How Children who Speak Marathi Respond to the Introduction of Uncertain Language in a Statistical Investigation

Mitali Thatte
Quality Education Support Trust
mitali.thatte@quest.org.in

Katie Makar
The University of Queensland
k.makar@uq.edu.au

This study was conducted in Maharashtra, India with children studying in a regional medium (Marathi) government school. In Marathi, the translation of the word 'about' is not very commonly used. The aim of the study was to see how the children used uncertain language about prediction while engaged in a statistical investigation and how children would respond to the uncertain language introduced by the researcher. The findings suggest that children did not use the equivalent word for 'about' without prompting from the researcher. The study has the potential of exploring and impacting the influences of language on the learning of statistics in a non-Western culture.

Being able to use uncertain language has always been an important aspect of informal statistical inference (Makar & Rubin, 2009). This includes both acknowledgement of variation in the data and to express uncertainty regarding future events. However, language is a social phenomenon and social interactions play an important role in the development of language, particularly social interactions with a significant adult (Vygotsky, 2012). It is reasonable to assume that children 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 between a researcher and children a school in Maharashtra, India while the children were engaged in a statistical investigation. The medium of instruction was Marathi, which is the same language that the children spoke at home, although children's home dialect was slightly different than the standard dialect used in school textbooks. The children belonged to a socio-economically disadvantaged section of the society. 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.

Literature Review

In this section, we look at language difficulties in statistics that may influence how non-English learners articulate their understanding of likelihood. We also consider how language encodes culture, power and equitable access to learning.

Articulating Uncertainty

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 used, as well as navigating the variety of ways to express likelihood and uncertainty due to variability and estimations in future predictions (Karelitz & Budescu, 2004). "The situation in which random variation is met influences how people think about probability" (Pratt & Kazak, 2018, p. 214). Researchers noticed that students often articulated predictions under uncertainty by either being overly deterministic or overly relativistic (Ben-Zvi et al., 2012; Rubin et al., 1990). In making predictions from data, the primary students in Ben-Zvi et al.'s (2012) study oscillated between certainty-only (deterministic) and uncertainty-only (relativistic) statements before further exploration assisted them to express uncertainty with emerging probabilistic language (about, maybe). Therefore, attending to expressions of uncertainty requires attention to understanding of uncertainty, how it is expressed and importantly, the relationship with the context.

When considering how non-English language learners articulate uncertainty, challenges may be related to statistical concepts, distinctive statistical language, both or neither. It can be difficult to tease out where the challenges lie. For example, one challenge of exploring statistical ideas with uncertain language in Marathi (the language in this study) is that the use of some words in English do not translate well. For example, there is no commonly used translation of the word 'about' (such as in 'I will sleep for about 7 hours every day'). It is on this issue of linguistic diversity that the researchers conducted a statistical investigation with children.

In research with non-English speaking students, additional constraints arise as how one expresses statistical ideas is also related to the specifics of the language or dialect. Lesser and Winsor (2009) worked with Spanish-speaking pre-service teachers in an English-speaking environment. They argued that with second language learners, not enough emphasis is placed on statistical *register*, a subset of language relevant to a particular context and purpose. Learners' proficiency in everyday language may be quite different from de-contextualised academic situations, which can lag behind by three or four years. Lesser and Winsor's 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 attending to how a student's language articulates an idea and to incorporate their language, community, and culture into learning activities rather than see these as limitations. Other recommendations included using wait time, embedding instruction across several contexts, seeking multiple ways to express and represent a new idea, and encouraging collaboration. In related work, Lesser and his colleagues (2013) focused on different aspects of statistical register (field, mode, tenor) to identify how students in different language groups understood statistical concepts that were articulated in each language. 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.

Beyond Language as Articulation: Culture, Power, and Equitable Access

Cultural dimensions of language can contribute significantly to learning statistics. For example, Chauhan's (2013) research showed that children with fatalist 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 their command of the dominant language (Benson, 2013). In education, if a learner is not fluent in the dominant language of instruction 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.

Kulkarni (1981, p. 55), while talking about the barriers that children face learning scientific concepts, especially children from lower socio-economic communities, stated that "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 in school was an entirely different language, but when the kind of language (formality, syntax etc) used in school was different than what the child used at home. They noted that in changing the language of textbooks to better suit the home language, children's scholastic performance improved significantly; a significant finding was the removal of disparity in the performance of students coming from different social economic backgrounds.

Encouraging children to use uncertain language in statistics classrooms, especially in socioeconomically and linguistically diverse communities is a challenging process. 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, Delpit (1988) argued that we should accept children's home expression and also teach them standard (academic) language in order not to be disadvantaged. This issue is especially

relevant as learning is thought to be a largely social process where interactions with a significant adult has a large impact on 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 might be highly relevant when the children speak a different dialect than the standard one spoken by the teachers or expected in textbooks. This was a key issue that the researchers sought to address when it comes to natural expressions of uncertainty.

Methodology

The activity was conducted with four students in Grades 5 and 6 (ages 9–11). It 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. The competition would involve throwing a ball into a bucket and whoever got the ball in the greatest number of times would win the competition. The researcher told the 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. With a nudge from the researcher, they decided that they would 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 actually landed in the bucket. Based on these data the children had a discussion facilitated by the researcher in order to decide who to send to the competition.

The data collection consisted of audio and video of the activities, artefacts that were created (e.g., tallies or graphs) and field notes. The analysis was based on an adaption of Powell et al.'s (2003) analysis of video data. 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). 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 a challenge or insight into the children's articulation of uncertainty. The pieces of transcript were 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.

Results

The following is an interaction between the researchers and students after they had drawn a time series graph for Nikhil (Figure 1). The researcher has asked, "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

Plot for Nikhil: The x Axis Shows the Number of the Trial While the y Axis Shows the Number of Balls That Went in the Bucket in That Particular Trial
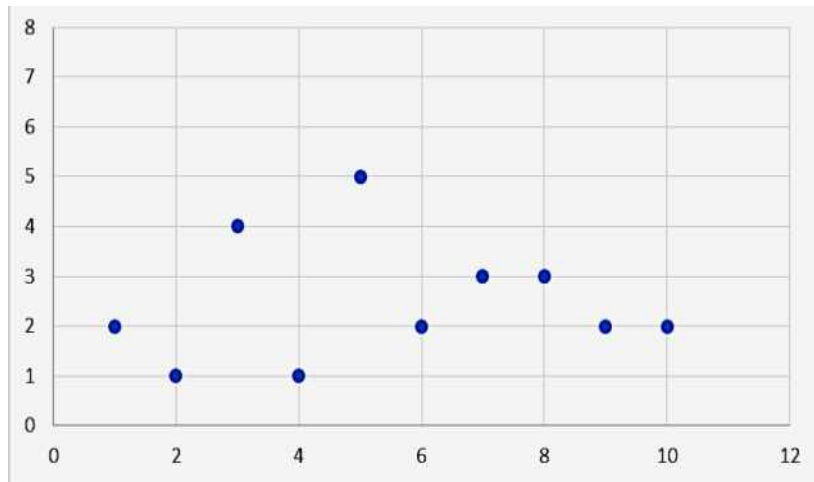
Researcher:	About (sadharan) how many balls does Nikhil throw? He has ten chances right?	साधारण किती टाकतो. त्याला १० chances दिलेत ना?
Vandana:	25	25
Researcher:	He has 10 chances right? About (sadharan) how many balls will he get in?	त्याला १० chances दिले ना? तो दहापैकी किती टाकेल?
Vandana:	He got 2 in the first round, 3 in the second round, 4 in the third round, 1 in the fourth round, 5 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 questions in textbooks generally ask students to count frequency, or add up values, it seemed likely that children were 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 daily use in their dialect. Considering that children did not adopt the word sadharan, the researcher tried to use another word as a substitute for sadharan which is the word '*andaje*' 'अंदाजे'. 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). Andaje is a broader word better suited to expressing uncertainty about events rather than data, sometimes used as a combination of guess and estimate. For example, the predictive statement, "It will take me about (andaje) 2 hours to reach my destination", but not the descriptive statement "An average child scored about (andaje) 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:	4	4
Priya:	5	5
Researcher:	Why do you think 5?	तुला का असं वाटतंय ५ टाकेल?
Priya:	His aim is very good	त्याचा नेम चांगला आहे

The children used their experience seeing Nikhil throw the ball and were optimistic in predicting how many Nikhil would throw next time. Since the children used neither word as a translation of the word 'about' as the researcher expected, the researcher decided to explicitly introduce the word sadharan and model how it is used in statistics. The following excerpt shows how the researcher led the discussion to create good 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:	5	5
Priya:	3 (Immediately changed her answer from 5 to 3)	3
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 would throw, the children gravitated towards using the central tendency of the data when the researcher phrased the question about making a bet (preduction). The researcher and the children had a discussion on which number to bet on, if Nikhil gets 10 chances to throw the ball in the bucket. Then the researcher introduced the word sadharan as they were curious to see how the children would react to the introduction of the word (which is reasonably common in Marathi but apparently not used by the children). The following conversation is from the later half of the activity.

Researcher:	Suppose I bet that Nikhil throws 7. Will I win or lose?	समजा अशी मी पैज लावली की निखिल सात टाकेल. तर मी जिंकी का 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 हरिन
Chorus:	Win	जिंकेन
Researcher:	Can I say Nikhil throws about (sadharan) 6 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 6 even once.	कारण त्यांनी एक पण सहा टाकलेला नाही आहे
Researcher:	About (sadharan) how many balls does Nikhil throw?	हा तर निखिल साधारण किती टाकतो?
Priya:	2 or 3	दोन नाहीतर तीन
Researcher:	Yes, we can say that Nikhil throws about (sadharan) 2 or 3 balls. Meaning, he throws 2 or 3 balls, give or take. (Around 2 or 3 balls)	निखिल साधारण दोन किंवा तीन टाकतो असा आपण म्हणू शकतो. म्हणजे काय की दोन तीन च्या आसपास टाकतो

The researcher 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 data of the second child, Priya (Figure 2).

Figure 2

Graph of the Number of Successes (y Axis) in Each Round (x Axis) for All the Children (Colour)
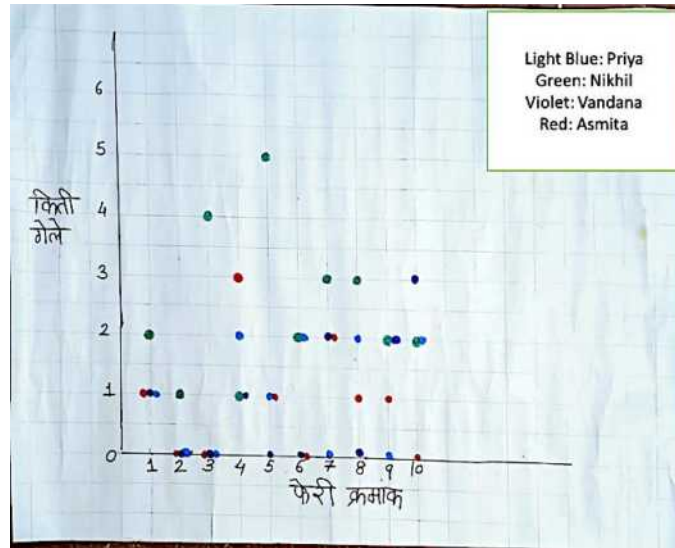
The following is a conversation that takes place after the researcher has drawn the time series graph for Priya.

Researcher	About (sadharan) how many balls does Priya throw	प्रिया साधारण किती टाकते बॉल
Asmita	2	2
Vandana	2	2
Nikhil	1 to 2	एक ते दोन
Sadhana	Even I think 1 to 2	मला एक ते दोनच वाटतय
Vandana	I think 2	मला दोन वाटतं
Researcher	Now don't say about 1 to 2. Which is better. About (sadharan) 0, about (sadharan) 1, about (sadharan) 2, about (sadharan) 3	आता एक ते दोन असं नाही द्यायचा उत्तर. साधारण शून्य साधारण एक साधारण दोन किंवा साधारण तीन
Vandana	(Practising the phrasing) About 2	साधारण 2
Asmita	About 0	साधारण शून्य

The researcher discouraged the use of "1 to 2" and asked that children use "about *(sadharan)* 0" or "about *(sadharan)* 1" in order to see if the children would adopt the implied uncertainty in the word "about *(sadharan)*". But it would appear that the children likely used "about" as a label; that is, they just attached the word "about" *(sadharan)* before 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). Its about give or take or around (using scale, plot and gestures). If I say about (sadharan) 2, are her dots around 2. They are down (gesture) but are they up (gesture) With scale | बघा हा आपण साधारण शब्द कसा वापरणार होतो. की आसपास gestures. 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 2. Can I say about 0?	मग मी साधारण दोन नाही म्हणू शकणार. साधारण शून्य असं म्हणू शकीन का?
Nikhil:	No	नाही
Asmita:	We can say about 1	साधारण एक असं म्हणू शकतो
Researcher:	(Looking at Nikhil) Why?	(निखिल कडे बघून) का बरं
Nikhil:	Not many are near 0	कारण शून्याच्या जवळ नाही आहेत
Researcher:	Most of them are above 0, right. Then what can we say?	हा बाकीचे सगळे शून्याच्या वरचे आहेत. मग मी काय म्हणू शकते?
Asmita:	About 1	साधारण एक

This exchange suggests that the children were developing the concept of uncertain language, but not yet adopting the terminology, even when the researcher introduced the word in multiple ways. The following conversation followed after the researcher introduced the word using the graph. This is where the children seemed to be playing with the word to understand how it was being used in the context introduced by the researcher. They are using other language also which is going more towards uncertainty. They are looking more specifically at the data.

Researcher:	Then how much does Priya throw?	मग प्रिया किती टाकते?
Chorus:	About 1	साधारण एक
Researcher:	And Vandana?	आणि वंदना किती टाकत होती?
Nikhil:	1-2	एक ते दोन
Asmita :	(Correcting Nikhil) About 2. No, about 0	साधारण दोन. साधारण शून्य

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

Researcher:	About how many does Asmita throw	अस्मितासाधारण किती टाकते
Asmita:	1	1
Vandana:	1	1
Nikhil:	1	1
Chorus:	About 1	साधारण एक

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

Discussion

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 in 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 research conducted in

classrooms where English is the dominant language. The trajectory that the children followed while reasoning about data seemed to be the same one described in research. However, it may have been unreasonable to assume the children would pick up the word introduced by the researcher. This is a challenge while informally dealing with statistical ideas. It was clear that even though the researcher and children spoke the same language, a language barrier existed between the researcher and the children. In the state of Maharashtra there are multiple dialects of the state language of Maharashtra. There is a standard variant used in textbooks which is generally understood by a large percentage of the population. However, adults and children speak variants of this language while conversing with each other. Although the researcher did not use any technical terms, the language of the researcher was closer to the dominant language written in textbooks and the children's dialect did not appear to utilise the same idea. While we completely acknowledge that there will exist other ways of expressing uncertainty that the children may be familiar with, the lack of similarity to dominant language of school may pose pedagogical challenges in introducing the informal ideas such as informal inference.

Acknowledgements

References

Ben-Zvi, D., Aridor, K., Makar, K., & Bakker, A. (2012). Students' emergent articulations of uncertainty while making informal statistical inferences. *ZDM Mathematics Education*, *47*, 1107–1120. https://doi.org/10.1007/s11858-012-0420-3

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

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.

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

Kulkarni, V. G., & Gambhir, V. G. (1981). The effect of language barrier on science education. *Indian Educational Review*, *16*, 48–58.

Lesser, L., Wagler, A., Esquinca, A., & Valenzuela, M. (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., & Windsor, M. (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.

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, & 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. In D. Vere-Jones, S. Carlyle, & B. P. Dawkins (Eds.), *Proceedings of the third international conference on teaching statistics* (ICOTS-3) (pp. 314–319). Dunedin, New Zealand: International Statistical Institute.

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