LAWSON CLASSROOM TEST OF SCIENTIFIC REASONING AT ENTRANCE UNIVERSITY LEVEL

Tereza Hrouzková, Lukáš Richterek
Palacký University Olomouc, Czech Republic
E-mail: tereza.hrouzkova01@upol.cz, lukas.richterek@upol.cz

Abstract

The Lawson classroom test of scientific reasoning is a quite popular and widely used tool that measures the level and development of the student's scientific reasoning skills. In this contribution, the results of this test for the N=446 students of the Faculty of Science Palacký University Olomouc from the years 2018–2020 at the beginning of their study were analysed. Calculation of the standard characteristics of the test items (difficulty index, discrimination index and point biserial coefficient) confirms that the test meets the required criteria for the included questions and its overall consistency. Performing the Mann-Whitney rank test with a standard significance level α = .05, the statistically significant difference between males and females was found, on the other hand, no statistically significant difference between the students preparing for their prospective teacher careers and those with technical specializations in chemistry or physics was identified. This contradicts a common belief that students choosing teacher preparation programmes are generally worse in scientific performance and technical subjects. The results also show that about a quoter of the students come to university with a low level of their scientific reasoning skills which should be therefore supported and further developed in lecturing and teaching.

Keywords: *scientific reasoning, test evaluation, Lawson Classroom Test of Scientific Reasoning*

Introduction

In science, technology, engineering, and mathematics education (STEM) there has been increased emphasis not only on the learning of content knowledge but also on the development of skills connected with the skills important for practical science, including scientific reasoning. Generally, scientific literacy is currently considered one of the central goals for the development of 21st-century citizens and scientific reasoning ability is determined as an important factor for fostering student performance in science learning and has gained increasing attention from science educators and researchers. Moreover, scientific reasoning represents a hidden variable that can influence students' understanding of learning new concepts. Therefore, mapping the level of scientific reasoning skills within a class or a course may be useful to reveal students who have more difficulty in learning and may provide the information to develop an appropriate method which overcomes the pedagogical issues.

In the literature, there are many definitions of scientific reasoning. Following (Bao et al., 2009a) we accept the position that scientific reasoning represents the cognitive skills necessary to understand and evaluate scientific information, which often involves understanding and evaluating theoretical, statistical, and causal hypotheses.

https://doi.org/10.33225/BalticSTE/2021.74

Broadly defined, it includes the thinking and reasoning skills involved in the inquiry, experimentation, evidence evaluation, inference, and argumentation that support the formation and modification of concepts and theories about the natural and social world.

In 1978, Lawson designed an assessment instrument that measures students' level of scientific reasoning development (Lawson 1978). The paper and pencil style addressed the need for a reliable, convenient tool enabling to examine larger numbers of respondents that would be more practical for classroom use, compared to the Piagetian tasks. In 2000, building on previous work, Lawson developed an improved version of the assessment instrument, named Lawson's Classroom Test of Scientific Reasoning (LCTSR). It is a two-tier, multiple-choice test with 24 items (Lawson, 2000). A two-tier multiple-choice item pairs (questions 1–22) contain a question with some possible answer choices, followed by another question proving some possible reasons for the response to the previous question. According to (Zhou et al., 2021), a correct answer with incorrect reasoning indicates an intermediate level of reasoning development, whereas an incorrect answer with correct reasoning is likely a result of guessing. All the answer choices were designed based on previous studies on student misconceptions with free response tests, interviews, and relevant literature (for a more detailed recent review of the LCTSR development see e.g. (Zhou et al., 2021)). After more than 20 years we may conclude that the LCTSR is a practical tool assessing a unidimensional scale of scientific reasoning with good overall reliability, though inspections of individual question pairs revealed some validity concerns for several question pairs (Bao et al., 2018). Nevertheless, correlations were calculated between LCTSR scores and other measures of reasoning, and adequate values found. The Lawson test gained popularity among STEM educators and researchers, so far it has been given to several thousands of students and results were published in tens of papers, some of which are also referred to in this text. It is appropriate for a wide spectrum of contexts, namely intro colleges (which is also our case), high schools and middle schools.

In this research, the LCTSR, which assesses students' abilities in six dimensions including conservation of matter and volume, proportional reasoning, control of variables, probability reasoning, correlation reasoning, and hypothetical-deductive reasoning, was also used. All of the above-mentioned skills were identified as necessary for a successful STEM career.

The LCTSR scores at the beginning of the course may tell us a lot about our students' initial reasoning levels so that we get a better sense of which types of reasoning our students need more help with and can adjust appropriate teaching methods. According to the scores divided into 3 equal intervals, Lawson suggested classifying students into three formal reasoning categories: concrete operational (scoring up to 8 points, i.e., 33%), transitional (scoring between 9 and 16 points, i.e., 34% and 67%), and formal operational (scoring above 16 points, i.e., 67%). For the comparison of the score results, it is important to mention some other similar studies. For instance, (Bao et al., 2009a) report an average score of 75% based on the $N = 5760$ freshmen science and engineering students in four U.S. and three Chinese universities of medium ranking. On the contrary, (Sriyansyah & Saepuzaman, 2017) found among 29 first-year prospective physics teachers who attended introductory physics courses in 2015 at Indonesia University of Education no formal reasoner and 31% concrete reasoners.

A strong positive statistically significant correlation between students' pre-instruction scores for reasoning ability (measured by the LCTSR) and some tools measuring the conceptual knowledge in a particular field of physics has been reported in high school and university introductory courses, e.g., learning of forces and Newtonian laws (measured by the Force Concept Inventory) and basic concepts of electricity and magnetism (measured by Brief Electricity and Magnetism Assessment); see (Nieminen et al., 2012; Pyper, 2011; Sriyansyah & Saepuzaman, 2017). For the original Lawson 1978 test administered to secondary school students in Israel and the U.S.A. there was only a small correlation between achievement in science and maths and the Lawson test; and the Israeli population achieved significantly higher than the US students on the Piagetian skills measured by the test (Hofstein & Mandler, 1985). Also, (Maloney, 1981) reported statistically significant differences between the science-major and non-science major students (in the original Lawson 1978 test), with all six students with the lowest scores (concrete operational level) dropping the intro-physics course without completing exams. Similarly, (Moore, & Rubbo, 2012) have found that non-STEM (science, technology, engineering, and mathematics) majors taking either a conceptual physics or astronomy course at two regional comprehensive institutions scored significantly lower on the LCTSR in comparison to national average STEM majors (the majority of their non-STEM students can be classified as either concrete operational or transitional reasoners, whereas in their STEM population formal operational reasoners are far more prevalent). Scores on the LCTSR were then correlated with normalized learning gains on various concept inventories. The correlation was strongest for content that can be categorized as mostly theoretical, meaning a lack of directly observable exemplars, and weakest for content categorized as mostly descriptive, where directly observable exemplars are abundant. These results further demonstrate that differences in student populations are important when comparing normalized gains on concept inventories, and the achievement of significant gains in scientific reasoning requires a re-evaluation of the traditional approach to physics for non-STEM students.

Another question is, to which extent the scientific reasoning skills are related to students' performance and are supported within university introductory courses. For instance, (Bao et al., 2009a, Bao et al., 2009b) found that the rigorous learning of physics knowledge in middle and high schools makes a significant impact on the ability of students in China to solve physics problems, but this knowledge does not seem to have a direct effect on their general scientific reasoning ability. This suggests that current education and assessment in the STEM disciplines often emphasize factual recall over a deep understanding of scientific reasoning. Similarly, based on the results from two Chinese universities (Ding, 2013) shows that, regardless of their major, student reasoning skills measured by the LCTSR remained largely constant across the four years of higher education in accordance with (Pyper, 2011) and his data from several years and several different classes that have shown that Lawson test scores do not change much over the course of a single semester. Students of average and lower reasoning abilities struggle in solving problems that depend on conceptual understanding and may depend more readily on memorization of simple procedures to solve problems (Fabby & Koenig 2015). As pointed by (Piraksa, Srisawasdi & Koul, 2014), there is a critical area for improvement of students' scientific reasoning ability. In biology (Susilowati & Anam, 2017) successfully verified the 5E-learning model (Engagement, Exploration, Explanation, Elaboration, Evaluation) to be effective in scientific reasoning and problem-solving ability of students.

This all indicates, that LCTSR can provide valuable information and feedback for the course instruction, at least in some context. Therefore, it makes sense to ask, whether the results can be affected by some parameters, like gender or study programmes of the respondents.

Research Methodology

The LCTSR was tested with the students of the Faculty of Science, Palacký University Olomouc within the years 2018–2020 with the total number of $N = 446$ respondents (158 in 2018, 133 in 2019 and 155 in 2021) including chemistry-major and physics-major students (technical programmes, called non-teachers further in the text) and students in prospective science teachers' programmes (called teachers below). Let us remark that in the Czech Republic future teachers are typically specialized in two-subject combinations like e.g., maths-physics or chemistry-biology. The test was typically taken within introductory courses in the first week of the semester, thus the scores were not affected by any university-level instruction and rather correspond to the level of scientific reasoning at the end of their high schools. To motivate the students to think about questions seriously and not just randomly mark the answers, we gave them some extra points for the overall scores of the courses within which we have organized the testing. The organizing and scoring the test also followed the recommendations and instruction provided for LCTSR by PhysPort network support developed and maintained by the American Association of Physics Teachers (https://www.physport.org/assessments/).

After the scoring, the standard characteristics for test items and the whole test were determined as described below within the results sections. The standard characteristics of the test items and the test as a whole were determined according to (Ding et al., 2013; Eaton et al., 2019) where the corresponding formulas and recommended ranges are introduced. The item difficulty index P is a measure of the difficulty of a single test question. It is calculated by taking the ratio of the number $N1$ of correct responses on the question to the total number N of students who attempted the question ($P = N1/N$). The recommended values are in the interval <.3, .9>.

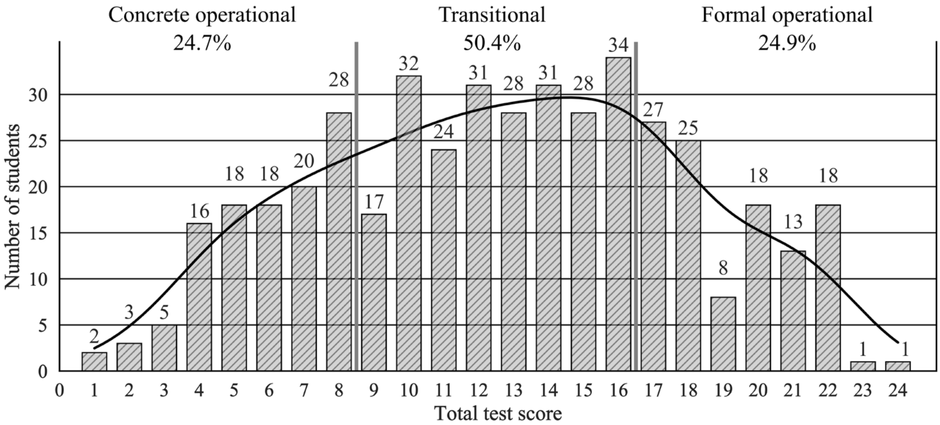
The item discrimination index D is a measure of the discriminatory power of each item in a test, it measures the extent to which a single test item distinguishes students with the better total scores well from those with lower ones. The possible range for the item discrimination index D is <–1, +1> with negative values for the questions answered better by the students with lower total test scores. There are more ways how to determine the groups with higher and lower scores based either on median or quartiles. In our case, following (Ding et al., 2013), we divided the whole sample of students according to quartiles and proceed with two different groups of equal size, a high group H and a low group L (using the top 25% as the high group and the bottom 25% as the low group). For a specific test item, one counts the number of correct responses in both H and L groups: namely, N_H and N_L. If the total number of students who take the test is N, then $D = (N_H - N_L)/(N/4)$. An item is typically considered to provide good discrimination if $D \geq .3$.

The point biserial coefficient (sometimes referred to as the reliability index for each item) is a measure of the consistency of a single test item with the whole test. It reflects the correlation between students' scores on an individual item and their scores on the entire test and is basically a form of the correlation coefficient. The point biserial coefficient has a possible range of <−1, +1>. Ideally, all items in a test should be highly correlated with the total score, but that is somewhat unrealistic for a test with a large number of items. The criterion widely adopted for measuring the "consistency" or "reliability" of a test item is rpbs ≥ .2 (Ding et al., 2013).

The Kuder-Richardson reliability index is a measure of the self-consistency of a whole test without administering one test twice or designing two tests, i.e., by using the information from one test administered just once. Using the formula KR-21 (Ding et al., 2013), possible values for the KR-21 reliability index fall into the range <0,1>. Different tests for various purposes have different criteria. A widely accepted criterion is that tests with a reliability index higher than .7 are reliable for group measurement and tests with a reliability index higher than 0.8 are reliable for individual measurement. Under most circumstances in physics education, evaluation instruments are designed to be used to measure a large group of students, so if a certain physics test has a reliability index greater than .7, one can safely claim it is a reliable test.

Figure 1

Histogram of Total Test Scores with a Density Estimate to Smooth the Distribution Shows that Nearly One Quarter of the Students Fall into the Lowest Piagetan Level of Concrete Operational Reasoning



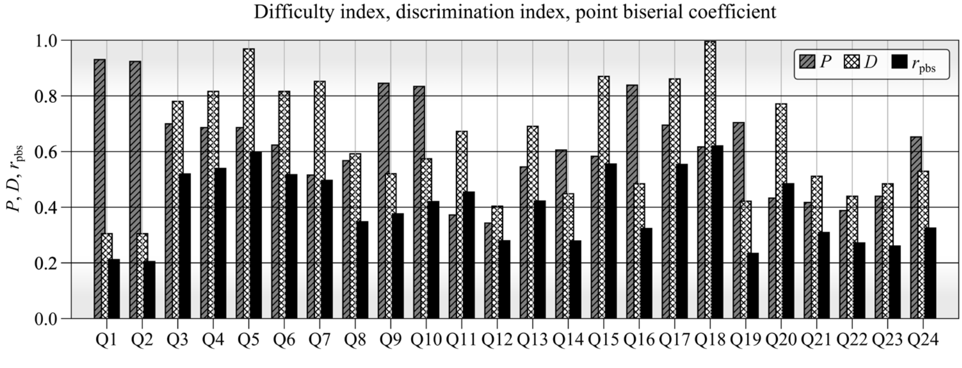
Ferguson's delta is another whole-test statistic measuring the discriminatory power of an entire test by investigating how broadly the total scores of a sample distribute over the possible range. If a test is designed and employed to discriminate among students, one would like to see a broad distribution of total scores. The possible range of Ferguson's delta values is <0,1>. If a test has Ferguson's delta greater than .9, the test is considered to offer good discrimination.

The scores in the six dimensions of scientific reasoning covered by the LCTSR were compared and a correlation matrix between them found to identify, which of them are most important for the overall test ranking. Based on the obtained results two hypotheses were tested – first, that the males over-score females and second, whether the students choosing the science teacher preparation study programmes reach a different level of scientific reasoning. Both hypotheses were checked by the Mann-Whitney U test with a significance level α = .05 and a two-sided *t*-test for identical means with the less strict assumption of different variances. The data and results from the years 2018 and 2019 were studied in a bachelor thesis, later the data from the autumn 2020 were included and previous conclusions confirmed. The statistical calculations were performed by the *Python* programming language using its open-source data analysis and manipulation library *pandas* and its module for statistical functions *scipy.stats*.

Research Results

The histogram of LCTSR scores for all N = 446 students is in Figure 1. The corresponding average is 12.7 (52.8% of the possible maximum), median 13 with a standard deviation of the distribution 5.13. It can be seen that nearly one quarter of the students entering our courses (24.7%) rank into the lowest category of concrete operational reasoning and approximately the same part (24.9%) into the highest category of formal operational, the rest (50.4%) belong to the transitional category. Only one student reached the full score of 24 points. In the context of the Czech Republic, among prospective physics teachers in the first year of their study at the Faculty of Mathematics and Physics, Charles University, Prague in 2011 there were only 9.3% concrete reasoners, and 59.3% students reached the highest level of formal operational reasoning. Therefore, our proportion of the students in the low concrete operational category is surprisingly high which may indicate that some students in this category could face problems in learning scientific concepts.

Figure 2
Item Characteristics for All 24 Questions in the LCTSR: Difficulty Index P, Discrimination Index D and Point Biserial Coefficient r_{pbs}



The density estimate to smooth our LCTSR scores distribution in Figure 1 shows the asymmetry in the direction of lower scores, which also confirms the negative skewness of the distribution – 0.0386. A negative kurtosis –0.817 means that the distribution is flatter than a normal distribution with the same mean and standard deviation.

The question-item measures are summarized in Figure 2. With the average difficulty index $\bar{P}$= .623 and the minimum value .343 for question 12, we can conclude

https://doi.org/10.33225/BalticSTE/2021.74

that there are no extremely difficult questions with $P < .3$ in LCTSR. The maximum P values 0.930 and .924 belong to questions 1 and 2 respectively; those two easiest questions dealing with the conservation of mass are out of the suggested range, and we can mark them as too easy. Question 12 belongs to the pair of questions combining identification and control of variables and probabilistic thinking as a second question providing a reason for the answer to question 11.

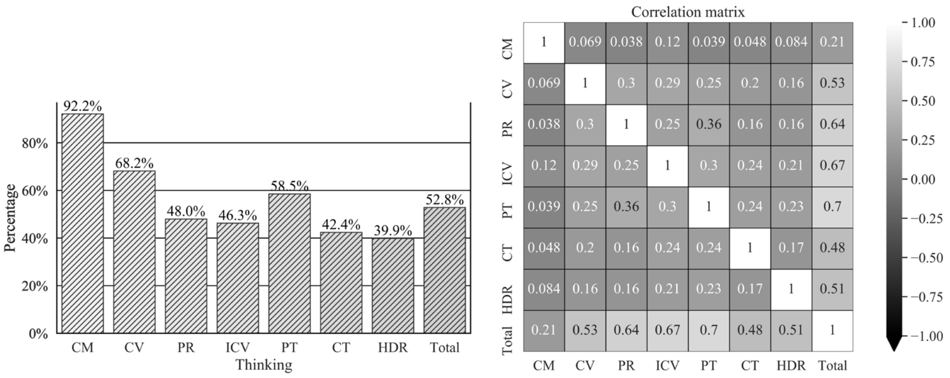
In our case, the average discrimination index $\overline{D} = .630$, with a minimum value .305 for questions 1 and 2 with the highest P values; the easiest questions naturally have the lowest discriminatory power (but still $D \geq .3$). The highest values of D correspond to questions 5 ($D = .969$) and 18 ($D = .996$) which represent proportional thinking and advanced probabilistic thinking, respectively. Similarly, our average point biserial coefficient $\overline{r_{\mathrm{pbs}}} = .400$ is satisfactory, even the lowest values .212 and .205 for the easiest questions 1 and 2 respectively are above the limit .2.

The Kuder-Richardson reliability index value .847 obtained for LCTSR is again compatible with the criterium to be over .7 and shows the test can be considered as reliable. Obtained value for the Ferguson's delta .985 is also satisfactory (> .9).

In Figure 3 we can see the scores evaluated according to the categories of scientific reasoning and the correlation matrix among them and the total test scores. The most successful is the conservation of mass (92.2%), which corresponds to the easiest questions 1 and 2 as discussed above. Not surprisingly, the lowest scores correspond to hypothetico-deductive thinking and correlation thinking (39.9% and 42.4% respectively). It is in partial agreement with the Thailand study (Piraksa et al., 2014), in which the lowest scores correspond to hypothetical-deductive reasoning, control of variables and proportional thinking.

Figure 3

Scores in the components of scientific reasoning defined by Lawson (in percent to the maximum possible scores in those parts) and the corresponding correlation matrix (CM – Conservation of mass, CV – Conservation of volume, PR – Proportional reasoning, ICV – Identification and control of variables, PT – Probabilistic thinking, CT – Correlation thinking, HDR – Hypothetico-deductive reasoning, Total – Total test score)



Correlation matrix

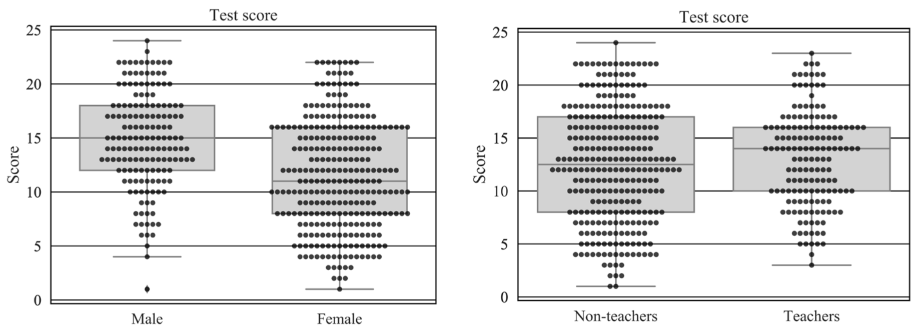
	CM	CV	PR	ICV	PT	CT	HDR	Total
CM	1	0.069	0.038	0.12	0.039	0.048	0.084	0.21
CV	0.069	1	0.3	0.29	0.25	0.2	0.16	0.53
PR	0.038	0.3	1	0.25	0.36	0.16	0.16	0.64
ICV	0.12	0.29	0.25	1	0.3	0.24	0.21	0.67
PT	0.039	0.25	0.36	0.3	1	0.24	0.23	0.7
CT	0.048	0.2	0.16	0.24	0.24	1	0.17	0.48
HDR	0.084	0.16	0.16	0.21	0.23	0.17	1	0.51
Total	0.21	0.53	0.64	0.67	0.7	0.48	0.51	1

Also, (Moore & Rubbo, 2012) reported their non-STEM students to demonstrate significant difficulty with proportional and hypothetico-deductive reasoning. From our correlation matrix in Figure 3, we can see, that the mutual correlation among the scores in the categories are loose (all Spearman correlation coefficients below .5), a little bit stronger correlation can be found between the total test scores and identification and control of variables category, probabilistic and proportional thinking (Spearman correlation coefficients above .6). Recently, it has been shown by (Hejnová et al., 2018), that for the 14–15-year students the ability to use the existing knowledge strongly correlates with three dimensions of the scientific reasoning structure: proportional reasoning, control of variables and probability thinking, which demonstrates the highest correlation with the total test scores in our research.

Working with the test scores we noticed that there could be a significant difference in the performance of male and female students (see the box plot in Figure 4), which was also indicated in the data we privately obtained from the Faculty of Mathematics and Physics, Charles University, Prague. Among our respondents, there were 286 females with a mean score 11.5 (median 11, minimum 1, maximum 22) and 160 males with a mean score 14.9 (median 15, minimum 1, maximum 24). Mann-Whitney U test for the distributions and *t*-test of the scores for the mean score differences confirmed, that the difference is statistically significant (*p*-value .000). The results of other studies are not unequivocal in such a conclusion. For the original Lawson 1978 test administered to secondary school students in Israel and the U.S.A. was also found that, in general, males outperformed females (Hofstein & Mandler, 1985). On the other hand, the results of the Thailand LCTSR study (Piraksa et al., 2014) did not indicate that gender should significantly impact students' scientific reasoning ability. Similarly, the results of Indonesian research (Novia et al., 2018) showed that no statistical difference between male and female scientific reasoning mean scores were observed and that gender and age do not significantly impact students' scientific reasoning ability in the first year of university education. Thus, besides the fact the LCTSR was designed by a male, the gender difference in the mean scores may also have roots in the local cultural and historical backgrounds.

Figure 4

Box Plots Showing the Quartiles of the Datasets and the Whisker of the Dataset Extends for the Total Scores with Respect to Gender and the Kind of a Study Programme of the Respondents



Another important question we wanted to answer relates to a common belief and opinion, that the carrier of prospective teachers is mostly chosen by students with worse high school results who give up or fail to proceed with more demanded study programs (e.g., law, economy, medicine, engineering, or technical science programs). Therefore, we checked whether there is a significant difference between "future teachers" and "non-teachers" in scientific reasoning skills measured by LCTSR. In our sample, the 156 prospective teachers reached a mean score 13.1 (median 14, minimum 3, maximum 23), 290 non-teachers mean score 12.5 (median 13, minimum 1, maximum 24). Both the Mann-Whitney U test (*p*-value .154) and the *t*-test for the different mean scores (*p*-value .234) confirmed, there is no statistically important difference in the distributions of the scores for those two groups. Thus, we can conclude, that at least within our Faculty of Science, the students of science education programmes should not be automatically taken as worse or with lower prerequisites to acquire, learn and understand scientific concepts and in solving problems. We certainly realize that might be rather a faculty-specific conclusion, and we cannot compare it with any similar study at this point.

Conclusions

It was demonstrated and verified the LCTSR is a relatively reliable tool fulfilling the demanded values for item characteristics like item difficulty index, discrimination index and point biserial coefficient (maybe except for the easiest questions 1 and 2, which may be considered as motivating). Also, both the Kuder-Richardson reliability index and Ferguson's delta measuring the test self-consistency and overall discrimination power meet usual criteria. We were also able to identify the identification and control of variables category, probabilistic thinking, and proportional reasoning as the key dimensions of the scientific reasoning correlating most with the total test scores. In our sample, there is a statistically significant gender bias of the score distributions, but students entering science education programmes achieved a comparable mean level of scientific reasoning as their colleagues enrolling in technical-scientific programmes.

The open problems for some further research should answer the relation between the LCTSR scores and possible learning gains within introductory courses, which is not generally straightforward. Also, it is not clear, to which extent scientific reasoning is developed within some types of courses.

On the other hand, training in scientific reasoning may also have a long-term impact on student academic achievement. With respect to relatively low mean scores for our sample, there is evidently demanded to develop and strengthen the scientific reasoning also within the introductory university courses, as most of our students belong to the transitional reasoners' category. Our results agree with (Bao, et al. 2009a), that students start to fully develop the basic reasoning abilities around their college years. The take-home message might be that the sorts of activities that require students to develop, explain and defend their reasoning help them to develop cognitively in ways that lectures or even peer interactions in a lecture-like environment cannot provide much effectively. We cannot assume that our incoming students are able to reason in the sophisticated ways which they need to really understand abstract and complex concepts like force and energy. If we do really care that our students get these concepts, we need to provide explicit opportunities to improve their reasoning. Hence, it is necessary to make reasoning explicit in the instruction. Although the reasoning abilities tested in the LCTSR may appear simple to expert scientists, these are crucial fundamental components for more sophisticated skills. However, to assess the reasoning abilities of senior college students and graduate students, we need to develop questions that involve more advanced reasoning components. This also implied that instructional pedagogy in science classrooms and university or college lecture rooms should be more emphasized on the way of teaching how to reason casually based on hypothesis generation, how to design well fair science experiments, and how to determine the correlation between target variables, to enhance the development of students' scientific reasoning ability.

Acknowledgement

The authors thank a lot to RNDr. Irena Dvořáková, Ph.D. (Department of Physics Education, Faculty of Mathematics and Physics, Charles University, Prague) for the introduction into the topic of the LCTSR, the Czech translation of the test and sharing her results collected within several years across the Czech Republic. The authors are also very grateful to the colleagues RNDr. Renata Holubová, CSc., RNDr. Marta Klečková, CSc., doc. RNDr. Roman Kubínek, CSc., Mgr. Zdeněk Pucholt, Mgr. Jan Říha, Ph.D., Mgr. David Smrčka, Ph.D., and Mgr. Vlastimil Vrba, Ph.D., for their kind help during the testing and data collection.

Declaration of Interest

Authors declare no competing interest.

References

Bao, L. et al. (2009a). Learning and scientific reasoning. *Science*, *323*(5914), 586–587. https://doi.org/10.1126/science.1167740

Bao, L., Fang, K., Cai, T., Wang, J., Yang, L., Cui, L., Han, J., Ding, L., & Luo, Y. (2009b). Learning of content knowledge and development of scientific reasoning ability: A cross culture comparison. *American Journal of Physics*, *77*(12), 1118–1123. https://doi.org/10.1119/1.2976334

Bao, L., Xiao, Y., Koenig, K., & Han, J. (2018). Validity evaluation of the Lawson classroom test of scientific reasoning. *Physical Review Physics Education Research*, *14*(2), Article 020106. https://doi.org/10.1103/PhysRevPhysEducRes.14.020106

Ding, L., Chabay, R., Sherwood, B., & Beichner, R. (2006). Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment. *Physical Review Special Topics - Physics Education Research*, *2*(1), 010105. https://doi.org/10.1103/PhysRevSTPER.2.010105

Ding, L. et al. (2013). Detecting progression of scientific reasoning among university science and engineering students [online]. *Physics Education Research Conference 2013* (Portland, Oregon: July 17-18, 2013), 125–128. https://www.per-central.org/items/detail.cfm?ID=13125

Eaton, P., Johnson, K., Frank, B., & Willoughby, S. (2019). Classical test theory and item response theory comparison of the brief electricity and magnetism assessment and the conceptual survey of electricity and magnetism. *Physical Review Physics Education Research*, *15*(1), Article 010102. https://doi.org/10.1103/PhysRevPhysEducRes.15.010102

Fabby, C., & Koenig, K. (2015). Examining the relationship of scientific reasoning with physics problem solving. *Journal of STEM Education: Innovations and Research*, *16*(4). https://jstem.org/jstem/index.php/JSTEM/article/view/1904

Hejnová, E., Eisenmann, P., Cihlář, J., & Přibyl, J. (2018). Relations between scientific reasoning, culture of problem solving and pupil´s school performance. *Journal on Efficiency and Responsibility in Education and Science*, *11*(2), 38–44. https://doi.org/10.7160/eriesj.2018.110203

Hofstein, A., & Mandler, V. (1985). The use of Lawson's test of formal reasoning in the Israeli science education context. *Journal of Research in Science Teaching*, *22*(2), 141–152. https://doi.org/10.1002/tea.3660220205

Lawson, A. E. (1978). Development and validation of the classroom test of formal reasoning. *Journal of Research in Science Teaching*, *15*(1), 11–24. https://doi.org/10.1002/tea.3660150103

Lawson, A. E. (2000). *Classroom test of scientific reasoning*. Arizona State University.

Maloney, D. P. (1981). Comparative reasoning abilities of college students. *American Journal of Physics*, *49*(8), 784–786. https://doi.org/10.1119/1.12676

Moore, J. C., & Rubbo, L. J. (2012). Scientific reasoning abilities of nonscience majors in physics-based courses. *Physical Review Special Topics - Physics Education Research*, *8*(1), Article 010106. https://doi.org/10.1103/PhysRevSTPER.8.010106

Nieminen, P., Savinainen, A., & Viiri, J. (2012). Relations between representational consistency, conceptual understanding of the force concept, and scientific reasoning. *Physical Review Special Topics - Physics Education Research*, *8*(1), Article 010123. https://doi.org/PhysRevSTPER.8.010123.

Novia, N., Syamsu, S., & Riandi, R. (2018). Student's achievement in Lawson's classroom scientific reasoning (LCTSR): The effect of gender and age on scientific reasoning ability. *International Conference on Mathematics and Science Education of Universitas Pendidikan Indonesia*, *3*, 542–547. http://science.conference.upi.edu/proceeding/index.php/ICMScE/article/view/82

Piraksa, C., Srisawasdi, N., & Koul, R. (2014). Effect of gender on student's scientific reasoning ability: A case study in Thailand. *Procedia - Social and Behavioral Sciences*, *116*, 486–491. https://doi.org/10.1016/j.sbspro.2014.01.245

Pyper, B. A. (2011). Changing scientific reasoning and conceptual understanding in college students. [online]. *Physics Education Research Conference 2011* (Omaha, Nebraska: August 3-4, 2011), *1413*, 63–65. https://www.per-central.org/items/detail.cfm?ID=11813

Sriyansyah, S. P., & Saepuzaman, D. (2017). Prospective physics teachers' consistency and scientific reasoning in the learning of force concept. *Proceedings of the 2016 International Conference on Mathematics and Science Education (Advances in Social Science, Education and Humanities Research)*, 21–24. https://dx.doi.org/10.2991/icmsed-16.2017.5

Susilowati, S. M. E., & Anam, K. (2017). Improving students' scientific reasoning and problem-solving skills by the 5E learning model. *Biosaintifika: Journal of Biology & Biology Education, 9(3)*, 506–512. https://doi.org/10.15294/biosaintifika.v9i3.12022

Zhou, S.-N., Liu, Q.-Y., Koenig, K., Li, Q.-Y., Xiao, Y., & Bao, L. (2021). Analysis of two-tier question scoring methods: A case study on the Lawson's classroom test of scientific reasoning. *Journal of Baltic Science Education*, *20*(1), 146-159. https://doi.org/10.33225/jbse/21.20.146

Received: *May 28, 2021* Accepted: *August 14, 2021*

Cite as: Hrouzková, T., & Richterek, L. (2021). Lawson classroom test of scientific reasoning at entrance university level. In V. Lamanauskas (Ed.), *Science and technology education: Developing a global perspective. Proceedings of the 4th International Baltic Symposium on Science and Technology Education (BalticSTE2021)* (pp. 74-85). Scientia Socialis Press. https://doi.org/10.33225/BalticSTE/2021.74