION

Being able to use uncertain language has always been an important aspect of informal statistical inference (Makar & Rubin, 2009). This includes both the ability to acknowledge the variation in the data and to express uncertainty regarding future events. However, language is a social phenomenon and Vygotsky's (2012) theory states that social interactions play an important role in the development of language. This is especially true for social interactions including an adult. It stands to reason then that this may also be true for development of uncertain language. Children likely negotiate the usage of syntax and semantics of uncertain language through various interactions in contexts where they have to reason about uncertainty.

This paper explores the interactions of a researcher with children from two schools in Maharashtra, India while the children were engaged in a statistical investigation. The medium of instruction for both schools was Marathi, which is the same language that the children spoke at home. However, in the first school the children's home dialect was slightly different than the standard dialect used in school textbooks. The children from the first school also belonged to a socioeconomically disadvantaged portion of society while the second school was a regional medium private school considered to be high quality by residents of the city. Studies in India have shown that children from socioeconomically disadvantaged families face challenges in dealing with the formal language expected in school textbooks (e.g., Kulkarni, 1981). In this study, we explore the challenges faced by the researcher and the children in arriving at a shared understanding of the uncertain language used by the researcher.

2. LITERATURE REVIEW

In this section, we look at the conceptual difficulties in statistics that may influence how students articulate their understanding of likelihood including the challenges of language in statistics for non-English learners. We consider how language goes beyond what and how ideas are communicated and discuss how language also encodes culture, power, and equitable access to learning. Finally, we broadly discuss language as a technology.

2.1. ARTICULATING UNCERTAINTY

What does articulation of uncertainty entail? Expressions of uncertainty are often vague and subjective in nature. They can carry challenges conceptually in aligning the intended meaning with the context in which probabilistic phrases are interpreted and used (Bonnefon & Villejoubert, 2006; Karelitz & Budescu, 2004). They also can present difficulties with navigating the ways to express likelihood and the degree of uncertainty due to variability in estimations of chance and future predictions. "The situation in which random variation is met influences how people think about probability" (Pratt & Kazak, 2018, p. 214). Students have been observed to articulate predictions under uncertainty by either being overly deterministic or overly relativistic (Ben-Zvi et al., 2012; Rubin et al., 1990). When asked to make predictions from data, the primary students in Ben-Zvi et al.'s study (2012) oscillated between certainty-only (deterministic) and uncertainty-only (relativistic) statements before further exploration helped them express uncertainty with probabilistic language (about, maybe). Therefore, attending to expressions of uncertainty require attention to conceptual understanding of uncertainty, how it is expressed, and, more importantly, its relationship with the context.

Articulating statistical ideas can also stem from students struggling to coordinate ideas of uncertainty, variability, distribution and prediction in statistics. Variability is tied to how one thinks about distribution and prediction.

> Until a data set can be thought of as a unit, not simply as a series of values, it cannot be described and summarized as something that is more than the sum of its parts. An average is a measure of the center of the data, a value that represents aspects of the data set as a whole. An average makes no sense until data sets make sense as real entities. (Mokros & Russell, 1995, p. 35)

Average is the "signal" among the noise in a distribution that allows one to estimate an unknown in an inference or an expected value (Konold & Pollatsek, 2002). When children talk about a statistical concept, they may not yet have the experience to consider it holistically. Before students develop a sense of distribution as an aggregate, they may see data as separate values and not consider the average as representative of the whole (Konold et al., 2015; Mokros & Russell, 1995; Watson, 2006). This phenomenon may also reflect Konold and his colleagues' (2015) identification of a data as a pointer or case, where children who collect data may focus on an attribute from their experience of collecting the data but may not be present in the data (data as a pointer) or focus on an individual value in a distribution such as a maximum or minimum (data as a case). In both cases, children focused on either an attribute of a data set or an attribute of the phenomenon from which the data were collected. Therefore when asked to describe a distribution, children may focus on an individual point in a distribution, ignoring other values, jumping between values, or responding in ways that were inconsistent with the rest of the distribution (Konold et al., 2015; Torok & Watson, 2000).

Articulation of uncertainty for non-English speakers. The articulation of uncertainty by non-English speaking learners can be further complicated by their lack of experience with English. Those learners may experience challenges related to statistical concepts, to the distinctive language features in statistics, or both or neither. It can be difficult to tease out exactly where the challenges lie. For example, one of the challenges facing the exploration of statistical ideas using uncertain language in Marathi (the language used in this study) is that the use of some words in English do not translate well into Marathi. For example, there is no commonly used translation of the word "about" (used in sentences like, "There are about 1000 leaves on this branch."). It is with this context of linguistic diversity that the researchers conducted the statistical investigation with children.

In research with non-English speaking students, additional constraints arise as a function of how one expresses statistical ideas related to particularities of the language or dialect. Lesser and Winsor

(2009) worked with Spanish-speaking pre-service teachers in an English-speaking learning environment. They argued that in statistics education with second-language learners, not enough emphasis is placed on statistical *register*, a subset of language relevant to a particular context and purpose. Focusing on register is unlike lexicon (learning a list of words) as learning is more than simply substituting words, but requires the capacity to differentiate between situations.

> The academic meaning of a term may be the same as the everyday meaning, different from the everyday meaning, or not have an everyday counterpart at all. … The most difficult of these three possibilities is arguably when the everyday and statistical meanings differ. (p. 8)

Learners' proficiency in everyday language may be quite different from the more complex, decontextualised academic situations, lagging behind by three or four years. Lesser and Winsor's (2009) research found that even if learners are exposed to an academic term, they may be more likely to draw on their everyday register than to adopt the academic term. Their research recommended that in statistics, introducing metaphors for academic terms or contextualising the use of the word is likely more useful than introducing single words, particularly if the statistical content expressed by the term is not available in a student's language. They recommended attending to how a student's language can express an idea and to subsequently incorporate their language, community, and culture into learning activities rather than see these as limitations. "This is an important place to warn of the pitfall of focusing unduly on disadvantages, as this can lead to 'deficit models'" (p. 22). Other recommendations Lesser and Winsor discuss include the use of wait time, embedding instruction across a number of contexts, seeking multiple ways to express and represent a new idea, and encouraging collaboration. In addition, they recommended recognising informal ways that learners express ideas can be valuable in supporting them to adopt statistical concepts.

In a related study, Lesser and his colleagues (2013) focused on the different dimensions of statistical register (*field*, *mode*, *tenor*) to develop a survey that identified differences in how students that speak different languages understood statistical concepts that were articulated differently in each language. The field dimension of register referred to how statistical language use was affected by the topic in which statistical ideas were used. They also identified the mode dimension, which referred to how statistical ideas varied between spoken and written forms, as well as the tenor dimension, which captured the social relationship in which language was used. In the tenor dimension, for example, they found that in a learning environment, students from a different language group may feign understanding to avoid drawing attention to themselves. These multiple dimensions of register further illustrate the complexity of research on how learners articulate statistical ideas when working in a language other than English. Unfortunately, all of these studies took place in English-language contexts, even when the students were non-English speakers. This creates a major obstacle when seeking to locate and build on research in non-English speaking contexts.

2.2. BEYOND LANGUAGE AS ARTICULATION: CULTURE, POWER, AND EQUITABLE ACCESS

As Lesser and his colleagues suggested, cultural dimensions of language can contribute significantly to learning statistics. For example, Chauhan's (2013) research with children in India showed how cultural beliefs about destiny influenced their perceptions and articulations of chance, often describing randomness and luck as pre-determined. Her research showed that children with these beliefs may not see the purpose of estimating likelihood as "ultimately everything depends on luck" (p. 153).

Language evokes tacit assumptions about the speaker, particularly in students' command of the dominant language (Benson, 2013). In education, if a learner is not fluent in the dominant language of instruction and use of language in school, the "monolingual habitus causes us to view a learner in deficit" (p. 284). The difference in status between the language of society and the language of school can exacerbate this issue, particularly if content is taught and assessed in the dominant language. In content areas, school language privileges technical terms without considering whether and how the language of society articulates ideas differently. Schools may not even be aware of their monolingual

habitus. "Adopting a multilingual habitus in educational policy and practice means giving value to existing linguistic and cultural resources, not ignoring them" (p. 295).

Kulkarni (1981, p. 55) wrote of the barriers that children—especially those from lower socioeconomic communities—face while learning scientific concepts, "More relevant is the role of language in classroom instruction at the school level and the comparison of language in and out of school." He makes this point not when the medium of instruction was an entirely different language, but when the kind of language (formality, syntax, etc.) used in school was different than the one used at home. He noted that changing the language of science textbooks to better suit children's home language helped children's conceptual understanding. As expected, students' ability to read simplified texts improved significantly. In addition, students' interactions with teachers improved when teachers' adopted simple language. However, the author's most significant finding was the removal of disparity in the performance of students coming from different socioeconomic backgrounds. The gap between the achievement levels of children from different socioeconomic backgrounds was reduced by simplifying the language in the textbooks.

The Indian state of Maharashtra attempted help children learn number names by simplifying the irregularity of the number names in the language of Marathi (गणित, 2019). In English, number names follow a definite pattern after 20; once a child memorizes the number names up to 20 as well as the names for 30, 40, 50, 60, 70, 80, 90, and 100 (28 terms overall), she does not have to put in significant effort to memorise all the remaining number names up to 100. She only had to recognise and use the pattern. In contrast, the way the number names are structured in Marathi and other north Indian languages makes it difficult to find concrete patterns. As a result, children essentially have to memorise all number names up to 100. Balbharati, the state textbook authority, recommended that teachers should start teaching acceptable alternative number names based on the English number-naming pattern. They explicitly simplified the language to reduce the irregularity of number names, and thus the number of terms children needed to memorise, thereby helping children with numbering, especially those from disadvantaged communities (गणित, 2019). Encouraging children to use uncertain language in statistics classrooms, especially in socioeconomically and linguistically diverse communities is a challenging process. This problem has been well recognized in other contexts. Teachers are faced with the dual challenge of encouraging the natural expression of children along with introducing the children to the language used by dominant groups, who hold more power. For example, Subtirelu (2019) outlined an ongoing movement in the field of education to destigmatise diverse expressions and encourage native speakers to speak in their own language. In contrast, Delpit (1988) argued that we should accept children's home expression and then also teach them standard (academic) language in order not be to disadvantaged. This issue is especially relevant as learning is thought to be a largely social process where the kind of interactions that you have with a significant adult has a large impact on your learning (Vygotsky, 2012). It stands to reason that children's ideas about statistics concepts are influenced by the language used by their teachers in the classroom and therefore is likely highly relevant to children that speak a different dialect than the standard dialect spoken by the teachers or used in textbooks. It is this key issue that researchers eventually hope to address when it comes to natural expressions of uncertainty. Various strategies can be used to facilitate children's expression in a dominant language. Moschkovich (1999) described revoicing as a possible technique for scaffolding children's expression. Such techniques may well be useful in contexts where the dialect is slightly different than the children's home dialect.

2.3. LANGUAGE AS TECHNOLOGY

This special issue examines the relationships between technologies and equity issues. Our study could be construed as only addressing equity. We, however, do not see it this way and argue that it is also relevant to technologies. John Dewey (1938) considered language a technology; "Words mean what they mean in connection with conjoint activity" (p. 53) to generate some effect, just as tools do. Technologies and, by extension languages, therefore both seek to purposely shape outcomes. Distinguishing between the machine and the "web of human activities surrounding the machine" as part of life and culture (Pacey, 1994, p. 3), gives rise to different perceptions of technology.

Colloquially, "technology" is often synonymous with digital technologies (computers, mobile devices, software). However, the academic view of technology is broader, additionally encompassing tools and artefacts, and recognises how technologies are situated in and rise out of culture (Pacey, 1994). Even more broadly are the "distinctions between technology-as-knowledge, technology-as-process, and technology-as-product – or thoughts, activities, and objects" (Hickman, 1990, p. 14), with the latter referring to the physical tools and digital devices that align with the historical perspective. Technological activities are the processes of design, invention, and operation that include planning, teaching and engineering. Knowledge as a technology is related to the Greek *techne*, referring to a productive skill that grows out of practice (Heidegger, 1998; Hickman, 1990). Technological objects rise out of a need to transform a problematic situation, with this transformation driven by thoughts rather than objects. Dewey identified "inquiry as a form of technological activity … [as] the mathematician nevertheless uses abstractions and symbols that are just as much tools as are hammers and microscopes" (Hickman, 1990, p. 46). Therefore language is likewise an abstraction with meaning situated in context, just as technology is a mode of language with meaning for its user.

> A tool or machine, for example, is not simply a simple or complex physical object having its own physical properties and effects, but is also a mode of language. For it says something, to those who understand it, about operations of use and their consequences. (ibid)

However, scientific or technical language is standardised to represent abstract meaning that does not depend on local, social use. As argued above, this can have a substantial impact on equity issues in student learning. There are few studies, however, that explore statistical ideas of uncertainty in languages or dialects that do not articulate uncertainty as is done in everyday English. Based on the literature and concerns of equity in neglecting non-dominant languages and dialects, the research question under investigation in this paper is: In the context where Marathi is the school language, how are children who speak different variants of Marathi able to express uncertain language in statistics?

3. METHODOLOGY

The activity presented in this paper was conducted at two schools, each with two groups of four children in grades 5 and 6. The schools granted permission for researchers to contact the children and their parents. The activity was conducted outside the school with the permission of the parents. Each activity took approximately 2.5 hours to complete.

The researcher told the children that there was to be a competition in the district where multiple schools would participate. Every school in the district had to send one representative to this competition. The competition would involve throwing a ball in a bucket multiple times from a distance and whoever got the ball in the bucket the most number of times would win the competition. The researcher told the four students that they had to recommend one student among them who would represent their school in the competition. The children spent the first part of the activity discussing what they would have to do in order to recommend someone to the competition. This activity lasted for about 20 minutes. With a nudge from the researcher, they decided that they would have to collect data several times and then make a decision. The children played 10 rounds each with 10 chances per child and collected data for the number of times out of 10 the ball landed in the bucket and stayed there. Based on the data the children had a discussion facilitated by the researcher on who gets the most balls in the bucket in order to decide who to send to the competition.

The data collection consisted of the activities, which were audio and video taped, artefacts that were created (e.g., tallies or graphs of the data) and field notes. The analysis was based on an adaption of Powell et al.'s (2003) analysis of video data with the following principles:

1. Viewing attentively the video data
2. Describing the video data
3. Identifying critical events
4. Transcribing
5. Coding
6. Constructing storyline

7. Composing narrative

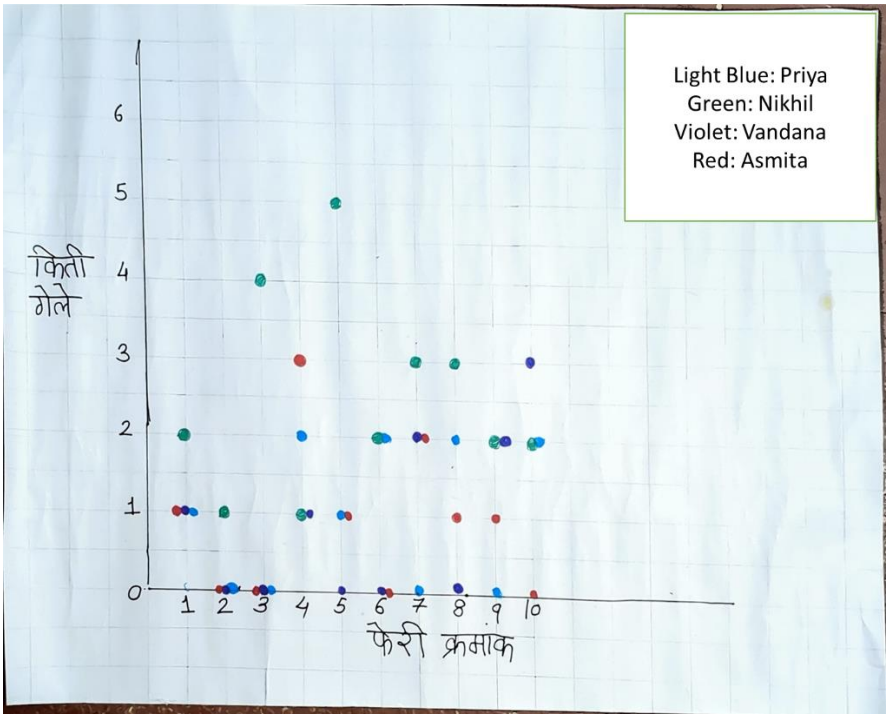
The focus of the analysis was to identify excerpts in the data in which the language of uncertainty arose or did not arise but was expected to arise. Initially, the audio and video data were reviewed and briefly summarised with timestamps for content. Potential critical events were highlighted during this process and then transcribed. A critical event was identified as an interaction in which a language feature raised an issue or insight into the children's articulation of uncertainty. They were generally of the type where children's expression of uncertainty was in line with previous literature or where expression is unexpectedly different than expected. The researchers also identified events as being critical when there was a significant difference in the expression of uncertainty for the two different groups of children. After introducing the word for 'about' in Marathi, the events where the children were either playing around with the word or using the word in way different than what the researcher intended generally qualified as critical events. It is important to note that the focus of this particular study was on exploring the expression of uncertainty and not about the entirety of the statistics activity. All conversations where the word for 'about' was described were also categorized as critical events. Since two of the researchers were familiar with the Marathi language, both of them were responsible for translating the transcripts. Typically a translation was done by one of the researchers and checked by the other researcher who was familiar with the language. The pieces of transcription were reviewed and coded to identify concepts that arose. These excerpts were discussed to locate a storyline—an articulation of key ideas told through a coherent linking of the transcripts. The transcripts were reviewed again to edit them for brevity of expression, adding further insights of analysis to create a narrative, before final editing of the results section presented below.

4. RESULTS

We present results of our investigation from School 1 in which students were first informally, then more explicitly, introduced to ways of articulating uncertainty using the word "about." The children in School 1 spoke a different dialect at home than what was used in school. The expression of uncertainty in this dialect was different than the standard dialect. We follow these excerpts with brief excerpts from School 2, where the home and school dialects were aligned. As previously mentioned, we therefore conducted an exploratory qualitative study where we investigated whether there was a difference in how the students in these two contexts expressed uncertainty in a statistical investigation involving estimation and prediction. We conclude with a discussion of the results.

4.1. SCHOOL 1

The following is an interaction between the researchers and students after they have drawn the time series graph for Nikhil (all names are pseudonyms). The researcher has asked the question, "About how many balls does Nikhil get in?" While asking the question, the researcher has used the Marathi rough equivalent for the word "about" which is "sadharan (साधारण)". In the researcher's experience, the word "sadharan" is commonly understood by adults who speak the standard dialect/variant of Marathi.



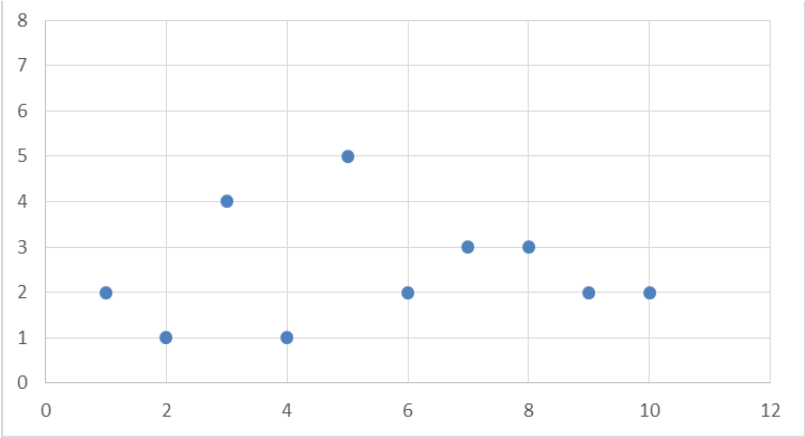


Figure 1. The results of ten throws each from four children (top). Nikhil's data is shown in green, which was the only visible data at the initial point of discussion. Nikhil's data has been reproduced in Excel (bottom) for ease of interpretation.

Researcher:	About (sadharan) how many balls does Nikhil throw? He has ten chances right? साधारण किती टाकतो. त्याला १० chances दिलेत ना?
Vandana:	25 पंचवीस
Researcher:	He has ten chances right? About (sadharan) how many balls does he get in? त्याला १० chances दिले ना? तो दहापैकी किती टाकले?
Vandana.	He got two in the first round, three in the second round, four in the third round, one in the fourth round, five in the fifth round…. पहिल्या फेरी २ मिळाले, दुसऱ्या फेरीत ३. तिसऱ्यात ४, चौथ्या फेरीत १, पाचव्या फेरीत ५...
Priya:	So he got 25 overall. सगळे मिळून २५ झाले.

Because the questions in textbooks generally ask students to count the frequency, or add up different values, it seems likely that children are trying to apply the same technique in this context. It was apparent that the children were not used to hearing the word "sadharan," likely because it was not a part of their daily use. This event was categorized as a critical event as the researcher did not expect the word to be interpreted the way that it was interpreted by the children. Considering that children did not adopt the word "sadharan," the researcher tried to use another word, "andaje" (अंदाजे), which might work as a substitute. The word "sadharan" is almost an exact translation of the word "about" for describing the data, as well as for expressing uncertainty about events (including future events). The word "andaje" is better suited to expressing uncertainty about events rather than for describing data. "Andaje" is also a broader word which is sometimes used as a combination of "guess" and "estimate." For example the predictive statement, "It will take me about (andage) two hours to reach my destination," but not the descriptive statement, "An average child scored about (andage) 30 marks on the test."

Researcher: My question is a little different. Out of ten, about (andaje) how many balls will Nikhil get in?
माझा प्रश्न थोडासा वेगळा तो दहापैकी अंदाजे किती बॉल टाकेल.

Nikhil: Four.
चार.

Priya: Five.
पाच.

Researcher: Why do you think five?
तुला का असं वाटतंय ५ टाकेल?

Priya: His aim is very good.
त्याचा नेम चांगला आहे.

Researcher: Why do you think he will throw *five*?
तुला का वाटतं की तो ५ टाकेल?

Vandana: Because he has gone most times and his aim is also good.
कारण की तो सगळ्यात जास्त गेलाय आणि त्याचा नेमही चांगला आहे.

The children have used their experience of seeing Nikhil throw the ball and have also been optimistic in their prediction of how many Nikhil will throw next time (as five was his maximum). This is also an important event as the children did not consider the word "andaje" to be equivalent to the word "about" either. Considering that the children used neither word as a translation of the word "about" in the way the researcher expected, the researcher decided to more explicitly introduce the word "sadharan" to the children and model how it is used in statistics. This was a turning point in the study. In the beginning, the researcher switched from trying to understand the natural expression of uncertainty and to see how the children responded to the word for "about" used by the researcher. The following conversation took place almost immediately after the previous conversation. The researcher then tried to introduce the word to the children and tried to understand how the children would respond to this introduction. The following interaction shows how the researcher conducted the discussion to create appropriate conditions for introducing the word.

Researcher: Let's have a bet on how many balls Nikhil will get in. What number will you bet on?
आपण पैज लावूया हं. निखिल किती बॉल टाकेल याची. तर कुठल्या नंबर वर पैज लावायची?

Vandana: Five.
पाच.

Priya: Three. (Immediately changed her answer from five to three)
तीन.

Researcher: Why do you think so?
तुला असं का वाटत आहे.

Priya: His aim is not that good.
कारण की जास्त नेम नाही लागलेला.

This time, rather than use their personal beliefs about how many Nikhil could throw, the children gravitated towards using the central tendency of the data when the researcher phrased the question about making a bet (about a future event). The researcher and the children had a discussion on which number to bet on if Nikhil gets ten chances to throw the ball in the bucket. The researcher then introduced the word "sadharan."

Researcher: Suppose I bet that Nikhil throws seven. Will I win or lose?
समजा अशी मी पैज लावली की निखिल सात टाकेल. तर मी जिंकीनका हरीन?

Chorus: Lose.
हरेन

Researcher: Suppose I bet that Nikhil scores a two. Then will I win or lose?
समजा मी अशी पैज लावली की निखिल दोन टाकेल. तर मी जिंकीन का हरिण?

Chorus: Win.
जिंकेन

Researcher: Can I say Nikhil throws about (sadharan) six balls in?
तर निखिल साधारण दहा बॉल टाकतो असा पण म्हणू शकतो का?

Chorus: No.
नाही

Researcher: Why?
का?

Asmita: Because he has not thrown six even once.
कारण त्यांनी एक पण सहा टाकलेला नाही आहे

Researcher: Can I say he throws about (sadharan) five?
बरं मग साधारण पाच टाकेल असं म्हणू शकतो?

Chorus: Yes
हो

Researcher: About (sadharan) how many balls does Nikhil throw?
हा तर निखिल साधारण किती टाकतो?

Priya: Two or three.
दोन नाहीतर तीन

Researcher: Yes, we can say that Nikhil throws about (sadharan) two or three balls. Meaning, he throws two or three balls, give or take. (Around two or three balls).
निखिल साधारण दोन किंवा तीन टाकतो असा आपण म्हणू शकतो. म्हणजे काय की दोन तीन च्या आसपास टाकतो

Chorus: Ok
चालेल

Researcher: Sometimes he scores more and sometimes less. So he scores about two to three.
कधीकधी तो जास्त टाकतो आणि कधीकधी कमीपण टाकतो. म्हणजे तो साधारण २ ते ३ टाकतो

The researcher has explicitly introduced uncertain language (sadharan) to the children in order to see how they would respond to hearing such language and whether they would be able to use the language once the researcher modelled its use in the given context. The researcher then drew the time

series plot for the second child, Priya. The following is a conversation that took place after the researcher drew the time series graph (Figure 1, top).

Researcher: About (sadharan) how many balls does Priya throw?
प्रिया साधारण किती टाकते बॉल

Asmita: Two.
दोन

Vandana: Two.
दोन

Nikhil: One to two.
एक ते दोन

Priya: Even I think one to two.
मला एक ते दोनच वाटतय

Vandana: I think two.
मला दोन वाटतं

Researcher: Now don't say about one to two. Which is better? About (sadharan) zero, about (sadharan) one, about (sadharan) two, about (sadharan) three?
आता एक ते दोन असं नाही द्यायचा उत्तर. साधारण शून्य साधारण एक साधारण दोन किंवा साधारण तीन

Vandana: (Practising the phrasing) About (sadharan) two.
वंदना साधारण

Asmita: About (sadharan) zero.
अस्मिता साधारण शून्य

The researcher discouraged the use of "one to two " and insisted that children use "about (sadharan) zero" or "about (sadharan) one" in order to see if the children would adopt the implied uncertainty in the word "about (sadharan)." However, the children appeared to use "about" as a label; that is, they just attached the word "about" to any number or they used it to indicate some value in the range. From this it would seem that the children have not yet articulated a sense of the uncertainty implied by the word "about (sadharan)."

Researcher: Let's see how to use the word sadharan (about). It's about give or take or around (using scale, plot and gestures). If I say about (sadharan) two, are her dots around two. They are down (gesture) but then are they up (gesture with a scale).
बघा हा आपण साधारण शब्द कसा वापरणार होतो. की आसपास (gesture). मी साधारण दोन असं म्हणलं तर दोनच्या आसपास आहेत का तिचे. इकडे आहेत (gesture down) पण इकडे आहेत का.

Chorus: No.
नाही

Researcher: Then I can't say about (sadharan) two. Can I say about (sadharan) zero?
मग मी साधारण दोन नाही म्हणू शकणार. साधारण शून्य असं म्हणू शकीन का?

Nikhil: No.
नाही

Asmita: We can say about (sadharan) one.
साधारण एक असं म्हणू शकतो.

Researcher: (Looking at Nikhil) Why?
(निखिल कडे बघून) का बरं

Nikhil: Not many are near zero.
कारण शून्याच्या जवळ नाही आहेत

Researcher: Most of them are above zero, right. Then what can we say?
हा बाकीचे सगळे शून्याच्या वरचे आहेत. मग मी काय म्हणू शकते?

Asmita: About (sadharan).
साधारण एक

This exchange suggests that the children were developing the concept of uncertain language, but were not yet adopting the terminology, even when the researcher introduced the word in multiple ways. This was categorized as a critical event by the researchers as this would provide a good contrast to a similar conversation that happened with the children of the other school.

The following is the conversation that followed after the researcher introduced the word using the graph. This is a place where the children seemed to be playing with the word in order to understand how it was being used in the context introduced by the researcher. They used other language as well which approaches a meaning of uncertainty. They are looking more specifically at the data.

Researcher: Then how much does Priya throw?
मग प्रिया किती टाकते?

Chorus: About (sadharan) one.
साधारण एक

Researcher: And Vandana?
आणि वंदना किती टाकत होती?

Nikhil: One to two.
एक ते दोन

Asmita: (Correcting Nikhil) About (sadharan) two. No, about (sadharan) zero.
साधारण दोन. साधारण शून्य

Researcher: Can we say about (sadharan) zero?
साधारण शून्य असं म्हटलं तर चालेल का

Chorus: No, about (sadharan) one.
नाही. साधारण एक.

Researcher: Let's write it down.

आपण लिहून ठेऊया

The children were going back and forth between using the language, not using it, and using the word as a label. Rather than being an authentic use, the children were likely using it in response to the researcher's expectation. This experimentation may have allowed them to test the use of "about" (sadharan) in context. The following conversation took place after the researcher revealed Asmita's graph.

Researcher: About (sadharan) how many does Asmita throw?
संध्या साधारण किती टाकते

Asmita: One.
एक

Vandana: One.
एक

Nikhil: One.
एक

Sandhya: About (sadharan) one.
साधारण एक

Nikhil: (In unison) About (sadharan) one.
साधारण एक

Vandana: (In unison) About (sadharan) one.
साधारण एक

Even after modelling the use of the word "about" (sadharan), we did not see any spontaneous use of the word in children's expressions. They later remembered they were supposed to use the word, but likely used it as a label in response to the researcher's expectation.

4.2. SCHOOL 2

The following is a brief excerpt between students of School 2.

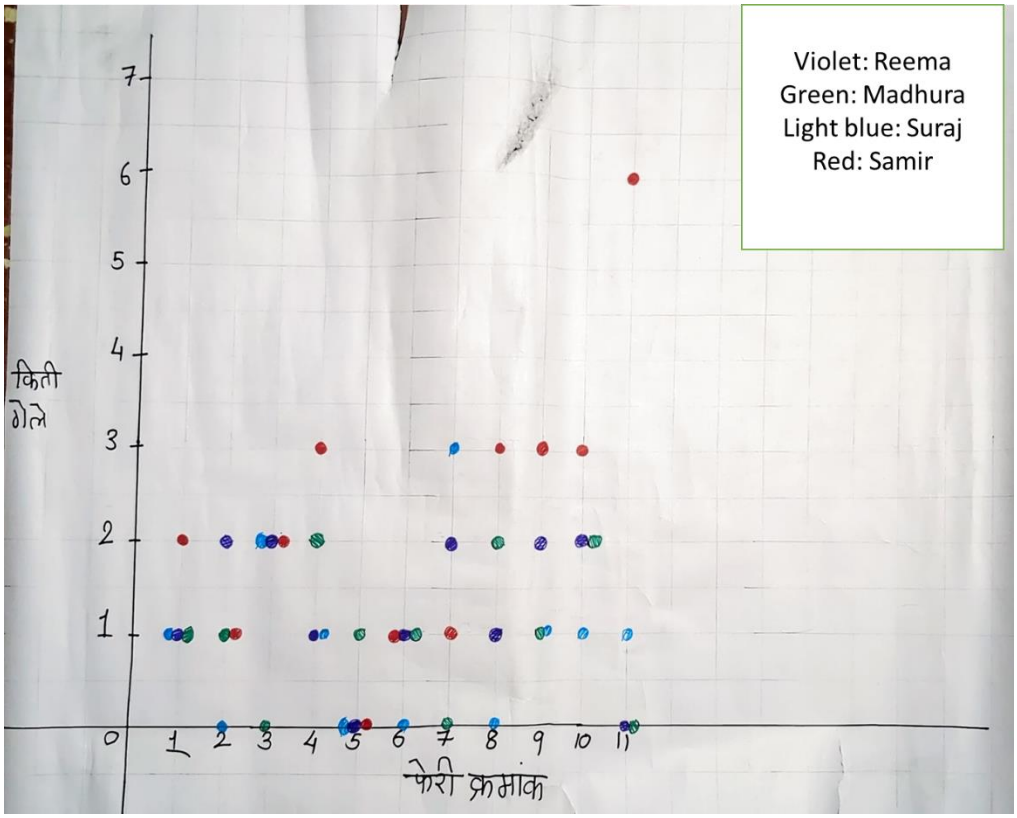


Figure 2. Outcomes of data collection in School 2.

Recall that the children in School 2 used a standard dialect of Marathi at home that aligned with standard Marathi in school (and textbooks). We provide this excerpt as a contrast to the children in the first school, whose Marathi dialect differed from standard Marathi used in school. The researcher followed a similar exercise of introducing the word "about" (sadharan) to these children as well. The researcher used the word "about" (sadharan) in a sentence and had the children use the time series.

Researcher: Can I say that Reema throws about (sadharan) three balls (holding the scale at three)?
साधरण तीन बॉल टाकते असं म्हणून चालेल

Chorus: No.
नाही

Researcher: Why not?
का बरं

Madhura: Because she has never thrown three.
कारण तीनच्याइथे कुठेच बॉल नाहीए.

Researcher: Can I say that she throws about (sadharan) two?
साधारण दोन बॉल टाकते असं म्हणून चालेल

Chorus: Yes.
हो

Researcher: Can I say that she throws about (sadharan) one ball?
साधारण एक बॉल टाकते असं म्हणून चालेल

Madhura: Yes.
हो

Samir: No, because she has thrown two 5 times (more times than she has thrown one). But maybe we can say about (sadharan) one.
नाही, कारण की पाच वेळा दोन बॉल टाकते. साधरण एक पण चालू शकतं

Madhura: Yeah, about (sadharan) one also works. Sometimes one sometimes two.
साधारण एक पण चालू शकतं कधीकधी एक कधीकधी दोन.

The children in this school appear to be trying to negotiate the meaning of the word "about." This event was categorised as critical as it contrasted to a similar event with the previous group of children. While the children in the first school simply used "about" as a label, the children here were likely trying to understand what the meaning of the word is and how it can be correctly used in the given situation.

Researcher: If I collect data again tomorrow, who do you think will score better? Suraj or Madhura? Who has a better chance of winning?
समजा मी उद्या डेटा घेतला तर सूरज जिंकेल का मधुरा जिंकेल. जिंकण्याची शक्यता कोणाची जास्त आहे

Chorus: Both.
दोघांचीही

Researcher: It's about (sadharan) the same?
सारखी आहे साधारण

Madhura: Yes, exactly, it's about (sadharan) the same. Even between the both of us (Madhura and Reema) it's about (sadharan) the same.
सारखी आहे साधारण. आमच्या दोघींमधेही सारखी आहे साधारण (मधुरा आणि रीमा)

In the last comment that Madhura makes, she has used the word "about" (sadharan) accurately and spontaneously.

5. DISCUSSION

The aim of the paper was to see how children who speak slightly different variants of the same language, Marathi, are able to express and understand uncertain language necessary for a statistics investigation. In Marathi the exact translation of the word "about" is "sadharan" and we could see that none of the children in the study were particularly comfortable using the word or any other alternative suggested by the researcher. This is an unexpected complication while navigating the expression of uncertainty that may not be easily foreseeable in the research conducted in classrooms where English is the dominant language. We see that although children from both schools followed roughly the same trajectories while reasoning about data, the experience of using an unfamiliar word was very different for both the sets of children.

The trajectory that the children followed while reasoning about data seems to be similar to what is described in the literature. However, it was much harder for children from School 1 to pick up the word when introduced by the researcher. There may several possible explanations for this. One possible explanation is that even a slight variation in the dialect can have large implications on the ease with which children are able to express uncertainty in the standard dialect. This might have long term

implications for understanding of different statistical ideas. Another explanation may be that the children who spoke the standard variant and were from privileged backgrounds know how to "play school" and were therefore able to accept the introduction of the word much quicker. They were quick to realise that the researcher had used the word "sadharan" in a way that was unfamiliar to them and were able to arrive at a common understanding of the word.

This study was conducted with two groups of children from two different in the same city and yet the core reaction to the word "sadharan (about)" was radically different. While there may be several factors as to why this is the case, it stands to reason that for a state like Maharashtra the diversity in the expression of uncertainty and the reaction to the uncertain language used by the researcher would be quite large. Large exploratory studies might be necessary to understand the extent of the differences as well as identify the commonalities that may be useful for teaching statistics. If we are going to use children's informal expression of uncertainty as a basis for introducing them to formal ideas, documenting this linguistic diversity in large qualitative studies might be a necessary first step. It may very well turn out that in some settings, explicit introduction to informal uncertain language may be extremely necessary.

One way to help children to develop an intuition for statistical ideas might be to introduce informal ideas and uncertain language in textbooks and in classrooms. Introducing informal uncertain language in suitable contexts might help students to link their informal notion of ideas like central tendency to the formalizations which are directly introduced in the higher grades. One challenge that may face such strategies would be that what constitutes an "easy" informal language may not be very obvious, especially in states like Maharashtra where there is a large variation in dialects. While the research of Kulkarni (1981) demonstrated that changing (simplifying) the language of the textbooks can bridge the gap between the achievements of socioeconomically well off and not so well of children, it might not be sufficient to just simplify the language in statistics textbooks. The children speaking a slightly different dialect might find it more difficult to use such language, especially in unfamiliar contexts. Therefore exploring this diversity while simultaneously allowing for explicit introduction of language strategies might be an important direction for future research.

ACKNOWLEDGEMENTS

We would like to acknowledge Quality Education Support Trust for providing funding to the project. We would also like to acknowledge the principals of schools for coordinating with the parents, the parents for encouraging their children, and the children themselves for participating enthusiastically in the study.

REFERENCES

Ben-Zvi, D., Aridor, K., Makar, K., & Bakker, A. (2012). Students' emergent articulations of uncertainty while making informal statistical inferences. *ZDM Mathematics Education, 44,* 913–925. https://doi.org/10.1007/s11858-012-0420-3

Benson, C. (2013). Towards adopting a multilingual habitus in educational development. In C. Benson and K. Kosonen (Eds.), *Language issues in comparative education* (pp. 283-299). Sense Publishers.

Bonnefon, J.-F., & Villejoubert, G. (2006). Tactful or doubtful? Expectations of politeness explain the severity bias in the interpretation of probability phrases. *Psychological Science, 17*(9), 747-751.

Chauhan, B. (2013). *Development of a mathematics curriculum: A culturally responsive approach.* Unpublished doctoral dissertation. Department of Education, University of Delhi.

Delpit, L. D. (1988). The silenced dialogue: Power and pedagogy in educating other people's children. *Harvard Educational Review,* 58(3), 280-298.

Dewey, J. (1938). *Logic - the theory of inquiry.* Henry Holt and Company.

Heidegger, M. (1998). Traditional language and technological language. *Journal of Philosophical Research, 23*, 129-145.

Hickman, L. A. (1990). *John Dewey's pragmatic technology.* Indiana University Press.

Karelitz, T. M., & Budescu, D. V. (2004). You say "probable" and I say "likely": Improving interpersonal communication with verbal probability phrases. *Journal of Experimental Psychology: Applied, 10*(1), 25-41. DOI: 10.1037/1076-898X.10.1.25

Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2015). Data seen through different lenses. *Educational Studies in Mathematics, 88*(3), 305-325.

Kulkarni, V. G. (1981). The effect of language barrier on science education. *Indian Educational Review*, 48-58.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education, 33*(4), 259-289.

Lesser, L. M., Wagler, A. E., Esquinca, A., & Valenzuela, M. G. (2013). Survey of native English speakers and Spanish-speaking English language learners in tertiary introductory statistics. *Statistics Education Research Journal, 12*(2), 6-31.

Lesser, L. M., & Windsor, M. S. (2009). English language learners in introductory statistics: Lessons learned from an exploratory case study of two pre-service teachers. *Statistics Education Research Journal, 8*(2), 5-32.

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal, 8*(1), 82-105.

Mokros, J., & Russell, S. J. (1995). Children's concepts of average and representativeness. *Journal for Research in Mathematics Education, 26*(1), 20-39.

Moschkovich, J. N. (1999). Supporting the participation of English language learners in mathematical discussions: The case of additive and multiplicative reasoning. *Mathematical Thinking and Learning, 1*(4), 284-307.

Pacey, A. (1994). *The culture of technology*. MIT Press.

Powell, A. B., Francisco, J. M., & Maher, C. A. (2003). An analytical model for studying the development of learners' mathematical ideas and reasoning using videotape data. *The Journal of Mathematical Behavior, 22*(4), 405-435.

Pratt, D., & Kazak, S. (2018). Research on uncertainty. In D. Ben-Zvi, K. Makar, and J. Garfield (Eds.), *International handbook of research in statistics education* (pp.193-227). Springer.

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 (ICOTS-3)* (pp. 314-319). Dunedin, New Zealand.

Torok, R., & Watson, J. (2000). Development of the concept of statistical variation: An exploratory study. *Mathematics Education Research Journal, 12*(2), 147-169.

Vygotsky, L. (2012). *Thought and language.* MIT Press.

Watson, J. (2006). *Statistical literacy at school: Growth and goals.* Lawrence Erlbaum Associates.

गणित: इयत्ता दुसरी (2019), Maharashtra State Bureau of Textbook Production and Curriculum Research.

MITALI THATTE
D-13, Flat no 4+5, Sarita Nagari 2, Sinhagad Road, Pune, India, 411030