Received: 20 June 2021 | Accepted: 30 June 2021

DOI: 10.1111/hir.12392

REGULAR FEATURE: DISSERTATIONS INTO PRACTICE

Health Information and Libraries Journal CILIP HLG WILEY

Algorithmic literacy in medical students – results of a knowledge test conducted in Germany

Philipp Kampa MA Subject Librarian[1] | **Felix Balzer MSc, PhD, MD Professor in Medical Data Science**[2]

[1]Universitäts- und Landesbibliothek Sachsen-Anhalt, Martin-Luther-Universität Halle-Wittenberg, Halle (Saale), Germany

[2]Institute of Medical Informatics, Charité – Universitätsmedizin Berlin, Corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany

Correspondence
Philipp Kampa, Martin-Luther-Universität Halle-Wittenberg, Universitäts- und Landesbibliothek Sachsen-Anhalt, Halle (Saale), 06098 Halle (Saale), Germany.
Email: philipp.kampa@bibliothek.uni-halle.de

For details on how to contribute to this Regular Feature please contact:
Frances Johnson
Manchester Metropolitan University
Department of Languages, Information and Communication
Geoffrey Manton Building, Rosamond Street West, Manchester M15 6LL
Tel: +44 (0) 161 247 6157
Email: f.johnson@mmu.ac.uk

Abstract

The impact of algorithms on everyday life is ever increasing. Medicine and public health are not excluded from this development – algorithms in medicine do not only challenge, change and inform research (methods) but also clinical situations. Given this development, questions arise concerning the competency level of prospective physicians, thus medical students, on algorithm related topics. This paper, based on a master's thesis in library and information science written at Humboldt-Universität zu Berlin, gives an insight into this topic by presenting and analysing the results of a knowledge test conducted among medical students in Germany.

KEYWORDS

education, medical; eHealth; information literacy; research, quantitative; students

INTRODUCTION

Algorithms play a continuously growing role in medicine and health care. The spectrum of algorithm based or algorithm related application scenarios ranges from image recognition of tumours, over genome analysis (Esteva et al., 2019) to the surveillance of epidemics (Radin et al., 2020), or even pandemics (Bragazzi et al., 2020). Additionally, as Liu et al. (2019) have shown, algorithms already perform with the same quality physicians

[The copyright line for this article was changed on 25 November 2021 after original online publication]
[Correction added on 8 December 2021, after first online publication: Appendix 1 and 2 have been replaced and added as supporting information in this version.]

do. Against this background, algorithms prove to be more than an emerging technology, but a *fact* doctors face. Therefore, literally the question arises, what do prospective physicians actually know about algorithms?

LITERATURE REVIEW

Cormen et al. (2009) define algorithms in a basic IT sense as sequences producing an output out of a given input. In this sense, algorithms have been a technology not only in IT or information science, but also in medicine for a very long time, probably from its beginning as a discipline. If a scheme is applied to a case, as protocols and 'reminders' (Komaroff, 1982), algorithms can be understood as helping to solve defined problems. However, this is not the prior sense of algorithms that concerns this study, rather it circles around machine driven algorithms, thus algorithms that are much more complex than 'simple', albeit helpful instructions or manuals to follow. Accordingly, we are concerned with algorithmic decision making and associated topics such as Big Data.

Debate on defining data literacy or digital literacy, that the competency *algorithmic literacy* belongs to, is not new (Koltay, 2015; Ridsdale et al., 2015; Schield, 2004). Nevertheless, defining and understanding algorithmic literacy, and respectively its parent category *data literacy*, tend to circle around the soft skills helping data literates when they try to solve a problem – skills such as self-management and data management (Calzada Prado & Marzal, 2013; Haendel et al., 2012). Such approaches thereby underestimate the role of knowledge and give little attention to testing people's knowledge about the subjects that may be seen as an integral part of algorithmic literacy.

Empirical findings on data literacy or algorithmic literacy are scarce. Whereas statistical literacy of medical students has been researched (Jenny et al., 2018), studies in data literacy or algorithmic literacy of medical students tend to be policy papers, evaluations of best practice courses or attempts to theoretically framework data literacy by identifying corner points of the concept – rather than studies that penetrate into surveying existing, and not only perceived, but also actual knowledge and skills (Calzada Prado & Marzal, 2013; Kuhn, 2018; Kuhn et al., 2018; Poncette et al., 2020; Schüller & Busch, 2019). 'Real' testing (with all its inherent challenges) therefore is a desideratum.

Studies in Germany and in Europe have shown that nearly half of the population was completely at a loss when asked to say something about algorithms (Fischer & Petersen, 2018; Grzymek & Puntschuh, 2019). A recent study among European medical students suggests similar (Machleid et al., 2020) in which personal skills in eHealth were perceived deficient by a majority of those who took part in the survey.

METHODS

Ethical approval of this study was given by the ethics committee of the Charité – Universitätsmedizin Berlin (EA1/066/20). Subjects participated voluntarily and anonymously in the study.

We developed a multiple choice knowledge test, consisting of nine questions, with each four answer items. The test covered a broad spectrum of algorithm associated topics, ranging from statistical questions (e.g. causation vs. correlation), over the estimation of data quality, to ethical matters and pitfalls (cf. Appendices 1–2). Questions consisted of a vignette usually presenting a case or a clinical situation, and four answer options, one being the correct one. Chance performance thus was 25%. The questions were designed in a closed and standardized manner with respect to medical assessment in Germany, and this was to minimize hurdles and to motivate subjects, by structurally imitating real testing situation (IMPP, 2020). Each knowledge question was accompanied by a three point scaled self-estimation scheme wherein subjects could declare how sure they were when answering the related knowledge question. Stages ranged from 'I don't know' over 'I suppose' to 'I'm sure'. Knowledge questions were formulated out of a selective literature research identifying major topics and current use cases such as algorithm controlled insulin pumps (Bergenstal et al., 2013).

We added three statements to the knowledge test: Subjects were asked to position themselves in the field of algorithms and medicine in a broader sense, including a statement on the status quo of digital health teaching in medical school. The statements were integrated into the survey to better understand the attitudes and the stance of the participants towards algorithms in medicine and health care. The collection of this data enabled us to check possible connections between habits and the results in the test, too. Here, subjects responded on a five point Likert scale ranging from 1 (Strongly agree) to 5 (Strongly disagree). A 'No idea' option was attached to avoid a situation in which persons with no specific attitude would be forced to, more or less coincidently, choose one category and therefore distort the results (Menold & Bogner, 2015). The first statement reflected attitudes towards medical education in terms of eHealth, respectively, digital health in Germany, the second towards the estimated objectivity of algorithmic decision making and the last one towards possible conflicts between the result of algorithmic processes and human diagnosis. The survey ended up with the query of the semester in which subjects studied – this to better classify the sample (Appendices 1–2).

Algorithmic literacy was operationalised by determining core fields of the topic *algorithms in medicine*. We tested algorithmic literacy in its cognitive manner, largely excluding its pragmatic dimension, although being covered in passing through a question system that was situation and action based, and therefore partly problem solving.

Data collection

The survey was conducted online, over Lime Survey (version 4.1.10+200311) hosted at Humboldt Universität zu Berlin, among medical students in Berlin and Halle, Germany. Two identical surveys, one for the Berlin subjects and one for the Halle ones, were unlocked for two weeks, in May and in June 2020. We invited subjects via e-mail and via a digital learning platform, partly including a reminder; communication channels were adapted to different organizational preconditions in Berlin and Halle. All subjects participated in the study without any sort of recompensation. Nor was any incentive launched. Initially, we aimed at surveying the 10th semester in paper in the scope of courses linked to the content. However, due to the SARS-CoV2 pandemic, the survey was conducted online, independently from courses. Therefore, the target group was widened to medical students from all semesters within the two medical schools to reach more potential subjects. Invitations were made via mailing lists and information platforms, a restriction to a certain semester therefore would not have made any practical sense.

Data analysis

We analysed the data with IBM® SPSS® Statistics (version 25) [computer software], Armonk, NY, USA: International Business Machines Corp. Above all, data were analysed descriptively by calculating frequencies, and comparing them, especially to contrast correct answer rates with the self-estimations, and respectively the confidence levels. Figures were created with Microsoft Excel (version 14.0.7265.5000 (32 bit)) [computer software], Redmond, WA, USA: Microsoft Corp.

RESULTS AND DISCUSSION

Initially, 225 persons took part in the survey; 98 of them filled in the questionnaire completely. Concerning complete questionnaires, 60 subjects were Berlin medical students; the rest, thus 38 persons, were students from Halle. Students from nearly all semesters including practical year ('Praktisches Jahr', PJ) took part in the study; the majority of the subjects were from advanced semesters – which is probably linked to the case based framework of the questions – with a clear emphasis on the 10th semester. Besides, it may be related to the acquisition of subjects: conducting of the research was affected and adjusted to the pandemic and mailing in Halle was focused on the 10th semester.

The median of correctly given answers over all questions was 64.3% (IQR = 53.6–82.4, $n = 98$). The two questions with the lowest rate of correct answers were two and five, the first covering basic algorithmic procedures on the example of search engine algorithms and the second focussing on the opacity of algorithms including related ethical pitfalls. Question two was answered correctly by 40.8% of the participants, question five by 42% ($n = 98$). Questions with the highest correct answer rate were – not surprising because of its low difficulty level as an opener – question one and question six (Table 1).

Question six is a striking part of the study due to another aspect: a fairly high correct answer rate in the study is often subdued by the self-estimation of the subjects queried in parallel. Here, almost every subject gave the correct answer,

TABLE 1 Overview of correct answer rates and the linked self-estimation, each in % ($n = 98$), by question

Question	Topic (noted retrospectively, not in the test)	Correct answer rate in % ($n = 98$)	Self-estimation in % ($n = 98$)		
			'I don't know'	'I suppose'	'I'm sure'
Question 1	[Secondary data]	87.8	6.1	65.3	28.6
Question 2	[Processing of a query]	40.8	38.8	55.1	6.1
Question 3	[Training data]	64.3	14.3	61.2	24.5
Question 4a	[Risk and algorithms 1]	65.3	13.3	48.0	38.8
Question 4b	[Risk and algorithms 2]	57.1	15.3	48.0	39.8
Question 5	[Limits of algorithms]	42.0	42.9	45.9	11.2
Question 6	[Types of algorithms]	95.9	19.4	51.0	29.6
Question 7	[Algorithmic concepts]	63.0	56.1	32.7	11.2
Question 8	[Logical relations]	80.6	17.3	60.2	22.4
Question 9	[Algorithmic performance]	59.2	48.0	45.0	5.0

but 19.4% of the subjects chose 'I don't know'; 51% only supposed to have picked out the correct item. Discrepancies between the correct answer rate and the confidence level are striking concerning 3 other questions of the survey (seven, eight and nine). 40% of the questions thus lead to discrepancies; 20% of the questions were answered on average false; taking this together, the overall performance rate, which was relatively high, has to be questioned and classified in a slightly different way. Discrepancies indicate knowledge gaps, as well as low confidence levels, although subjects are likely to be interested in algorithms; otherwise, they would not have taken part in the study (Table 1).

Analysis of the first statement (Figure 1) showed that the majority of subjects do not feel sufficiently prepared for the digital challenges in health and medicine by their studies. Given the statement, 43.9% of the participants chose 4 on a five point Likert scale (5 = strongly disagree); 24.5% marked 5. The results coincide with former research into this topic (Machleid et al., 2020; Offergeld et al., 2019) showing again that medical students perceive teaching of digital skills in medical studies as marginal and non-sufficient.

Response to the second statement (Figure 2) reflected indifference, for 30.6% of the subjects were undecided, even 10.2% of them picked out the 'No idea' category. Apart from this, a slight tendency towards a pro-attitude concerning algorithm related objectivity can be determined – 10.2% of the subjects (n = 98) strongly agreed to the statement ('Algorithms as decision support make medicine more objective'.).

Finally, response to the third statement (Figure 3) was split: 43.9% of the subjects chose 1 or 2; that means they

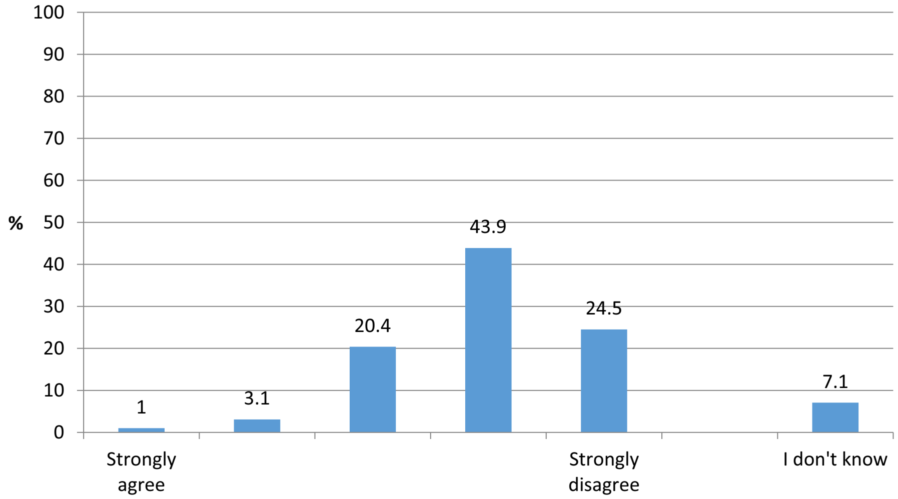

FIGURE 1 Statement 1 (Throughout my studies in medicine, I feel prepared for the increasing digitisation of health care in a sufficient manner.), answer rates in % (n = 98) [Colour figure can be viewed at wileyonlinelibrary.com]

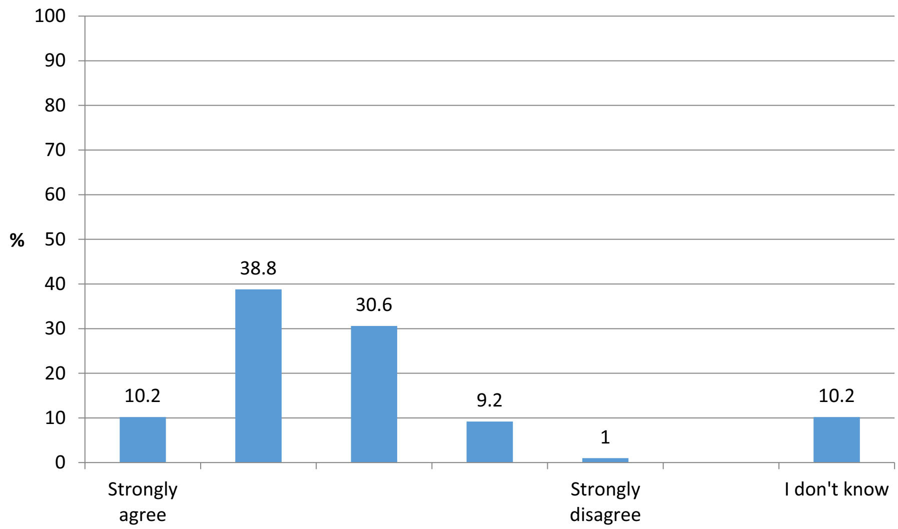

FIGURE 2 Statement 2 (Algorithms as decision support make medicine more objective.), answer rates in % (n = 98) [Colour figure can be viewed at wileyonlinelibrary.com]

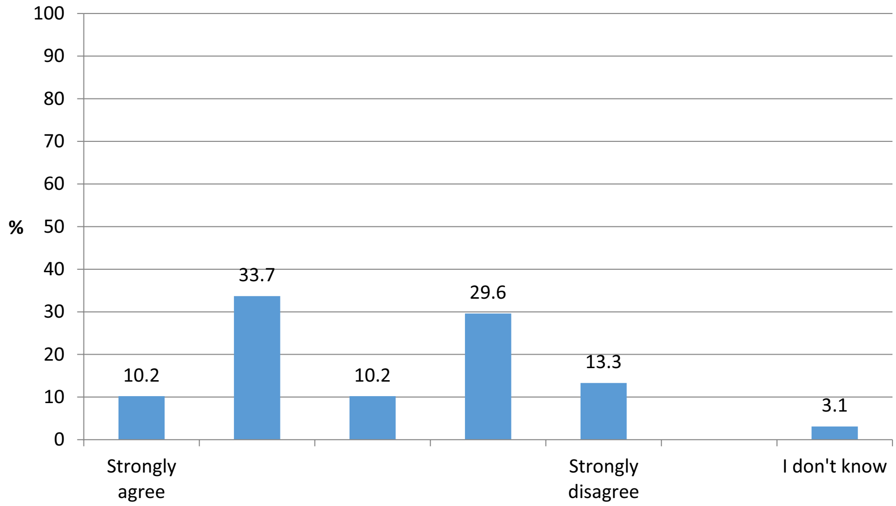

FIGURE 3 Statement 3 (The routinely derived ECG of a patient, who is cardiac symptom-free, seems to be inconspicuous. However, the automatic ECG-analysis indicates that a myocardial infarction is suspected. In this situation, I would register further laboratory diagnostics to rule out ischemia.), answer rates in % (n = 98) [Colour figure can be viewed at wileyonlinelibrary.com]

would arrange further diagnostics although being sure on their decision differing from the algorithm's one. 42.9% chose 4 or 5 (the scale ranged from 'Strongly agree' [1] to 'Strongly Disagree' [5]). A sub-analysis of the 10th semester (n = 28) – the population initially serving as target – in comparison with the 1st and 2nd semester (the two are covered together because there was only one participant in the 1st semester) suggests a connection between study progress and trust in one's own medical judgement. Median within the 10th semester was 4 (5 is 'Strongly disagree', n = 28), median with the 1st and 2nd semester, regarded, as novices, was 2 (n = 10), and two of the ten subjects chose the 'I don't know' category.

Furthermore, we compared results in Berlin with Halle. Although students in Berlin take part in a so-called Modellstudiengang, thus a curriculum that essentially includes praxis based learning formats, medians of the correct answer rates were nearly identical: in Berlin at 67.8% (IQR: 54.2–82.7), in Halle at 64.6% (IQR: 51.3–82.1).

Limitations

The sample was quite small due to a high, but not unusual dropout rate (El-Menouar & Blasius, 2005). Besides, survey participants came only from two selected medical schools in Germany. The similarities in the results were nevertheless striking. Distortions that might have been caused by different strategies of subject acquisition in the two study places are therefore cushioned. A weakness of the study lies in the relative vagueness of the sample concerning demographic data. Questions were formulated case and scenario based; advanced students were more likely to understand more of the medical background and thus having an advantage over novice students. Limitations also include the strong cognitive dimension of the survey: as every knowledge test, the test applied here, queries external competencies such as reading skills and the ability to concentrate in a way that might distort the results. Another limitation lies in the design of the study as a convenience sample based on self-selection of subjects: it is very likely that subjects interested in and familiar with this field, and therefore having prior knowledge, participated in the survey rather than people without any affirmative relation to this field. Taking this into consideration, knowledge gaps among medical students are probably much bigger, in reality. Results are also limited by the fact that the knowledge test was newly developed; no proven instrument has been used.

IMPLICATIONS FOR PRACTICE AND CONCLUSION

This was the first study in algorithmic literacy as a subfield of data literacy in medical students. Striking discrepancies between correct answer rates and self-estimation imply the existence of knowledge gaps.

We determined search engine algorithms as well as structural components of machine learning algorithms as knowledge gaps. Furthermore, subjects were uncertain in their knowledge when it came to even simple distinctions between different algorithm types. Special curricular instruction on those topics could help in creating a more solid knowledge.

Algorithmic literacy as a part of data literacy is a key concept in medical education. Our findings suggest that knowledge gaps exist basically throughout all related fields. Although the median of the correct answer given was 64.3% (IQR = 53.6–82.4, $n = 98$), deep discrepancies between relatively high correct answer rates and strong signs of guessing could be determined. Respectively, low confidence levels among subjects, concerning a wide range of algorithm related topics can be supposed. The study provides an empirical basis for more focussed teaching interventions in digital health. An implementation of fixed and basic curricular offers concerning digital health, that go beyond sporadic approaches, might improve student's knowledge and confidence when it comes to algorithms in medicine and health care.

ACKNOWLEDGEMENTS

The authors would like to thank all participants who made this study possible. Philipp Kampa would like to thank Felix Balzer, the academic supervisor, as well as Michael Seadle, the academic co-supervisor of this study. Furthermore, he would like to thank Niklas Keller, Malte Schmieding, Jochen Kruppa, the Fachschaftsinitiative Medizin der Charité – Universitätsmedizin Berlin, the Fachschaftsrat Medizin Halle, the offices of student affairs in Berlin and Halle and Rafael Mikolajczik for (organizational) advice and support. He then would like to thank Universitäts- und Landesbibliothek Sachsen-Anhalt, Halle (Saale), for supporting him. Open Access funding enabled and organized by ProjektDEAL.

ORCID

Felix Balzer https://orcid.org/0000-0003-1575-2056

REFERENCES

Bergenstal, R. M., Klonoff, D. C., Garg, S. K., Bode, B. W., Meredith, M., Slover, R. H., Ahmann, A. J., Welsh, J. B., Lee, S. W., & Kaufman, F. R. (2013). Threshold-based insulin-pump interruption for reduction of hypoglycemia. *The New England Journal of Medicine*, *369*(3), 224–232. https://doi.org/10.1056/NEJMoa1303576

Bragazzi, N. L., Dai, H., Damiani, G., Behzadifar, M., Martini, M., & Wu, J. (2020). How big data and artificial intelligence can help better manage the COVID-19 pandemic. *International Journal of Environmental Research and Public Health*, *17*(9), 3176. https://doi.org/10.3390/ijerph17093176

Calzada Prado, J., & Marzal, M. Á. (2013). Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri*, *63*(2), https://doi.org/10.1515/libri-2013-0010

Cormen, T. H., Leiserson, C. E., Rivest, R. L. & Stein, C. (2009). *Introduction to algorithms* (3rd edn). MIT Press.

El-Menouar, Y., & Blasius, J. (2005). Abbrüche bei Online-Befragungen: Ergebnisse einer Befragung von Medizinern. *ZA-Information / Zentralarchiv Für Empirische Sozialforschung*, *56*, 70–92. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-198541

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, *25*(1), 24–29. https://doi.org/10.1038/s41591-018-0316-z

Fischer, S. & Petersen, T. (2018). Was Deutschland über Algorithmen weiß und denkt: Ergebnisse einer repräsentativen Bevölkerungsumfrage. Retrieved from https://doi.org/10.11586/2018022

Grzymek, V., & Puntschuh, M. (2019). Was Europa über Algorithmen weiß und denkt: Ergebnisse einer repräsentativen Bevölkerungsumfrage. Retrieved from https://doi.org/10.11586/2019006.

Haendel, M. A., Vasilevsky, N. A., & Wirz, J. A. (2012). Dealing with data: A case study on information and data management literacy. *PLOS Biology*, *10*(5), e1001339. https://doi.org/10.1371/journal.pbio.1001339

IMPP (2020). Praktische Hinweise zur Durchführung der schriftlichen Prüfungen nach der Approbationsordnung für Ärzte. Mainz. Retrieved from https://www.impp.de/pruefungen/allgemein/praktische-hinweise.html?file=files/PDF/Praktische%20Hinweise/Medizin_58Auflage.pdf.

Jenny, M. A., Keller, N., & Gigerenzer, G. (2018). Assessing minimal medical statistical literacy using the Quick Risk Test: A prospective observational study in Germany. *British Medical Journal Open*, *8*(8), e020847. https://doi.org/10.1136/bmjopen-2017-020847

Koltay, T. (2015). Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science*, *49*(1), 3–14. https://doi.org/10.1177/0961000615616450

Komaroff, A. L. (1982). Algorithms and the "art' of medicine. *American Journal of Public Health*, *72*(1), 10–12. https://doi.org/10.2105/ajph.72.1.10

Kuhn, S. (2018). Transformation durch Bildung: Medizin im digitalen Zeitalter. *Deutsches Ärzteblatt*, *115*(14), A633–A638.

Kuhn, S., Kadioglu, D., Deutsch, K., & Michl, S. (2018). Data Literacy in der Medizin. *Der Onkologe*, *24*(5), 368–377. https://doi.org/10.1007/s00761-018-0344-9

Liu, X., Faes, L., Kale, A. U., Wagner, S. K., Fu, D. J., Bruynseels, A., Mahendiran, T., Moraes, G., Shamdas, M., Kern, C., Ledsam, J. R., Schmid, M. K., Balaskas, K., Topol, E. J., Bachmann, L. M., Keane, P. A., & Denniston, A. K. (2019). A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis. *The Lancet Digital Health*, *1*(6), e271–e297. https://doi.org/10.1016/S2589-7500(19)30123-2

Machleid, F., Kaczmarczyk, R., Johann, D., Balčiūnas, J., Atienza-Carbonell, B., von Maltzahn, F., & Mosch, L. (2020). Digital health in medical education: Findings from a mixed-methods survey among european medical students. *Journal of Medical Internet Research*. https://doi.org/10.2196/19827

Menold, N., & Bogner, K. (2015). Gestaltung von Ratingskalen in Fragebögen. https://doi.org/10.15465/gesis-sg_015

Offergeld, C., Neudert, M., Emerich, M., Schmidt, T., Kuhn, S., & Giesler, M. (2019). Vermittlung digitaler Kompetenzen in der curricularen HNO-Lehre: Abwartende Haltung oder vorauseilender Gehorsam? *HNO*. Advance online publication. https://doi.org/10.1007/s00106-019-00745-8

Poncette, A.-S., Glauert, D. L., Mosch, L., Braune, K., Balzer, F., & Back, D. A. (2020). Undergraduate medical competencies in digital health and curricular module development: Mixed methods study. *Journal of Medical Internet Research*, *22*(10), e22161. https://doi.org/10.2196/22161

Radin, J. M., Wineinger, N. E., Topol, E. J., & Steinhubl, S. R. (2020). Harnessing wearable device data to improve state-level real-time surveillance of influenza-like illness in the USA: A population-based study. *The Lancet Digital Health*, *2*(2), e85–e93. https://doi.org/10.1016/S2589-7500(19)30222-5

Ridsdale, C., Rothwell, J., Smit, M., Bliemel, M., Irvine, D., Kelley, D., & Ali-Hassan, H. (2015). Strategies and best practices for data literacy education knowledge synthesis report. https://doi.org/10.13140/RG.2.1.1922.5044

Schield, M. (2004). Information literacy, statistical literacy and data literacy. IASSIST Quarterly, Summer/Fall, 6–11. Retrieved from https://iassistdata.org/sites/default/files/iqvol282_3shields.pdf

Schüller, K., & Busch, P. (2019). Data Literacy: Ein Systematic Review. https://doi.org/10.5281/zenodo.3484583

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Kampa, P., & Balzer, F. (2021). Algorithmic literacy in medical students – results of a knowledge test conducted in Germany. *Health Information & Libraries Journal*, 38, 224–230. https://doi.org/10.1111/hir.12392