English versus Native Language for Higher Education in Computer Science: A Pilot Study

Taj Muhammad Khan
Faculty of Computer Sciences and Engineering,
Ghulam Ishaq Khan Institute
Topi, Pakistan
taj.khan@giki.edu.pk

Syed Waqar Nabi
School of Computing Science, University of Glasgow
Glasgow, United Kingdom
syed.nabi@glasgow.ac.uk

ABSTRACT

While the official language of instruction for higher education programs in Pakistan is English, in practice it can vary from English-only, to some combination of English and a native language, to native language only. There is a lack of consistency not just across institutions, but also across classrooms in the same institution. We have conducted a pilot study based on a small cohort of computer science students, who happened to have some of the lectures of the same course delivered in Urdu, and the others in English. Based on a questionnaire response with both quantitative and free text questions, we investigated students' self reported oral and written communication skills, as well as their preference of language for lectures. We found, albeit based on a limited, pilot study, that using Urdu, the first language of most of our students, as the medium of oral communication (lectures, general classroom communication) should be preferred, whereas English should remain the language of choice for written communication. We expect to expand this study to multiple classrooms across institutions in Pakistan and other countries operating in a similar context, so that more definitive conclusions can be derived, potentially leading to better informed, evidence-based policies and practices in the future.

KEYWORDS

higher education, language of instruction, english vs native language

1 INTRODUCTION

There is a strong emphasis on English being the preferred language of instruction in the higher educational institutions (HEIs) in Pakistan mostly for two reasons: 1) English is the *lingua franca* if one wishes to participate in science and technology milieus or the job market at international level; 2) For historical reasons (British colonial legacy) English is considered the language of the elite and has a social prestige associated with it. Therefore the government as well as parents want the students to be educated in English. We posit the thesis that while the goal itself is worthwhile, the current state of its implementation is failing to achieve its target and may be doing more harm than good.

Pakistani population is an amalgamation of multiple different ethnic groups. Urdu, originally the language of immigrants from India after the 1947 partition, was declared as the national language to serve as a binding force for the nation. Apart from Urdu, there are at least 4 other major regional languages: Punjabi (and its close cousin, Siraiki), Pashto, Sindhi, and Balochi, and many other smaller languages. English is the second or third language for most people.

While there are efforts towards mother tongue based education at primary level, the language of instruction in Pakistan, especially high school and above, is a convoluted mixture of English and Urdu. Very broadly, schooling can be categorized in two groups: private sector – English medium and expensive, and public sector – Urdu medium and inexpensive. These two streams (English and Urdu medium) are the main sources of students joining the HEIs.

In HEIs the medium of instruction, due to reasons cited earlier, is officially English. This does not take into account the fact that students coming from Urdu medium background, and even many from English medium, are not sufficiently well versed in the English language. Thus the medium of instruction (English) can become *extraneous cognitive load* for these students and adversely affects their learning. This phenomenon can be particularly acute in the first year or so, when the foundational concepts are taught.

Although Pakistan has made great progress in the field of higher education and research, an evidence-based comprehensive policy about medium of instruction in higher education is absent. Apart from some work at the primary and secondary education levels, briefly discussed in the next section, there is little to no prior work or study – anything that could be counted as "evidence" – to guide the formulation of policies at the higher education level. As a result, the policies about medium of instruction, e.g., enforcing English without taking into account students' and even the teachers' English language skills, are accompanied by patchy and inconsistent compliance. If the evidence suggests that English presents an extraneous cognitive load for the students, then room could be made in the policies to use Urdu explanations as *scaffolding*, for the time necessary to adapt to the new medium of instruction. Such informed policies will prevent the students from losing out on basic concepts which are crucial to success of their education and career.

We present here findings from a pilot study of a small cohort of students showing some clear trends that lead to tangible observations – resonating with our own experiences as students as well as teachers – but the size of the study is currently too small to warrant more generalized conclusions. Our purpose here is: firstly, to present this to the wider community as a step towards scaling this study , and secondly, to encourage dialogue amongst academics in Pakistan and other countries with a similar language conundrum, so that practices and policies are well-informed and evidence based.

2 RELATED WORK

Since 1953 UNESCO [1] has been promoting the use of mother tongue in the early education of children. There is extensive literature [2] showing not only that children are more likely to succeed in school when taught in their mother tongue [3] but also that the parents' participation improves by being able to communicate with the child as well as the teacher about educational topics [4].

In the Pakistani context it has been acknowledged that the use of English at primary level hinders learning and that children's education should start in their mother tongues and other languages can be introduced later in a step-wise fashion [5], [6], [7].

Sabiha Mansoor [8] catalogues the language shift from regional languages to Urdu and English especially in the major cities of Pakistan. She highlights the public sector, where the medium of instruction is Urdu, and private sector, where the medium of instruction is English, education system bifurcation in Pakistan.

Another perspective on this work is to view it from the perspective of cognitive workload theory [9], which divides cognitive load into three types: *Intrinsic* - inherent to the problem, *Germane* - for constructing schemas to store knowledge, and *Extraneous* - the manner in which the information is presented to the learner. The first two are useful but the extraneous cognitive workload plays the role of *noise* and should be minimized. We believe the use of English as medium of instruction to teach non-native learners adds to the extraneous cognitive workload and thus impedes the learning.

To estimate the gamut of difficulties and barriers that non-native English speakers face when they learn programming, Philip J. Guo [10] surveyed the users of a very popular introductory programming website (www.pythontutor.com). The respondents reported problems when reading instructional materials, doing technical communication, reading and writing code; all are essential steps when learning to program. They wished for technical materials to be available in simplified English free of cultural slang, use of more visuals, and code examples which are culturally agnostic.

John Airey [11] did his Ph.D. thesis on teaching undergraduate Physics courses in Swedish as well as in English. He found that when taught in English, students tend to ask and answer fewer questions. They also reported that in English they were not able to follow the lecture and take notes simultaneously since it took them time to process the English sentences in their minds.

To gauge the effect of native language on learning programming in the classroom environment versus the use of screencasts, Pal and Iyer [12] [13] conducted 4 six-day workshops teaching basic programming concepts to students coming from Hindi-medium schooling backgrounds: 2 groups were taught in a live classroom environment (one in Hindi and the other in English) and 2 groups were taught using screencasts recorded in Hindi and English respectively. In both the live classroom as well as screencasts they found statistically significant evidence that when the students hailing from Hindi-medium background are taught in English, there is an adverse impact on their learning. They also found that when these Hindi-medium students were taught in English, having Hindi-language screencasts for offline viewing helped improve their learning.

On the role of native language when learning CS concepts, Soosai Raj et al [14] [15] experimented teaching linked-lists and data structures to two different groups of students: to one in pure English and to the other in a mixture of English and Tamil, using *code switching* as well as *translanguaging*. They did not find a statistically significant improvement in the results of students who were taught in both languages. They did, however, note that the class interaction in the English-Tamil group was more animated and more than 90% of the questions in that group were asked in the Tamil language indicating that students tend to engage more when they are in their comfort medium. All the students who were exposed to the bilingual teaching methodology preferred it over English.

We are unaware of any study targeting the use of native languages in the Computer Science Education, in Pakistan.

3 RESEARCH QUESTION (RQ)

What language do students of a CS undergraduate program in Pakistan, an ex-colonial country with English as an official but not native language, prefer as the medium of instruction, oral and written?

4 CONTEXT AND METHODOLOGY

This study took place at the end of a course on Object-Oriented Programming in the 2nd semester of a BS in Computer Engineering program offered at an engineering institute. This 4-credit (3 lecture hours and 3 lab hours per week) course was a follow up to the introductory Computer Programming course.

The Spring 2021 semester started physically for about 5 weeks, with Urdu as the main language of oral communication in the classroom. Then, due to the 3rd COVID-19 wave, the classes continued online for the last 9 weeks, accompanied by a shift to using English-only for all online lectures. Live lectures, were conducted on MS Teams, recorded, and shared with students on YouTube.

Instructor and students' video sharing was off, for both bandwidth and cultural reasons. The students would see the instructor's screen and hear his voice. On their end the students could either unmute their mics to ask questions or comment, or do the same by typing in the shared MS Teams chat. Commenting in the chat was the dominant mode of communication from the students' side.

This study assesses student preferences in view of the change in the language of instruction from Urdu to English midway through the course. The official language of instruction in the institute is English. But during the physical instruction in classrooms, due to cultural reasons, both the instructor and the students would often slip into Urdu because they felt more at home in it.

In transitioning online, due to official bureaucratic reasons, the medium of instruction was deliberately kept English. It was during these lectures when some students asked for certain concepts to be repeated in Urdu that the seeds for this study were sown.

These students underwent two transitions simultaneously: going from physical-to-online, they also experienced a language switch from Urdu-to-English. We polled the students with a questionnaire to quantify how did their learning get affected with these simultaneous transitions, keeping the focus on the language transition.

The population comprises 33 students (out of a class of 35) who agreed to fill the questionnaire. All the students sat in the same course, taught by the same instructor, using a combination of Urdu and English as discussed earlier.
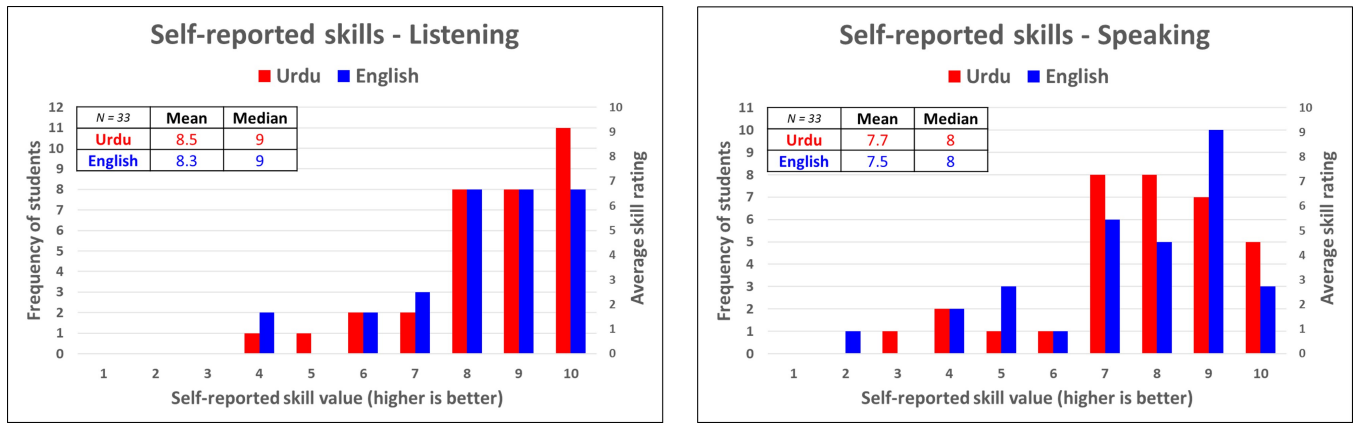
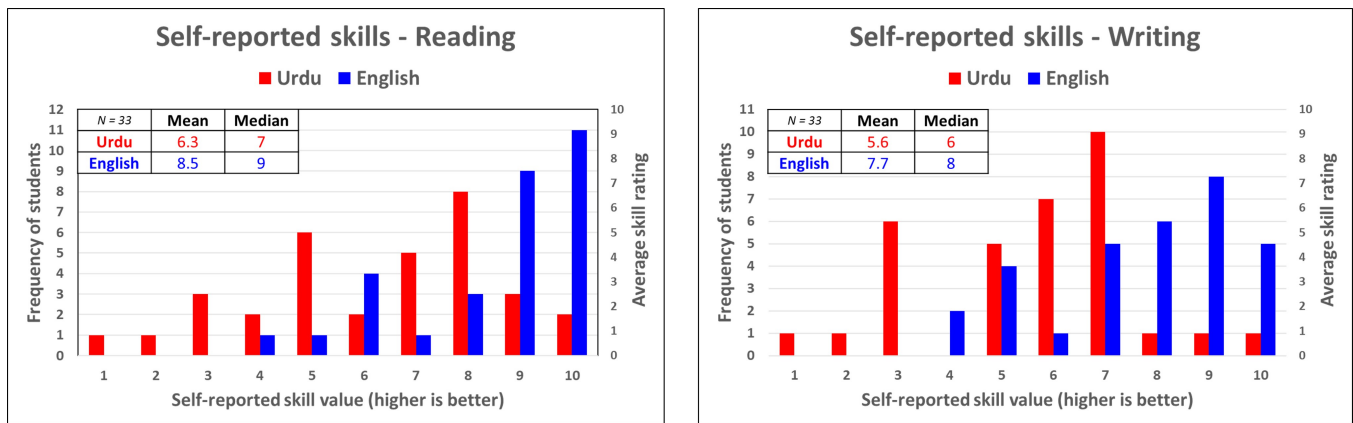**Figure 1: Self-reported language skills - oral communication**

Figure 2: Self-reported language skills - written communication

The online anonymous questionnaire consisted of three types of questions: (1) Quantitative questions about the students' language skills on a scale of 1-10; (2) Questions structured for the Likert scale; (3) Free text questions. The questions related to transition from Urdu to English, as well as from physical to online classes.

5 RESULTS AND ANALYSIS

Three quarters of the students reported Urdu as their first language, while 15% listed Pashto, a prominent local language of Pakistan, native in the area where the host university is based. Two students out of 33 indicated English as their first language, unusual in Pakistan.

All students did study English at school, as is the norm in Pakistan, with most students (73%) started studying English in pre-school, or class 1 or 2 (15%).

5.1 Self-reported language skills

We asked all students to assess their own language skills, both in English and Urdu, and do this separately for their oral (listening and speaking) as well as written (reading and writing) communication skills. Majority reported Urdu as their first language, but still 8 (24%) students reported Pashto or English as their first language.

The result of self evaluation of oral (listening and speaking) skills, in Urdu and in English, are shown in Figure 1. A number of interesting observations can already be made:

Across languages, while the mean of both listening and speaking skills were slightly higher for Urdu, the differences are minor, and the median value is actually the same. So at least when it comes to self-perception of oral communication skills, Urdu may be the slightly stronger language for most, but not by much. This was a somewhat surprising finding for us, as we were expecting a more significant difference between Urdu and English skills. We will revisit this when we look at the language preference of students when listening to lectures, where we see a much clearer preference.

When comparing listening to speaking, the mean for both Urdu and English is one point lower for speaking skills; another way to look at this comparison is that relatively more students reported their speaking skills at 5 or less – 4 students for Urdu, 6 for English – compared to those reporting their listening skills at 5 or less – 2 each for Urdu and English. Students seem more confident of their listening skills than speaking skills, whether in Urdu or English. The result of self evaluation of written (reading and writing) skills,
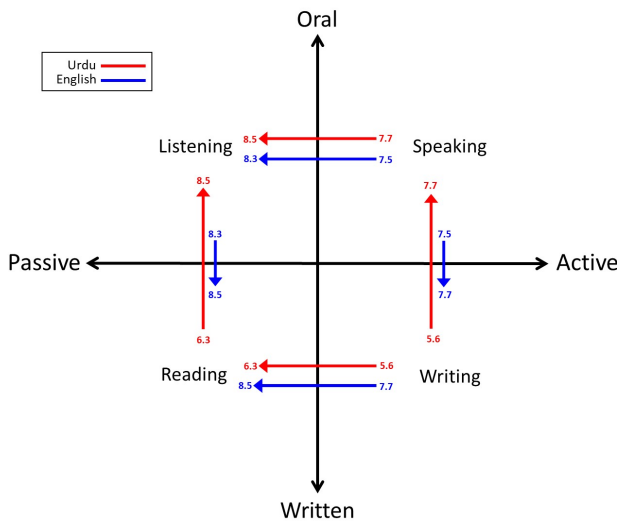
Figure 3: The Passive-Active and Written-Oral axes give rise to the four quadrants of speaking, listening, reading and writing. The direction of the arrows indicate increasing skill, and the length approximates difference in average scores.

Figure 4: Language preference for class lectures

in Urdu and in English, are shown in Figure 2. Looking at these results along with those in Figure 1, we can observe:

- While Urdu was the "winner" in self-reported oral communication skills, English is the more dominant language when it comes to written communication skills.
- English is not just the stronger language for written communication, its margin over Urdu is much more significant than that of Urdu over English for oral communication skills.
- A significantly larger number of students report their written language skills in Urdu as quite weak; 13 students score themselves at less than or equal to 5 for both reading and writing skills, as opposed to 2 and 4 students for listening and speaking skills.
- Very interestingly, this significant difference between oral and written communication skills that we observed in Urdu is absent in English. The mean of English oral communication skills and written communication skills is almost the same, and the number of students reporting their skills at less than 5 for listening and reading skills one one side (2 and 2), and reading and writing on the other (6 and 6) is exactly the same.

We found it instructive to visualize these four skills along two orthogonal dimensions: the written-oral dimension, and the active-passive dimension, leading to the four quadrants of speaking, writing, listening and reading, as shown in Figure 3. Once visualized this way, we can revisit some of our previous observations as follows:

- For both English and Urdu, passive communication skills are rated higher than active, whether oral or written.
- For Urdu, oral communication skills are rated higher than written communication, whether active or passive, whereas for English, the direction of change is opposite, with written communication skills rated higher than oral, but by a smaller margin.
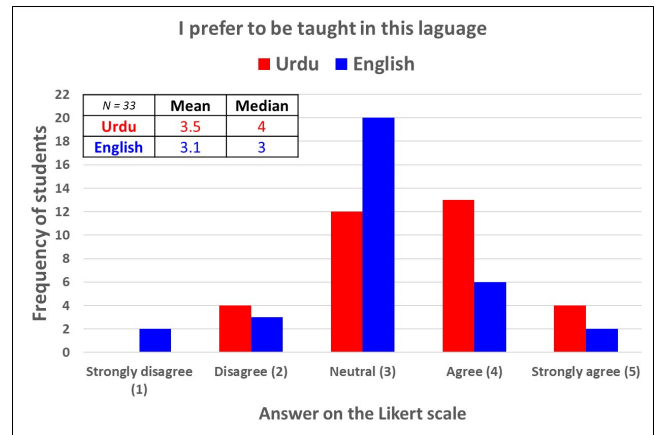
- The largest difference in any transition is when moving from written to oral skills in Urdu, whether active or passive.

We can now make some more general observations, being wary that the small size of study may have context-specific effects skewing the results that would have been evened out in a larger study.

Students feel more confident in passive communication – listening and reading – as compared to active communication – speaking and writing – irrespective of language. This in our view was an expected observation, as it is natural to feel more hesitant in any kind of active communication . For the native language for most, Urdu, oral skills were rated higher than English. This is inline with the observation that most of these students have spoken and listened to Urdu since childhood. When it comes to written skills though, English was not just rated higher, but by a higher margin. While this may look counterintuitive to some, this too will ring true to people from countries where English is not a native language, but where text books and written exams for technical subjects are in English. Thus it's natural that most students feel more confident of their written skills in English than in Urdu.

5.2 Language preference for lectures

We asked the students to rate how much they would like to be taught in a particular language, using the 5-point Likert scale. The results are shown in Figure 4. As expected, students showed a clear preference for use of Urdu as the language of communication for lectures. E.g. the median value for Urdu is "agree" and English is "neutral" when asked if they would prefer to be taught in this language. Recall that these students were taught both in English and then in Urdu for the same course and by the same teacher. It should be clarified here that when we talk about Urdu as the language used for lectures, the technical terms are still said in English.

5.3 Class engagement

We had hypothesized that students will be less willing to ask questions in class and otherwise engage with lectures if they were restricted to talk in English. Such restrictions are quite common in many schools, colleges and universities in Pakistan. We placed a
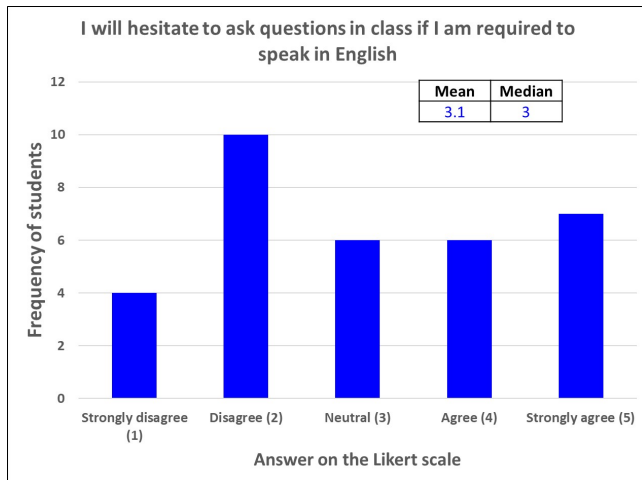
Figure 5: Gauging the dependence of class engagement on whether or not students are restricted to speak in English

question in the questionnaire specifically to get more evidence for this hypothesis, and the results are shown in Figure 5.

This corroborates previous research [14] [15]. There are a significant number of students, 13 out of 33, or 39%, who agreed or strongly agreed that they would indeed hesitate to ask questions if required to speak in English. It is our view that the agreement of more than a third of the respondents is a significant percentage of the class that would likely end up disenfranchised if forced to communicate in English in the classroom.

6 DISCUSSION AND CONCLUSION

While the limited scale and context of this study means it is not possible yet to derive any definitive, statistically significant conclusions, we can offer a tentative answer to the RQ as follows: The students prefer oral communication in Urdu over English, and may hesitate to engage in class if forced to speak in English. On the other hand, students reported higher levels of skills for English in written communication, and may prefer it for reading and writing.

It is our hypothesis that the following approach for Pakistani computer science classrooms of higher education will lead to better class engagement and improved learning:

- English should not be enforced on either the teacher or the students for oral communication in the classroom. A common native language, Urdu or even a regional language, should be allowed, and perhaps even encouraged.
- The established practice of using English texts, as well as conducting written exams in English seems a sensible choice, and based on our study, there is no reason to revisit it.

It should be noted here that there are other, wider considerations around the choice of language too, e.g. the impact on prospects of Pakistani graduates to pursue higher studies or find jobs internationally, that have not taken into account.

We expect that our hypothesis will find some resonance not just amongst other academics in Pakistan teaching computer science subjects at university, but also teachers of other technical subjects,

not just in Pakistan, but in many other ex-colonial countries also. The prior work we discussed in section 2 supports this expectation. What is now required is many more similar and larger studies that support and/or improve our hypothesis, to the point where it may be counted as *evidence*. We feel that with Pakistan and many other developing countries looking to grow their technological expertise in the next few years and decades, basing the higher education policy around language of instruction on sound evidence is crucial.

Our main contribution in this work is that we have begun a line of investigation which we believe is crucial to the success of higher education in Pakistan in general and STEM areas in particular. We have observed some clear trends based on quantitative data, albeit from a relatively small cohort of students in a very specific context. It remains to be seen if these trends are maintained for larger studies that cancel out context-specific variability, to the point where generalizations can be made at the national level at least, and possibly even across nations. If they are maintained, or even if new trends emerge, they should have far reaching implications on the choice of language and policies for higher education sector in Pakistan. We plan to continue this investigation.

Data Availability Statement. The questionnaire used in this study, along with the data, is available in the Open Science Framework repository here: https://osf.io/zbdmx/

REFERENCES

[1] Unesco, *The Use of Vernacular Languages in Education*, ser. Bilingual-bicultural education in the United States. Unesco, 1953. [Online]. Available: https://books.google.co.bw/books?id=Ilc_MQAACAAJ

[2] J. Ball, "Educational equity for children from diverse language backgrounds: mother tongue-based bilingual or multilingual education in the early years: summary," 2010.

[3] K. Kosonen, "Education in local languages: Policy and practice in south-east asia," *Asia-Pacific Programme of Education for All First Language First: Community-based Literacy Programmes for Minority Language Contexts in Asia. Bangkok: UNESCO Bangkok, 2005.*, p. 96, 2005.

[4] C. J. Benson, "Real and potential benefits of bilingual programmes in developing countries," *International Journal of Bilingual Education and Bilingualism*, vol. 5, no. 6, pp. 303–317, 2002.

[5] H. Coleman, "Teaching and learning in pakistan: The role of language in education," *Islamabad: The British Council*, pp. 1–56, 2010.

[6] H. Khalique, "The urdu-english relationship and its impact on pakistan's social development," 2007.

[7] F. Shamim and U. Rashid, "The english/urdu-medium divide in pakistan: Consequences for learner identity and future life chances." *Journal of Education and Educational Development*, vol. 6, no. 1, pp. 43–61, 2019.

[8] S. Mansoor, "The status and role of regional languages in higher education in pakistan," *Journal of Multilingual and Multicultural Development*, 2004.

[9] P. Chandler and J. Sweller, "Cognitive load theory and the format of instruction," *Cognition and instruction*, vol. 8, no. 4, pp. 293–332, 1991.

[10] P. J. Guo, "Non-native english speakers learning computer programming: Barriers, desires, and design opportunities," in *Proceedings of the 2018 CHI conference on human factors in computing systems*, 2018, pp. 1–14.

[11] J. Airey, "Science, language, and literacy: Case studies of learning in swedish university physics," Ph.D. dissertation, Acta Universitatis Upsaliensis, 2009.

[12] Y. Pal and S. Iyer, "Classroom versus screencast for native language learners: Effect of medium of instruction on knowledge of programming," in *Proceedings of the 2015 ACM conference on innovation and technology in computer science education*, 2015, pp. 290–295.

[13] Y. Pal and S. Iyer, "Effect of medium of instruction on programming ability acquired through screencast," in *2015 International Conference on Learning and Teaching in Computing and Engineering*. IEEE, 2015, pp. 17–21.

[14] A. G. Soosai Raj, K. Ketsuriyonk, J. M. Patel, and R. Halverson, "Does native language play a role in learning a programming language?" in *Proceedings of the 49th ACM technical symposium on computer science education*, 2018, pp. 417–422.

[15] A. G. S. Raj, H. Zhang, V. Abhyankar, S. Mukerjee, E. Zhang, J. Williams, R. Halverson, and J. M. Patel, "Impact of bilingual cs education on student learning and engagement in a data structures course," in *Proceedings of the 19th Koli Calling International Conference on Computing Education Research*, 2019, pp. 1–10.