Broadening participation of biology students in computing: a mixed methods study among bioinformatics students

1st Thom Kunkeler
Department of Information Technology
Uppsala University
Uppsala, Sweden
thom.kunkeler@it.uu.se

2nd Aletta Nylén
Department of Information Technology
Uppsala University
Uppsala, Sweden
aletta.nylen@it.uu.se

***Abstract*—Despite the rapid evolution of bioinformatics over the past decades, a lack of computing skills and a widening skill gap among biologists can be observed. It is therefore of increasing importance to understand why biologists are motivated in studying bioinformatics and computing, and what the challenges are that they face. To this end, a computing capital mixed methodology has been developed, inspired by the concept of science capital. A total of 13 master's level bioinformatics students with a biology background participated in the computing capital survey, of which 10 agreed to a follow-up interview. Descriptive statistics and thematic analysis suggest that, while biology students are driven by career opportunities, as well as positive beliefs and values when it comes to computing, they perceive barriers such as low levels of computing confidence, and a gap between their skills and future careers. Future research can build onto the developed methodology, and explore other aspects of capital and their role in computing education. The aim of such efforts is to broaden participation in computing education by developing an understanding of why people engage with computing, and what they aspire to do with it.**

***Index Terms*—broadening participation, computing education, computing capital, motivation, challenges**

I. Introduction

The field of bioinformatics can be traced back to 1952 when the role of DNA as a genetic information encoding molecule was validated [1, 2]. Some important milestones over the past few decades include the use of computers to determine protein primary structure [3-5], DNA sequencing [6], a complete genome sequencing of a free-living organism [7], and the publication of the human genome at the beginning of the 21st century [8,9]. Over the past decades, bioinformatics evolved as an interdisciplinary field, incorporating aspects of computing, mathematics, and statistics in order to store, manage, and analyse biological data, making it fundamentally data driven and computational [10]. While the field's evolution has been fast paced, there has been a lack of computing skills among biologists [11,12], and a widening computing skill gap between biologists can be observed [13-17]. It is therefore of increasing importance to understand why biologists are motivated to study bioinformatics and computing.

This paper addresses these concerns through a computing capital research methodology, which has been adapted from already established science capital research [18-20]. Using a sequential mixed methods research design [21], the goal is to understand why people engage with computing education, and what they aspire to do with it. The initial findings of this research suggest that while the biology students are driven by career opportunities, as well as positive beliefs and values when it comes to computing, they perceive barriers such as low levels of computing confidence, and a gap between their skills and future careers.

Our current results suggest that biology curricula should include bioinformatics or computing at an earlier stage, particularly if the goal is to broaden participation of biology students in computing. In addition, career information provided to biology students can help them make better informed decisions about picking up computing skills. Future work can further explore aspects of computing capital, and how they relate to participation in computing education.

II. Related work

Previous research in the United Kingdom introduced the concept of science capital to "help understand why some young people participate in post-16 science and others do not" [20, p.2]. By considering science capital a "holdall", the researchers proposed that science capital includes all the "science-related knowledge, attitudes, experiences, and resources that you acquire through life". In particular, the researchers looked at the following eight dimensions:

1) Scientific literacy
2) Science-related attitudes, values and dispositions
3) Knowledge about the transferability of science (that science 'opens doors' to many careers)
4) Science media consumption
5) Participation in out-of-school science learning contexts
6) Family science skills, knowledge and qualifications
7) Knowing people in science-related roles
8) Talking about science in everyday life

Depending on the score of an individual on each of these dimensions, the researchers argue, the more likely it is that this person is going to participate in science education and careers. However, since science capital focuses on all science-related disciplines, from biology to physics, it misses out on discipline-specific capital. As the researchers suggest, science capital, while "strongly related to engineering and physical science future study aspirations, was not strongly related to the pursuit of either maths or technology postsecondary study" [18, p.1228].

III. Methods

A. Context

This research was conducted as part of a first-year bioinformatics course in Information Management Systems at a Swedish university. As a second semester course, students were already familiar with introductory programming in Python and database design in MySQL. During the course, students were expected to design a laboratory information management system, while being free to choose the project management, communication, and software tools needed to fulfil their goals.

B. Survey instrument development

To make the science capital survey relevant to computing education, the decision was made to adapt the original survey [19]. This meant that the word "science" was changed to "computing" in many cases. However, since the science capital survey was originally intended for young people before their university studies, some of the statements had to be replaced or removed altogether. The goal of the survey design was to keep the core of the original survey, while changing the subject from science to computing. The survey was sent out, along with a project information sheet, to the students before the start of the course. It was also made clear to the students that participation in the survey was voluntary and that data provided in no way affected their course performance. Some personal identifying information was collected (e.g. gender), but results were anonymised during data analysis. A total of 13 students from the bioinformatics programme with a biology background participated in the survey, of which 8 were women.

C. Interview instrument development

Due to low sample size (n=13), the decision was made to further continue the research in a sequential mixed methods design [21, p.65]. This decision was further motivated from a methodological perspective as the researchers wanted to evaluate the usefulness of either research instrument for future studies. The interview design was informed by the initial results of the survey. Additional semi-structured questions were added, relating to background and motivation of the bioinformatics students, and challenges that they faced in their studies and career progression.

Out of the 13 students who filled in the survey, 10 agreed to participate in the follow-up interview. They were sent out an information sheet with more details on the research project, data collection and handling, and right to withdrawal. At the start of the interview, this information was repeated and participants were asked to sign a consent form. Interviews lasted around 30 minutes and took place in person or via video call.

D. Mixed methods data analysis

The survey results were analysed using descriptive statistics, as a low sample size (n=13) and uneven groups (e.g. gender) made it difficult to use inferential statistics. After the results were summarised and interpreted, they were divided into three subthemes. During qualitative data analysis, another subtheme emerged and was added. The themes and subthemes are visualised in Figure 1.

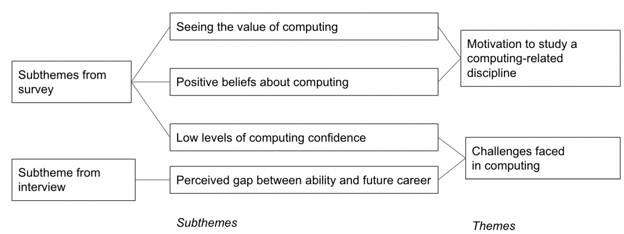

Fig. 1. Subthemes and themes identified

IV. Findings

Through analysis of the survey results and interview transcripts, it became clear that the motivation to study bioinformatics was related to career aspirations and positive beliefs and values when it comes to computing. At the same time, students reported mixed levels of computing knowledge and confidence, and a perceived gap between skills and future careers. These themes, motivation to study a computing-related discipline, and challenges faced in computing, are covered in more detail below.

A. Motivation to study a computing-related discipline

1) Career opportunities and general interest: At the start of the interviews, the students were asked when they decided to study bioinformatics, and what triggered the decision. The reasons provided were varied: a majority indicated an interest in bioinformatics and computing, and about half of the students based their decision on career-related reasons. For instance, three students specifically indicated not wanting to pursue a traditional biology profession: "I have some research experience before, only focusing on a wet lab. And it was extremely boring. And I kind of felt that I don't want to be working in wet labs in my future.. but I also think bioinformatics is really interesting" (int. 8). Similarly, two interviewees mentioned that job prospects after the master's degree were an incentive to study bioinformatics: "In evolutionary biology, you kind of have to continue to research, there's not so much actual industrial work you can do. While in bioinformatics, you have more of an option. You can both do the research or go to industry” (int. 6).

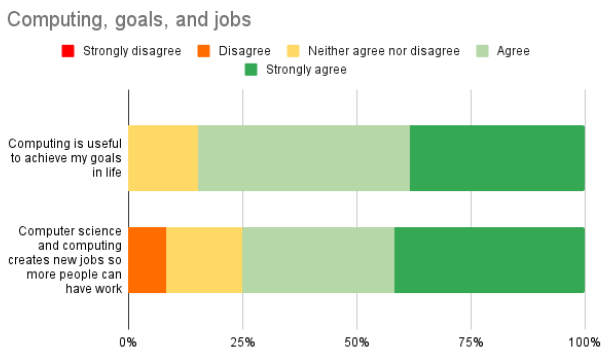

Fig. 2. Survey statements related to computing, goals, and jobs

To relate this back to the survey statements as can be seen in Figure 2, students agreed that computer science and computing creates new jobs, and also that it is useful to achieve goals in life. When asked to elaborate on this during the interviews, a majority of the students mentioned their careers as a goal, and how computing supports that: “I think that computing will be more and more present.. I guess I’ll apply that for my work whenever I get a job as well. Because it will hopefully be bioinformatics” (int. 1).

The students were also asked why they thought that students with a traditional biology background would go into a computing-related discipline or profession, such as bioinformatics or computer science. Again, most of the reasons provided were career related, such as the job market being more appealing, there being more opportunities, a higher salary, and finally, high levels of competition for traditional biology roles: “Because the job market is huge for anything computing related... You get more opportunities, and the pay is very high. It’s not only that, I feel like it’s exciting, because every day, there is new technology emerging” (int. 7).

What these findings seem to suggest is that while biology students are motivated by an interest in bioinformatics and computing, they also perceive the value of it in terms of their future. Some want to move away from traditional biology roles involving a laboratory, whereas others are encouraged by a higher salary and better opportunities.

2) Positive beliefs and values when it comes to computing: The students were also motivated to study a computing-related discipline because of the positive beliefs and values they held when it comes to computing, as can be seen in Figure 3.

During the interviews, when asked if everyone should learn computing, all of the students indicated that they would support this: “I think I would support that. Yeah. But again, I think it’s mostly just good to learn that type of thinking and learning, creative problem solving. So it’s more so the skills you pick up when you learn programming, rather than the programming itself actually being useful” (int. 3). Another student reflected on the way computing teaches her to think

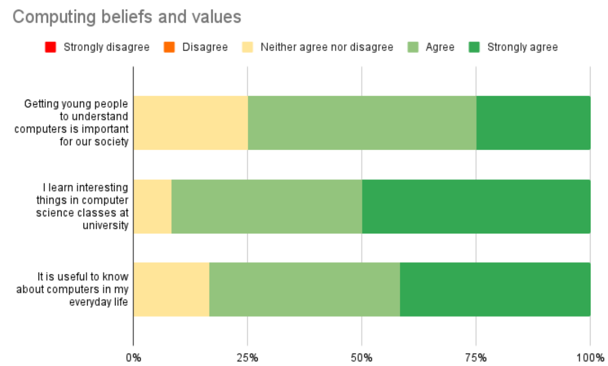

Fig. 3. Survey statements related to computing beliefs and values

in different ways: “It kind of feels that it makes you think in different ways, or thinking about a problem and how to solve it, for example. And it feels really useful to me” (int. 4). And another student mentioned the impact computing has on our lives: “I think that everyone should have a feeling for it. And with the increased impact that computers have on our lives, the increased amount of computers everywhere, the more important it becomes to know how they work” (int. 5). Additionally, most of the students supported the idea that computing courses should be introduced at the bachelor’s level of education of biology.

As can be seen in Figure 3, most of the students strongly agreed or agreed to the statement “it is useful to know about computers in my everyday life”. During the interview, the students were asked to provide an example of this. Two of the students indicated that they use programming for side projects in their free time, two indicated that they play a tech support role for friends and family, and another student mentioned that they were becoming better at finding solutions through Google. The remaining students talked about cybersecurity and awareness, and having better knowledge than others, for instance older generations.

B. Challenges faced in computing

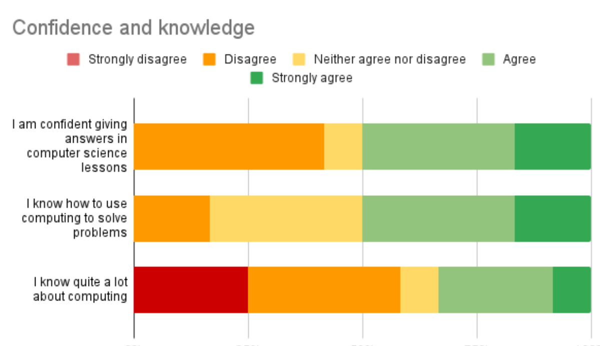

Fig. 4. Survey statements related to computing confidence and knowledge

1) Low levels of computing confidence: Despite being motivated by career opportunities and positive beliefs and values, the students involved in this study reported barriers when it comes to computing. The most evident one is the mixed levels of confidence and knowledge when it comes to computing, as can be seen in Figure 4. During the interviews, the participants were asked to further elaborate on their strengths and weaknesses when it comes to computing, and also their perceived difficulty of computer science courses in their master's programme. All but one student indicated that their weaknesses were related to computing, mathematics, and technical skills: "My weaknesses are definitely the programming, the maths, so the understanding of what is going on behind the scenes" (int. 5). For two students, this meant that they had to spend a disproportionate amount of time on their coursework: "the computing, and [its] challenges also lead to spending a lot of time on solving it" (int. 1).

Since the genders of the participants were known, it was possible to see if there were any group differences between the male and female participants. While the male participants agreed mostly on the statements from the survey, the female participants disagreed for the most part. However, during the interviews, these gender differences were less noticeable. Students were asked to evaluate the difficulty of computing courses during their master's programme, and most of the students reported that the courses were doable. In addition, during the course in which this study took place, the researchers did not notice any differences in computing ability or grades among the students based on gender. It might therefore have been the case that while the male participants initially agreed more to statements related to confidence and knowledge in computing, their performance in the classroom was not different from the female participants.

2) Perceived gap between ability and future career: Students were also asked what their biggest challenges would be when it comes to pursuing bioinformatics or computing-related professions. Three of the students identified high levels of competition for jobs: "most biology labs, they want people with a computer science background... So it is quite hard for me to outcompete those people" (int. 8). Another of the students mentioned that jobs in bioinformatics and computing require a high level of skill, something which she thought not to be trained for.

Despite the students being appealed to the bioinformatics and computing job market, they perceived themselves not to be good enough to compete: "we can write our own little mini script to do some bioinformatics stuff. But I don't think that we'll be able to be a competitive developer" (int. 5). However, at the same time, some of the students were unable to indicate how much computing they needed for their future profession: "I think maybe all the courses we did here kind of scare me off with MATLAB, Python, HTML, and PHP. Maybe I don't need any of that at all" (int. 1). Although the student indicated to have technical ability in four programming languages, she feels that this is not sufficient enough for her future profession. Like other students, there seems to be uncertainty about their own ability, and what they think is required of them after their master's degree.

V. Discussion

Given the lack of computing skills among biologists, and a widening skill gap, it is important to understand the factors underlying their participation in computing. In line with previous studies, which highlight the increasing importance of computing in career trajectories [22-24], and high demand for computing skills in the current job market [25], participants in this study are motivated to study bioinformatics and computing for its potential career benefits.

Unlike previous studies, which report negative stereotypes associated with computing [26, 27], and a "geek culture" surrounding computing [28, p.360], the students in this study held positive beliefs and values when it comes to computing. They exemplified this through providing examples of using computing in everyday life, and all of them supported mandatory computing education, for instance in high school or during their biology education.

At the same time, the students reported levels of discouragement in their studies, by perceiving themselves as not good enough for computing. Related to previous research that there are differences between male and female self-evaluation and self-efficacy in computing [26, 29], this study found that while male students agreed more to computing confidence on the survey, there were not any noticeable differences during the interview or in classroom performance.

VI. Conclusion and further work

The aims of the current study were to evaluate how the science capital survey can be adapted to computing education, and how such research can be used to understand motivation and challenges faced among students. The biology students in this study reported career opportunities, as well as positive beliefs and values when it comes to computing, as a motivation to study a computing-related discipline. At the same time, they were discouraged by low levels of computing confidence, and a perceived gap between skills and future career. From a methodological perspective, the researchers developed a survey instrument and interview guide which can be used for future studies. For instance, by conducting this study over consecutive iterations of the course, it will become possible to use inferential statistics such as t-test between gender groups, or cronbach's alpha for measuring the internal consistency among the survey items. And, while more research is needed, this paper suggests that biology curricula should include bioinformatics or computing at an earlier stage, particularly if the goal is to broaden participation of biology students in computing. In addition, career information provided to biology students can help them make better informed decisions about picking up computing skills.

References

[1] Gauthier, J., Vincent A.T., Charette, S.J., Derome, N. (2019). A brief history of bioinformatics. *Briefing in Bioinformatics, 20*(6), p.1981-1996. DOI: 10.1093/bib/bby063.

[2] Hershey, A.D. Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. *The Journal of General Physiology, 36*(1), p.39-56. DOI: 10.1085/jgp.36.1.39

[3] Ledley, R.S. (1959). Digital electronic computers in biomedical science. *Science, 130*(3384), p.1225-234. DOI: 10.1126/science.130.3384.1225

[4] November, J.A. (2011). Early biomedical computing and the roots of evidence-based medicine. *IEEE Annals of the History of Computing, 33*(2), p.9-23. DOI: 10.1109/MAHC.2011.35.

[5] Dayhoff M.O. Ledley R.S. (1962). *Comprotein: a computer program to aid primary protein structure determination* [Paper presentation]. The December 4-6, 1962, fall joint computer conference. New York, New York, United States. DOI: 10.1109/AFIPS.1962.39

[6] Maxam, A. M., Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America, 74*(2), p.560-564. DOI: 10.1073/pnas.74.2.560

[7] Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M. (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science, 269*(5223), p.496-512. DOI: 10.1126/science.7542800

[8] Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., ... Zhu, X. (2001). The sequence of the human genome. *Science, 291*(5507), p.1304-1351. DOI: 10.1126/science.1058040

[9] Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., ... International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature, 409*(6822), p.860-921. DOI: 10.1038/35057062

[10] Attwood, T.K., Blackford, S., Brazas, M.D., Davies, A., Victoria Schneider, M. (2019). A global perspective on evolving bioinformatics and data science training needs. *Briefings in Bioinformatics, 20*(2), p.398-404. DOI: 10.1093/bib/bbx100

[11] Tan, T.W., Lim, S.J., Khan, A.M., Ranganathan, S. (2009). A proposed minimum skill set for university graduates to meet the informatics needs and challenges of the "-omics" era. *BMC Genomics, 10*. DOI: 10.1186/1471-2164-10-S3-S36

[12] Welch, L., Lewitter, F., Schwartz, R., Brooksbank, C., Radivojac, P., Gaeta, B., Victoria Schneider, M. (2014). Bioinformatics Curriculum Guidelines: Toward a Definition of Core Competencies. *PLoS Computational Biology, 10*(3). DOI: 10.1371/journal.pcbi.1003496

[13] Cummings, M. P., Temple, G. G. (2010). Broader incorporation of bioinformatics in education: opportunities and challenges. *Briefings in bioinformatics, 11*(6), p.537-543. DOI: 10.1093/bib/bbq058

[14] Carvalho, B. S., Rustici, G. (2013). The challenges of delivering bioinformatics training in the analysis of high-throughput data. *Briefings in bioinformatics, 14*(5), p.538-547. DOI: 10.1093/bib/bbt018

[15] Chang, J. Core services: Reward bioinformaticians (2015). *Nature, 520*, p.151-152. DOI: 10.1038/520151a

[16] The Association of the British Pharmaceutical Industry. (2022). *Bridging the skills gap in the biopharmaceutical industry*. ABPI. https://www.abpi.org.uk/media/ya2fjboi/bridging-the-skills-gap-jan-2022.pdf

[17] Brazas, M.D., Blackford, S., Attwood, T.K. (2017). *Nature, 544*, p.161. DOI: 10.1038/544161c

[18] Moote, J., Archer, L., DeWitt, J., Macleod, E. (2020). Science capital or STEM capital? Exploring relationships between science capital and technology, engineering, and maths aspirations and attitudes among young people aged 17/18. *Journal of Research in Science Teaching, 57*(8), p.1228-1249.

[19] Archer, L., Dawson, E., DeWitt, J., Seakins, A., Wong, B. (2015). "Science capital": A conceptual, methodological, and empirical argument for extending bourdieusian notions of capital beyond the arts. *Journal of Research in Science Teaching, 52*(7), p.922-948.

[20] Enterprising Science (2016). *Science Capital Made Clear* [White paper]. https://www.stem.org.uk/sites/default/files/pages/downloads/Science-Capital-Made-Clear.pdf

[21] Creswell, J.W. Plano Clark, V.L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.

[22] Promegger, B., Arshad, D., Krcmar, H. (2021). *Understanding Boundaryless IT Professionals: An Investigation of Personal Characteristics, Career Mobility, and Career Success* [Paper presentation]. Computers and People Research Conference 2021 (SIGMIS-CPR'21), virtual event Germany. DOI:https://doi.org/10.1145/3458026.3462162

[23] Joseph, D., Boh, W.F., Ang, S., Slaughter, S.A. (2012). The career paths less (or more) traveled: A sequence analysis of IT career histories, mobility patterns, and career success. *MIS Quarterly, 36*(2), p.427-452.

[24] Mourmant, G Voutsina, K. (2010). *From IT Employee to IT Entrepreneur: the Concept of IT Entrepreneurial epiphany* [Paper presentation]. International Conference on Information Systems 2010 (ICIS'10). St. Louis, Missouri, United States.

[25] Andreessen, M. (2011, August 20). Why Software Is Eating The World. *The Wall Street Journal*. https://www.wsj.com/articles/SB10001424053111903480904576512250915629460

[26] Kunkeler, T. & Leonard, H.C. (2021). Computing skills, beliefs and identities in young people from underserved communities. *Conference on Research in Equitable and Sustained Participation in Engineering, Computing, and Technology (RESPECT'21)*, p.1-5, DOI: 10.1109/RESPECT51740.2021.9620547.

[27] Wong, B. (2017). 'I'm good, but not that good': digitally-skilled young people's identity in computing. *Computer Science Education, 26*(4), p.299–317.

[28] R. Varma. (2007). Women in computing: The role of geek culture. *Science as culture, 16*(4), p.359–376.

[29] Kallia, M. Sentence, S. (2018). *Are boys more confident than girls? the role of calibration and students' self-efficacy in programming tasks and computer science*. The 13th Workshop in Primary and Secondary Computing Education (WiPSCE'18). Potsdam, Germany.