

Development of Statistical Literacy and Scientific Reasoning & Argumentation in Physicians

Felicitas Maria Schmidt (✉ Felicitasmariaschmidt@gmail.com)

Ludwig-Maximilians-Universität München Medizinische Fakultät <https://orcid.org/0000-0001-8284-5549>

Jan Zottmann

Ludwig-Maximilians-Universität München Medizinische Fakultät

Maximilian Sailer

Universität Passau

Martin R Fischer

Ludwig-Maximilians-Universität München Medizinische Fakultät

Markus Berndt

Ludwig-Maximilians-Universität München Medizinische Fakultät

Research article

Keywords: evidence-based practice, scientific reasoning and argumentation, statistical literacy, medical education, postgraduate medical education

Posted Date: April 8th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-20026/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Statistical literacy (SL) of physicians, i.e. the ability to understand and apply statistics, is an essential prerequisite for risk estimation and communication. Together with scientific reasoning and argumentation (SRA) skills, SL provides the basis for evidence-based practice. Several studies suggest that in medical students both skills are developed merely on a medium level.

Methods: The aim of the present study was to investigate these skills in practicing physicians (N = 71, M Age = 40.00, SD = 9.59) and when and how these skills were acquired. Biographical data was collected with an online survey tool. SL was assessed with multiple-choice items. SRA skills evidence evaluation (EE) and drawing conclusions (DC) were measured with a decision scenario.

Results: Study results indicated that physicians have medium to high levels of SL (M = 17.58, SD = 6.92, max 30 pts.) and SRA (EE-Score: M = 7.75, SD = 1.85, max 10 pts.; DC-Score: M = 37.20, SD = 5.35, max 60 pts.). Skills development via autodidactic learning activities (M = 4.78, SD = 1.125, range 1-6) was reported significantly more often than development during medical education (M = 2.31, SD = 1.456, t (71) = -9.915, p < .001) or in extracurricular activities (M = 3.34, SD = 1.869, t (71) = 4.673, p = .000). The active involvement in research seemed decisive: The number of publications and time spent in research predicted SL to a large extent ($\beta = .355$, p = .002; $\beta = .280$, p = .018). SRA skills were associated with the type of doctoral thesis ($\beta = -.380 \pm .154$, p = .016) and working in research ($\beta = 3.355 \pm 1.229$, p = .008).

Conclusion : The development of SL and SRA skills needs to be systematically fostered during medical education combined with active involvement in research.

1. Background

In a patient-physician relationship on equal footing, transparent information about risks is key. Statistical literacy (SL) of the physician is a prerequisite for effective risk communication (1, 2). Statistical literacy is not only the ability to understand statistical information, but also to apply it in decision-making (3). It comprises the aptitude of critical reflection about statistics as evidence in arguments (4) and is intertwined with scientific reasoning and argumentation skills (SRA), to provide the basis for evidence-based decision-making (5, 6).

However, a collective statistical illiteracy has been observed among physicians (3, 7) and SRA skills needed for evidence-based practice (8), such as evidence evaluation (EE) or drawing conclusions (DC), are underdeveloped (3, 9). The aim of the present study was thus to gain insights into the status quo of SL and SRA skills in medical doctors (physicians), and to identify influencing factors that should be considered in focused educational training to foster these skills.

Statistical literacy and Scientific reasoning & argumentation

Statistical literacy is based on numeracy, the aptitude of mathematical operations (10), and encompasses the use and interpretation of statistics and its results (11, 12) and (Walker, 1951) the ability to explain and critically evaluate them (13, 14). In a medical context, it ranges from relative risks and conditional probabilities (15, 16), over test constructs like sensitivity and specificity (3), to complex inferential statistical procedures applied in empirical research studies (17-19).

Several existing instruments for the assessment of SL have been employed in a variety of studies. Domain-general assessment tools, such as the Berlin Numeracy Test (20), or the Three Item Test (7) are focusing mainly on basic numeracy and are considered less reliable in highly educated samples, such as physicians, due to ceiling effects (20, 21). Such effects have not been observed in domain-specific tests, such as the Obstetrician-Gynecologist Statistical Literacy Questionnaire (OGSLQ) (9) which aims at assessing clinically relevant numerical facts, concepts, and relations in physicians, or the Medical Data Interpretation Test assessing the aptitude to weigh treatment alternatives based on evidence (22).

In a study by Anderson et al. (2014), 52% of the participating physicians answered two (or fewer) of four questions regarding statistical concepts correctly (9). This is in line with findings by Windish and colleagues (2007) reporting only 40% of resident physicians demonstrating adequate understanding of biostatistical concepts (23). Similar results were found by Gigerenzer and Wegwarth (2008), showing that 33% of gynecologists were not aware of the benefits of mammography screening, with 79% being unable to interpret the positive predictive value (PPV) (24). Gigerenzer and colleagues (2008) summarized various studies on the concept of PPV and its dependence on the prevalence, finding that 50% of participants were under the impression that false positive test results in HIV testing do not exist. They also found that only 2 of 20 urologists have sufficient knowledge about the reliability of a PSA-test. Overall, physicians' SL is not below-average (9). However, it can be considered comparable to other educated samples (7, 25) and was found superior to that of residents in research training (23) or medical students (26). The few studies that have assessed medical students' SL are supporting the findings of superiority of physicians (27). The authors (2020) compared medical students to those of social sciences and economics and found that medical students in their first years of study scored higher in comparison to social science students and comparable to economics students (28). This study not only assessed SL, but also the two skills: evidence evaluation (EE) and drawing conclusions (DC), as part of the conceptual framework for scientific reasoning and argumentation (SRA) by Fischer and colleagues (2014). Based on this framework, SRA can be defined as the competence of comprehending and applying scientific working methods and their results when solving problems (29, 30). The intertwining of SRA and SL was assumed by Anderson et al. (2013) stating that the latter is needed to evaluate scientific evidence (31) and by Franklin et al. (2005), who stated that SL itself encompasses also SRA skills, like formulating questions and answering them on an informed basis (32).

In a medical context, possible links between SL and SRA have been examined with the Medical Data Interpretation Test (22). Participating physicians scored higher overall than others with postgraduate degrees (89 out of 100 score points). Johnson et al. (2014) assessed numeracy of medical students and residents and found students with poor numeracy being more likely to misjudge risks of different

treatment alternatives and that the confidence in treatment recommendation increased during medical school (26). In the study by the authors (2020), the skills did not improve over the course of undergraduate studies (28).

In summary, SL and SRA seem to be underdeveloped in medical students and better developed in physicians. There are indicators that the development of SL and SRA does not necessarily happen within formal education and the question remains how, where, and when this development occurs. This study should contribute to the discussion when and how to best foster SL and SRA skills in lifelong learning of physicians. It aimed at providing further insights into these skills in physicians and to identify demographic factors and learning opportunities that may be associated with the development of these skills. Our guiding research questions were:

1. SL and SRA

1. To what extent are SL and SRA skills developed in physicians?
2. To what extent does SL predict SRA skills of physicians?

2. Education and Development:

1. How, where, and when do physicians develop SL and SRA skills?
2. Which demographic factors are related to the development of SL and SRA?

2. Method

2.1. Design and sample

Our study followed a quasi-experimental, causal-comparative design with two dependent variables: SL and SRA. We included N = 71 German and Austrian physicians (31 female, 34 male) from different work settings and locations in our study: hospital (N = 43), outpatient sector (N = 2), research (N = 16). A doctoral thesis was completed by 58 and 9 participants were currently working on it. The mean age of participants was 40 years (SD = 9.59, range = 26–65) (Table 1).

2.2. Test instrument / Measures

For the assessment of SL and SRA skills, we used a tool previously developed in the context of a study with medical students (28). The instrument combines multiple choice items to assess SL with a decision scenario to assess the participants' skills in EE and DC (SRA skills) (33). We added items on relevant demographic factors.

2.2.1. Demography

Demographic parameters of the participants were assessed with a special interest in their working history and environment (hospital, out-patient care, research). Questions were adapted from a study by Epstein

and colleagues (34) and comprised multiple choice items, some with the opportunity to fill in additional free text; Five items on the doctoral thesis (qualifications fostered meanwhile), three items on the professional career, two items on the publication record (type of authorship, number of publications), and three items on the current job description.

2.2.2. Statistical Literacy

Statistical literacy was measured with multiple choice items based on validated instruments (9, 20, 22), to assess a broad spectrum from basic numeracy to conditional probabilities and statistical concepts. Reliability of the SL test was .82 (Cronbach's α) in our sample with a maximum score of 30 points. All items were framed in a medical context; however, no medical content knowledge was necessary to answer them correctly.

2.2.3. SRA skills

The assessment of SRA skills focused on the two epistemic activities EE and DC being assessed with a decision-making scenario in a medical context (general medicine, out-patient care) and provided a separate overall score for EE and DC (maximum score 10 and 60, respectively, Cronbach's α for EE items .87; for DC items .74). For EE, participants had to evaluate four pieces of evidence in terms of their scientificness (35–37) including one authentic pharmaceutical brochure. Then, the participants rated the persuasiveness (Likert 1–6) of 20 arguments, which were extracted from the presented evidence and assigned a level of argument strength from 1 (lowest) to 4 (highest). The participants' ratings for scientificness were compared to the ratings of scientificness by the authors, resulting in a measure of similarity for EE and DC. The range of these measures was from 0–10 (EE-Score) and 0–60 (DC score) with zero indicating no similarity.

2.3. Procedure

The study was completed by the participants either online with LamaPoll (<https://www.lamapoll.de/>; accessed March 22, 2020), a survey tool optimized for mobile applications, or with pen and paper (return rate 16.5% online and 66.7% pen and paper). Average duration was 45 minutes. Participants were invited via mailing lists and personal contacts.

2.4. Statistical Procedures and Analyses

Statistical analysis was performed with IBM SPSS 25. Descriptive and frequency data were computed for primary analysis and Cronbach's α for internal reliability. Extensive outlier analyses were conducted and all required prerequisites for statistical analyses, such as normal distribution and homoscedasticity of variables, were tested. One-factorial ANOVAs and linear regression models were calculated to assess the association of demographic factors on SL and SRA. P values less than .05 were considered significant. Data in natural verbal language underwent thematic analysis to extract common themes.

3. Results

We included 71 complete cases (see Table 1). For 13 participants, we imputed missing values in the evaluation of the 20 arguments from the average of the respective item. The entire data set was checked for univariate outliers, skewness and kurtosis for all variables was within the ± 2 range (38). The prerequisites for ANOVA were fulfilled, all variables were normally distributed and homoscedastic.

3.1. Statistical literacy and SRA

The 71 physicians' average score in SL was $M = 17.58$, $SD \pm 6.92$, with a range of 5 to 30 out of 30 attainable points. On average, physicians evaluated the evidences concordantly with the authors' evaluation, EE-Score $M = 7.75$, $SD = 1.85$. The ratings for argument quality were in accordance with the authors' rating, DC-Score $M = 37.20$, $SD = 5.35$. Statistical literacy and DC were significantly inversely correlated (DC-Score $r(71) = -.272$, $p = .022$). However, no correlation was found between SL and EE, $r(71) = .198$, $p = .098$, nor EE and DC, $r(71) = .138$, $p = .256$.

3.2. Education & Development

We explored how, where, and when physicians developed scientific skills. Significantly more participants indicated to have acquired scientific skills in an autodidactic manner ($M = 4.78$, $SD = 1.125$, Likert 1–6 scale) rather than during their studies ($M = 2.31$, $SD = 1.456$, Likert 1–6 scale) ($t(71) = -9.915$, $p < .001$) or in extracurricular activities ($M = 3.34$, $SD = 1.869$, Likert 1–6 scale) ($t(71) = 4.673$, $p < .001$) (Fig. 1). In a free-text box, the participants added various other learning opportunities, such as massive open online courses, higher education, workshops, and learning through peer reviews and feedback (Fig. 2).

Regarding the doctoral thesis, a univariate ANOVA showed that having completed or working on a doctoral thesis had an effect on the scoring in EE, $F(3,71) = 10.494$, $p < .001$, partial $\eta^2 = .320$, but not on DC, $F(3,71) = 1.133$, $p = .342$, nor SL, $F(3,71) = 1.812$, $p = .153$, partial $\eta^2 = .075$. The fostering of critical scrutinization of study results presented by other researchers during the preparation of the doctoral thesis was positively correlated with the scoring in SL, $r(71) = .271$, $p = .033$.

Regarding the postgraduate phase, a one-factorial ANOVA showed a significant main effect of having worked in research on SL, $F(1,70) = 12.737$, $p = .001$, partial $\eta^2 = .156$. The type of authorship, $F(5,71) = 3.886$, $p = .004$, partial $\eta^2 = .230$, and the time spent in research, $F(23,71) = 2.262$, $p = .009$, partial $\eta^2 = .525$, were significantly associated with better SL, as well as the number of publications, $r(71) = .36$, $p = .002$.

Regarding SRA, linear regression models revealed that the corresponding score in EE increased by $\beta = .314 \pm .150$, $p = .041$, when the Likert value of the doctoral thesis supervisor's content-related support was increased by one point. Additionally, the form of the doctoral thesis (experimental, clinical, empirical, statistical, or literature review) was associated with EE, with experimental and clinical design being positively related to EE skills, $\beta = -.380 \pm .154$, $p = .016$, $R^2 = .187$, $F(1,59) = 4.353$, $p = .041$. DC was higher when participants indicated to have already worked in research, $\beta = 3.355 \pm 1.229$, $p = .008$, $R^2 = .314$, $F(1,68) = 7.448$, $p = .008$.

4. Discussion

4.1. Research question 1

We found average SL of physicians (59%), a rather high-level EE score (77%) and a medium-level DC score (62%). SL did not predict the SRA skills of physicians' which could indicate that these are independent concepts.

Due to the focus on SL rather than the combination of basic numeracy and SL(20), the test discriminated well and we did not find any ceiling effects as observed in other educated samples (7, 21). A comparison to other studies assessing SL of physicians is not easily done as every test focuses on different aspects of SL. Schmidt et al. (2017) focused on knowledge of 18 different statistical tests among pathologists and observed a rather low level of SL (39). Anderson et al. (2014) did not create an overall score but distinguished between fact, concept, and relation questions and found altering levels of SL (9). A study with Greek residents also concentrated on knowledge questions and reported a rather low SL (40). A similar tool was applied with German medical students (28). Comparing the two studies, the EE and DC score of medical students and physicians were almost on the same level. Riegelman and Hoveland (2012) found that residents struggled when critical reflection upon research was required (41), whereas the participants in our study showed medium to high levels of SRA skills.

SRA and SL scores were not correlated, and DC and SL correlated inversely. A possible explanation would be that the evidences provided in the scenario did not contain any statistical information, as suggested by Anderson et al. (2014). Nevertheless, in theoretical frameworks, SL has been regarded as prerequisite for SRA (42). Moreover, in a Dutch community-based study, more numerate participants showed enhanced performance in SRA due to increased evaluation of pros and cons in decision making and evaluation of judgments (43). Future research could incorporate numerical evidence in the assessment of SRA in order to further analyze this connection.

4.2. Research question 2

We explored how, where, and when physicians developed SL and SRA skills. They indicated to have acquired scientific skills mostly in an autodidactic manner, in higher education outside of medical studies or in extracurricular activities.

Better SL was associated with the fostering of critical thinking during the doctoral thesis, having worked in research, time spent in research, number of publications, and the type of authorship. Our findings are in line with Schmidt et al. (2017), who found that having an advanced degree other than MD or statistic courses were positively associated with SL. A study with physicians, residents, and final year medical students in Thailand showed – not surprisingly – that having recently completed a statistical workshop led to higher SL scores (44). However, additional courses are often hard to integrate in medical training. A study showed that 37% of American Ob-Gyn residents do not receive formal training (31), a study with neurology residents observed a lack of acceptance for interventions on SL (45).

In the present study, better EE was associated with having worked on a doctoral thesis (with favorably experimental or clinical design) and having content-related support by the supervisor. These findings are in line with the subjective impression of German medical graduates with a doctoral thesis ranking their scientific skills higher compared to those working on it (34). However, the participants in the study by Epstein and colleagues (2018) did not feel confident enough to conduct research on their own (34). This is particularly important because having already worked in research was associated with a higher SL and DC score in the present study and in Schmidt et al. (2017). Moreover, Epstein and colleagues (2018) found, that medical graduates self-estimate their scientific skills after medical school as rather low (34). In the United States, only 68.1% of medical students in their final year participated in research during medical school and only 42% had (co-)authored a paper submitted for publication (46). It seems important that medical students should be involved in the publishing process and learn how to write papers already during the completion of the doctoral thesis, as this might enhance their SL and SRA skills in the long run.

4.3. Strengths and Limitations

This study built upon an innovative approach by the authors (2020) to assess SL and SRA skills in physicians (28). As this participant group is not easily recruited, the sample size of N = 71 was satisfactory. The recruited sample may not have been representative of all physicians. Furthermore, sociocultural factors and specifics of the medical education system may limit generalizability of results. The addition of numerous demographic variables yielded insights on how, where and when scientific skills were acquired and helped to identify potential associated factors. Although first attempts were made to validate the assessment tool with medical students, it could be considered too extensive and time consuming for physicians (as was indicated by 11 participants in their feedback).

5. Conclusion/ Outlook

We assessed SL and SRA skills of German and Austrian physicians together with a thorough analysis of the predictive value of demographic variables. The active involvement in research seems to play an important role to foster these skills and might in consequence enhance adequate risk communication and evidence-based practice. As most participants indicated to have acquired these skills post-graduate and in an autodidactic manner, we argue to formalize and intensify the acquisition of these skills during undergraduate medical studies. Medical education curricula should include more and more effective statistical training and the active involvement of students in research.

Abbreviations

ANOVA	Analysis of Variance
DC	Drawing conclusions
DV	Dependent variable

EE	Evidence evaluation
EL	Evidence level
IV	Independent variable
LV	Likert value
MME	Master of Medical Education
MOOCS	Massive Open Online Courses
PPV	Positive Predictive Value
SR	Scientific reasoning
SRA	Scientific reasoning and argumentation

6. Declarations

Ethics approval and consent to participate

The study was approved with student subjects by the ethics committee of the Medical Faculty of LMU Munich (approval reference no. 527–16). The additional clearance to conduct the study with physicians was given by the Bavarian State Chamber of Physicians. The consent being obtained from study participants was written.

Consent for publication

Not applicable

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing Interests

The authors declare that they have no competing interests.

Funding

This work was supported by an intramural grant of the Medical Faculty of LMU Munich (WiFoMed). WiFoMed is a sponsorship association of LMU Munich's medical faculty promoting excellent research of early stage researchers. It provided financial means for participant acquisition, survey assistants, students support staff, and printing costs for the paper-based survey. The German Federal Ministry of

Education and Research (grant no. 01PB14004C) financed project members (one author full time, one author part time) until the 31st of March 2018, as well as costs for conference presentations and project meetings. However, none of these funding agencies provided any kind of content-related directive that could have influenced the research and its outcomes.

Authors' contributions

FS was the principal author in writing the manuscript and analyzed and interpreted the data. MB was a major contributor in writing the manuscript and supervising the data analyses and interpretations. All authors collectively designed the measurement instrument. In particular, JZ, MS and MB designed the statistical literacy test, while FS and MF were principally involved in writing the SRA decision scenario. All authors read and approved the final manuscript.

Acknowledgement

The authors would like to thank Wolfgang Gaissmaier for his valuable comments and suggestions to the present study.

Authors' information

Felicitas M. Schmidt MPH, attended medical school at the Ludwig-Maximilian University Munich. This paper constitutes a central part of her doctoral thesis "Statistical literacy as basis for scientific reasoning and argumentation in medicine". During her undergraduate degree in Philosophy, Politics and Economics (PPE) from the Free University of Bolzano and the HEC Lausanne, she developed a profound understanding of statistics and its value for the scientific community and everyday life. She holds a Master's degree in Public Health and is a scholar of the Friedrich Naumann-Foundation for Freedom.

Jan M. Zottmann holds a master's degree and a doctoral degree in Educational Sciences (LMU Munich, Germany). He worked at the Knowledge Media Research Center (Tübingen, Germany), the University of Tübingen, and LMU Munich before he became Assistant Director of the Institute for Medical Education at the University Hospital of LMU Munich in 2016. Dr. Zottmann is also a member of the Munich Center of the Learning Sciences at LMU. His areas of academic interest include clinical reasoning, case-based learning, instructional support for simulation-based learning, and computer-supported collaborative learning.

Prof. Dr. phil. Maximilian Sailer is a University Professor of Educational Sciences at the University of Passau, Germany. He studied Educational Sciences at the University of Eichstätt and earned his doctoral degree in 2002. After receiving his post-doctoral degree (Habilitation) in 2009, he worked as a senior researcher & Adjunct Professor at the University of Augsburg. He has been involved in teaching and research at LMU Munich since 2011, where served five years as an interim Professor for various positions. He is also a member of the Munich Center of the Learning Sciences. His areas of academic

interest include Technology-Enhanced Learning (TEL), Blended Learning and Empirical Educational Research.

Professor Dr. Martin R. Fischer is a board-certified internist and endocrinologist and a medical educator. He graduated from the Albert-Ludwigs-University in Freiburg, Germany and earned his doctoral degree in neuropharmacology in Freiburg. He did his residency at Medizinische Klinik-Innenstadt in internal medicine and endocrinology and earned a Master's Degree in Medical Education at the University of Berne in Switzerland. From 2008 until 2011 he was a Professor for Health Sciences Education at Witten-Herdecke University. Since 2011 until today he is the Director of the Institute for Medical Education at the University Hospital and Medical Faculty of LMU Munich. Since 2012 he also is Dean for Clinical Education at LMU.

Dr. Markus Berndt earned his PhD in Educational Psychology (Ludwig-Maximilians-Universität of Munich, Germany), a Master's degree in Educational Sciences with a major in adult education (University of Regensburg, Germany), and a Master's degree in Human Resource Management (Murray State University, USA). He has been involved in teaching and research at LMU Munich since 2010. Research foci are on clinical reasoning, scientific reasoning and argumentation, peer assessment and peer feedback, assessment and evaluation, and the use of eye tracking technology to investigate cognitive processes and interpretation of radiographic images and electrocardiograms. Currently, Dr. Berndt holds a position as research fellow at the University Hospital of LMU Munich and a position as Senior Contributing Faculty at Walden University.

7. References

1. Reyna VF, Nelson WL, Han PK, Dieckmann NF. How numeracy influences risk comprehension and medical decision making. *Psychol Bull.* 2009;135(6):943-73.
2. Nelson W, Reyna VF, Fagerlin A, Lipkus I, Peters E. Clinical implications of numeracy: theory and practice. *Ann Behav Med.* 2008;35(3):261-74.
3. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping Doctors and Patients Make Sense of Health Statistics *Psychological Science In The Public Interest.* 2008;8(2):53-96.
4. Shield M. Statistical literacy: thinking critically about statistics. *Of Significance.* 1999;1(1):15-20.
5. Sedlmeier P, Gigerenzer G. Teaching Bayesian Reasoning in Less Than Two Hours. *Journal of Experimental Psychology: General.* 2001;130(3):380-400.
6. Fischer F, Kollar I, Ufer S, Sodian B, Hussmann H, Pekrun R, et al. Scientific Reasoning and Argumentation: Advancing an Interdisciplinary Research Agenda in Education. *Frontline Learning Research.* 2014;2(3):28-45.
7. Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. *Med Decis Making.* 2001;21(1):37-44.
8. Monahan J. *Statistical literacy: A prerequisite for evidence-based medicine.* SAGE Publications Sage CA: Los Angeles, CA; 2007.

9. Anderson BL, Gigerenzer G, Parker S, Schulkin J. Statistical literacy in obstetricians and gynecologists. *J Healthc Qual.* 2014;36(1):5-17.
10. Peters E. Beyond Comprehension: The Role of Numeracy in Judgments and Decisions. *Current Directions in Psychological Science.* 2012;21(1):31-5.
11. Walker HM. Statistical literacy on the social science. *The American Statistician.* 1951;5(1):6-12.
12. Wallman KK. Enhancing statistical literacy: enriching our society. *Journal of the American Statistical Association.* 1993;88(421):1-8.
13. Ben-Zvi D, Garfield J. Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. . In: Ben-Zvi D, Garfield J, editors. *The challenge of developing statistical literacy, reasoning and thinking.* Dordrecht, The Netherlands: Kluwer; 2004. p. 3-25.
14. Watson JM. Assessing Statistical Thinking Using the Media. In: Gal I, Garfield JB, editors. *The Assessment Challenge in Statistics Education.* Amsterdam: IOS Press and The International Statistical Institute 1997. p. 107-21.
15. Covey J. A meta-analysis of the effects of presenting treatment benefits in different formats. *Med Decis Making.* 2007;27(5):638-54.
16. Naylor CD, Chen E, Strauss B. Measured enthusiasm: Does the method of reporting trial results alter perceptions of therapeutic effectiveness? *Annals of Internal Medicine.* 1992;117(916-921).
17. Sotos AEC, Vanhoof S, Van den Noortgate W, Onghena P. Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational research review.* 2007;2(2):98-113.
18. Gal I. Expanding conceptions of statistical literacy: An analysis of products from statistics agencies. *Statistics Education Research Journal.* 2003;2(1):3-21.
19. Wyatt CM, Fallar R. Medical Students Learn to Apply Bayesian Reasoning for Clinical Inference but Not Statistical Inference: A Mixed-Methods Study.
20. Cokely ET, Galesic M, Schulz E, Ghazal S, Garcia-Retamero R. Measuring Risk Literacy: The Berlin Numeracy Test. *Judgment and Decision Making.* 2012;7(1):25-47.
21. Hanoch Y, Miron-Shatz T, Cole H, Himmelstein M, Federman AD. Choice, numeracy and physician-in-training performance: The case of Medicare part D. *Health Psychology.* 2010;29:454-9.
22. Schwartz LM, Woloshin S, Welch HG. Can patients interpret health information? An assessment of the medical data interpretation test. *Med Decis Making.* 2005;25(3):290-300.
23. Windish DM, Huot SJ, Green ML. Medicine residents' understanding of the biostatistics and results in the medical literature. *JAMA.* 2007;298(9):1010-22.
24. Gigerenzer G, Wegwarth O. Risikoabschätzung in der Medizin am Beispiel der Krebsfrüherkennung. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen.* 2008;102(9):513-9.
25. Okamoto M, Kyutoku Y, Sawada M, Clowney L, Watanabe E, Dan I, et al. Health numeracy in Japan: measures of basic numeracy account for framing bias in a highly numerate population. *BMC Med Inform Decis Mak.* 2012;12:104.

26. Johnson TV, Abbasi A, Schoenberg ED, Kellum R, Speake LD, Spiker C, et al. Numeracy among trainees: are we preparing physicians for evidence-based medicine? *Journal of surgical education*. 2014;71(2):211-5.
27. Msaouel P, Kappos T, Tasoulis A, Apostolopoulos AP, Lekkas I, Tripodaki E-S, et al. Assessment of cognitive biases and biostatistics knowledge of medical residents: a multicenter, cross-sectional questionnaire study. *Medical Education Online*. 2014;19:23646-.
28. Berndt M, Schmidt FM, Sailer M, Fischer F, Fischer MR, Zottmann JM. Investigating Statistical Literacy and Scientific Reasoning & Argumentation in Medical-, Social Sciences-, and Economics Students. Submitted 2020.
29. Hetmanek A, Engelmann K, Opitz A, Fischer F. Beyond intelligence and domain knowledge: Scientific reasoning and argumentation as a set of cross-domain skills. In: Fischer F, Chinn CA, Engelmann K, Osborne J, editors. *Scientific reasoning and argumentation: The roles of domain-specific and domain-general knowledge*. New York: Routledge; 2018. p. 203-26.
30. Rudolph JL, Horibe S. What do we mean by science education for civic engagement? *Journal of Research in Science Teaching*. 2016;53(6):805-20.
31. Anderson BL, Williams S, Schulkin J. Statistical literacy of obstetrics-gynecology residents. *Journal of graduate medical education*. 2013;5(2):272-5.
32. Franklin C, Kader G, Mewborn D, Moreno J, Peck, R. , Perry M, et al. *Guidelines for Assessment and Instruction in Statistics Education (GAISE) report: A Pre-K-12 curriculum framework*. Alexandria: VA: American Statistical Association; 2005.
33. Trempler K, Hetmanek A, Wecker C, Kiesewetter J, Wermelt M, Fischer F, Fischer M, Gräsel C. Nutzung von Evidenz im Bildungsbereich – Validierung eines Instruments zur Erfassung von Kompetenzen der Informationsauswahl und der Bewertung von Studien. *Z Pädagogik*. 2015;61:144-66.
34. Epstein N, Huber J, Gartmeier M, Berberat P, Reimer M, Fischer M. Investigating on the acquisition of scientific competencies during medical studies and the medical doctoral thesis. *Journal of Medical Education* 2018;35(2).
35. Simon D. Chronische Herzinsuffizienz Apotheken Umschau 2016 [Available from: <http://www.apotheken.de/gesundheit-heute-news/article/chronische-herzinsuffizienz/>
36. *Ärzteblatt*. Neue Empfehlungen zum Einsatz von Weißdornpräparaten bei Herzinsuffizienz 2017 [Available from: <https://www.aerzteblatt.de/fachgebiete/kardiologie/news?nid=72259>].
37. Meissner T. Weißdorn: Effekt auf das Endothel im Fokus. *ÄrzteZeitung* 2017 15. März 2017.
38. Tabachnick BG, Fidell LS. *Using multivariate statistics*. 4th ed. ed. Needham Heights, MA: Allyn & Bacon; 2001.
39. Schmidt RL, Chute DJ, Colbert-Getz JM, Firpo-Betancourt A, James DS, Karp JK, et al. Statistical Literacy Among Academic Pathologists: A Survey Study to Gauge Knowledge of Frequently Used Statistical Tests Among Trainees and Faculty. *Arch Pathol Lab Med*. 2017;141(2):279-87.
40. Msaouel P, Kappos T, Tasoulis A, Apostolopoulos AP, Lekkas I, Tripodaki ES, et al. Assessment of cognitive biases and biostatistics knowledge of medical residents: a multicenter, cross-sectional

questionnaire study. Medical Education Online. 2014;19:10.

41. Riegelman RK, Hovland K. Scientific Thinking and Integrative Reasoning Skills (STIRS): Essential outcomes for medical education and for liberal education. Peer Review. 2012;14(4):10.
42. Watson JM, Callingham R. Statistical literacy: A complex hierarchical construct. Statistics Education Research Journal. 2003;2(2):3-46.
43. Ghazal S, Cokely ET, Garcia-Retamero R. Predicting biases in very highly educated samples: Numeracy and metacognition. Judgment and decision making. 2014.
44. Laopaiboon M, Lumbiganon P, Walter SD. Doctors' statistical literacy: a survey at Srinagarind Hospital, Khon Kaen University. Journal of the Medical Association of Thailand (Chotmai het thangphaet). 1997;80(2):130-7.
45. Leira E, Granner M, Torner J, Callison R, Adams H. Education research: the challenge of incorporating formal research methodology training in a neurology residency. Neurology. 2008;70(20):e79-e84.
46. Karani R, Ognibene FP, Fallar R, Gliatto P. Medical students' experiences with authorship in biomedical research: a national survey. Academic Medicine. 2013;88(3):364-8.

Tables

Table 1: Description of study group

Variable	Options	Frequency	Percentage
Gender	Male	34	47.9%
	female	31	43.7%
	No answer	6	8.4%
German mother tongue	Yes	68	95.8%
	No	3	4.2%
	NA	5	7.0%
MD in Germany	Yes	68	95.8%
	No	3	4.2%
Thesis	Yes	58	81.7%
	Currently working on	9	12.7%
	No thesis or skipped	4	5.6%
Academic qualification	Habilitation	12	16.9%
	Professorship	7	9.9%
Ever worked as researcher	Yes	36	50.7%
	No	35	49.3%
Working Environment	Hospital	2	2.8%
	Out-patient care	43	60.6%
	Research	16	22.5%
	NA	10	14.1%

Figures

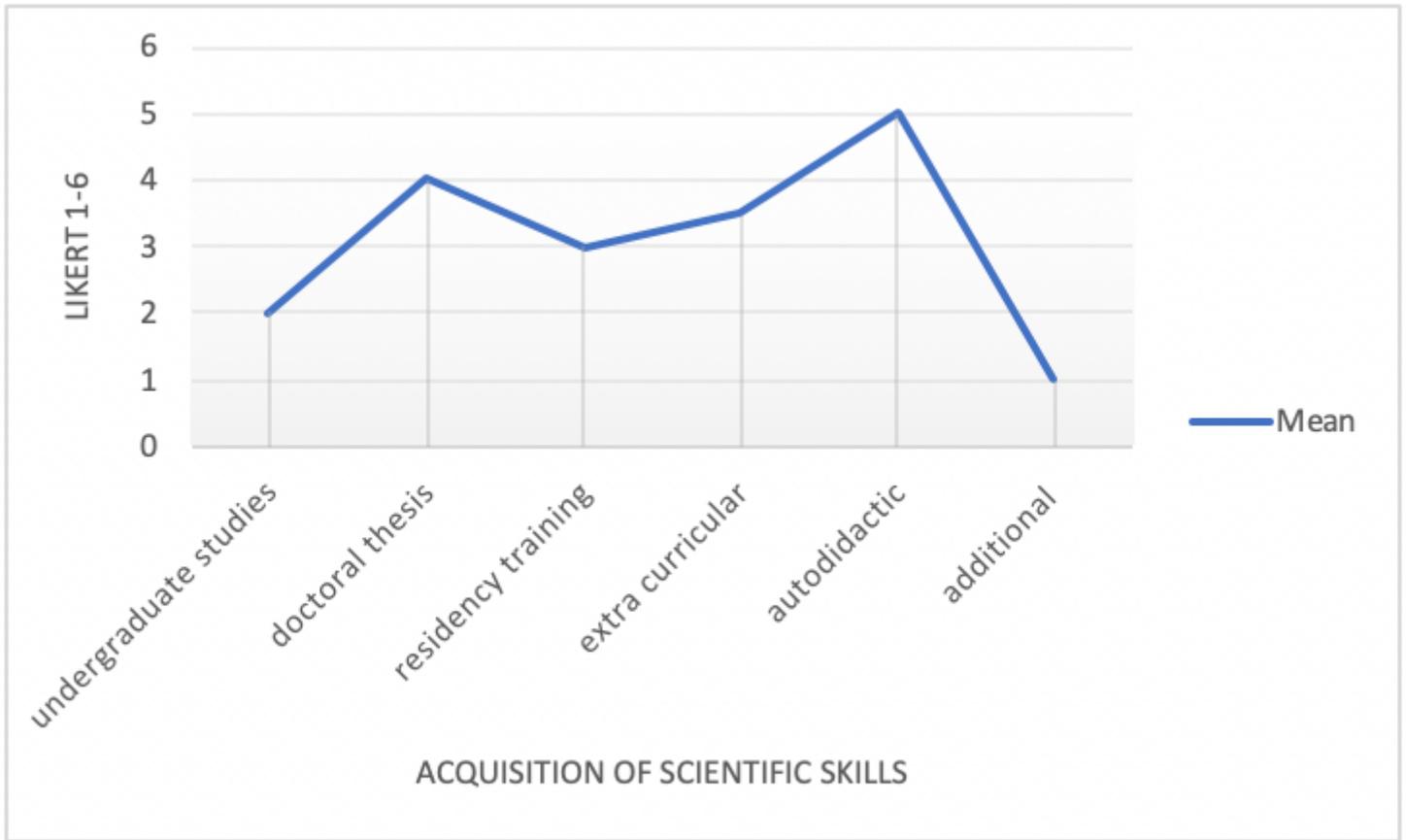


Figure 1

In which context have you acquired scientific skills?

Additional (Free Text Analysis)

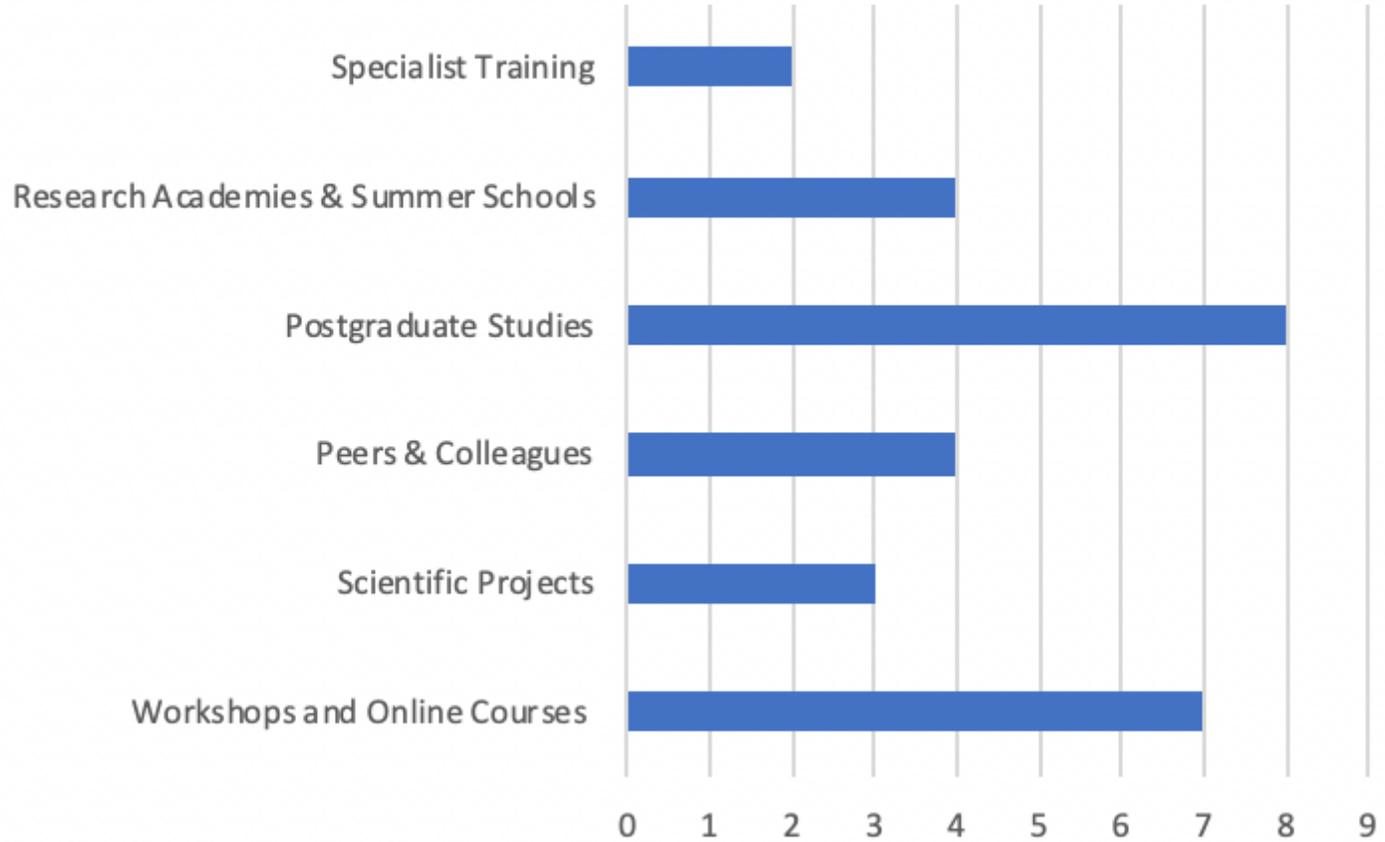


Figure 2

Where have you acquired scientific skills (free text analysis)?