Pre-service teachers’ vocabularies of the language of science in the context of learning about electrons and photons

Karoliina Vuola, Maija Nousiainen and Ismo T. Koponen

University of Helsinki, Finland

Teaching and learning the language of science is an important part of science education. Learning the vocabulary of science plays a key role in learning the language of science. The meaning of abstract scientific terms builds on their connections with other terms and how they are used. In this research, we study pre-service physics teachers’ physics-related vocabularies and investigate how rich a vocabulary they use and what similarities and differences there are in their vocabularies regarding electrons and photons. We investigate the connectedness of physics terms by categorizing them according to their role in explaining quantum physics and carry out a lexical network analysis for N=60 written reports. The analysis shows that vocabularies do not share much similarity and the reports reflect narrow images of photons and electrons. We conclude that science teacher education needs to pay attention to explicit teaching of the language of science for pre-service teachers.

ARTICLE DETAILS

LUMAT General Issue
Vol 11 No 2 (2023), 1–34

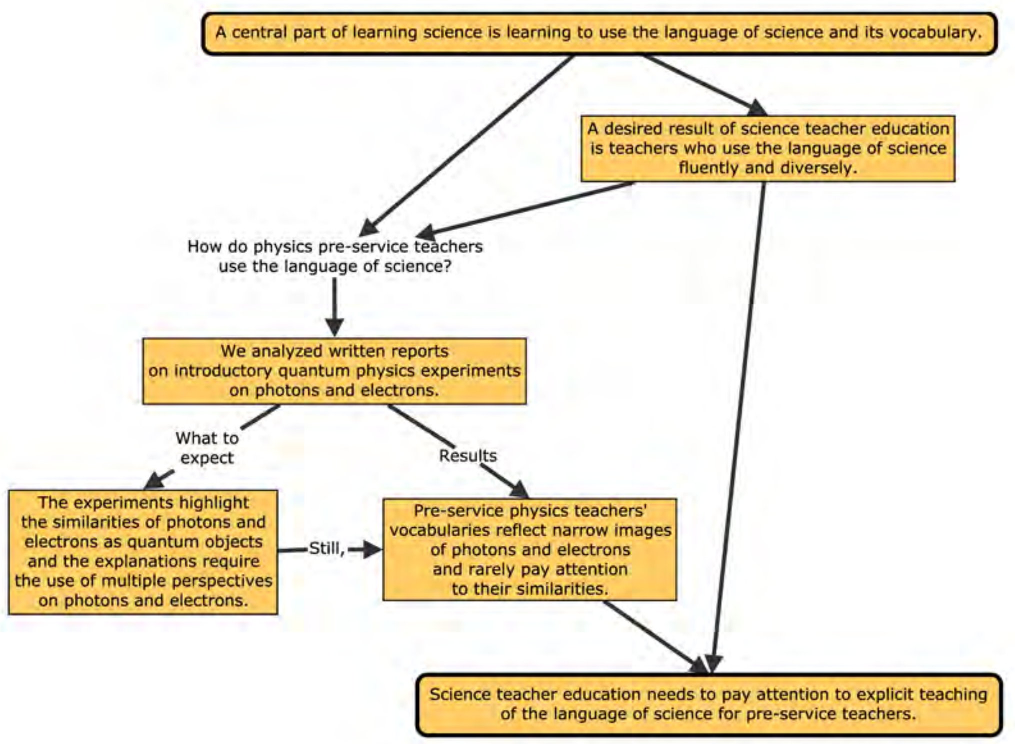
Received 1 February 2023
Accepted 7 June 2023
Published 22 June 2023

Pages: 34
References: 24

Correspondence:
karoliina.vuola@helsinki.fi

https://doi.org/10.31129/LUMAT.11.2.1924

Keywords: the language of science, pre-service teachers, quantum physics, vocabulary



1 Introduction

In learning science, content knowledge is obviously central and a major part of what is to learn scientific terms and concepts and the normative ways to use them; science has its own language with own vocabulary, semantics, and syntax, and therefore some researchers refer to the language of science (see e.g., Lemke, 1990; Yun & Park, 2018). The language of science and its vocabulary are intertwined with doing science from designing research to discussing its results and implications (see e.g., Bratkovich, 2018; Lemke, 1990). Science teaching often concentrates only on content knowledge with lesser explicit attention on the language of science itself, although teachers need to be able users of the language of science. Some researchers call this untaught and indirect demand to know the language of science a "hidden curriculum" (see e.g., Schleppegrell, 2004).

Science teachers and authors of science textbooks tend to forget that their students are not very fluent in understanding the language of science (Yun & Park, 2018). For example, the language used in science classroom differs greatly from everyday language (Fang, 2006) and science terms usually have a precise meaning that is far from the everyday terms that students may know from their daily lives, for example "force". Compared to everyday language, students use the language of science rarely, and they are not familiar with the stylistic norms of the language of science (see e.g., Lemke, 1990). In addition, when they advance to higher school levels, they might have difficulties to handle the abundance of new terminology in science textbooks (Yun, 2020). Consequently, students may find reading science texts difficult because their major problem in learning science is learning the language of science (Wellington & Osborne, 2001). The language of science consists of scientific terms that are very dense, but such very technical terms are used in science because they enable exact and compact description of phenomena and their relations. Compared to the texts that students are used to reading and writing, the characteristics of scientific language make scientific texts dense, abstract, and hard to read and write (Fang, 2006). Knowing these characteristics of scientific language is important and it might help students to better understand and learn the language of science. Learning to read science and reading to learn science are the key elements in developing scientific literacy (Glynn & Muth, 1994; see also Keys et al., 1999). It is important to support students in their attempts to learn the language of science and guide them to get used to scientific terms and the use of the language of science (Yun & Park, 2018).

In learning the language of science, the challenge is two-fold: first, there is a new vocabulary consisting of science specific words, terms, and concepts; second, one needs to assimilate new subjects (e.g., some phenomenon) using the new vocabulary. A comprehensive vocabulary is then a prerequisite of knowing the language of science, and it reflects students' ability to use the language of science (Koponen & Nousiainen, 2020; Nousiainen & Koponen, 2020; Södervik et al., 2021). Especially important part of the language of science is knowing science terminology: what terms and concepts mean, and how they are used in different contexts. Scientific terms usually do not have directly observable real-life referents (for example the terms "field" or "energy" in physics), and therefore, the meaning of such abstract scientific terms builds on their connections with other terms and how they are normatively used. Thus, by looking at the use of terminology and the connections between terms, we can get a grip on the language of science (Stenhouse, 1986).

In learning the language of science, it is important to facilitate multifaceted use of the language of science in science classes and in learning science (Glynn & Muth, 1994; Mercer, 2009). For example, teaching and learning science needs to enhance the ability to read, write, and speak science and about science because such activities are shown to increase students' understanding of science (cf. Glynn & Muth, 1994; Keys et al., 1999; see also, Lachner et al., 2021). Consequently, we need more techniques to increase students' systematic talking and writing during science classes to enculturate students' systematic use of scientific language (Schwarz, 2009). Thus, higher education should pay more attention to the explicit teaching of the language of science and guide university students to use it in communicating and discussing their scientific ideas and views. This will help university students to build and express their understanding of content knowledge better (Bratkovich, 2018; see also Keys et al., 1999) which is especially important for pre-service teachers. A desired result of science teacher education is teachers who use the language of science fluently and diversely. Therefore, science educators need skills to revise their teaching towards a more language-sensitive direction (Bratkovich, 2018).

2 Contextualizing the use of language of science in learning quantum physics

Quantum physics serves us a fruitful context to study how pre-service teachers use the language of science. Quantum physics is known for its elusive notion of “particles”, “waves” and “quanta”, with multiple context dependent meanings attached to them, and thus, too often used by students’ incoherently and unsystematically. Consequently, it is common to find much confusion and inconsistencies in how students use concepts electron and photon (e.g., being particle-like, wave-like or interpreted as field excitation) in making sense of different quantum phenomena. Thus, paying attention to the use of concepts of electron and photon as part of one’s QM vocabulary we get insight into the role of vocabulary and normative use of terms as part of students’ process of learning the language of science.

3.1 Teaching and learning about photon, electron, and the double-slit experiment

In teaching quantum physics, the double-slit experiment for dim light or dim electron beam is a much-used demonstration of how classical models fail at predicting the behavior of both light and electrons (Cheong & Song, 2014; Hobson, 2005). The experiment shows similar results for both objects: First, an interference pattern emerges in both cases though interference is an obvious expectation only in the case of light. Secondly, when the light is dim enough or the electron beam weak enough, the interference pattern can be observed building up from single hits, which is classically expected only in the case of particle-like electrons. The emerging interference pattern can be interpreted as a proof of wave nature and the single hits as a proof of particle nature. Thus, the outcome of the experiment problematizes both classical notions: light (photons) as classical waves and electrons as classical particles. From the perspective of quantum (field) theory, the double-slit experiments underline the symmetry between photon and electron as quantum objects: both can be considered as quanta of continuous fields. In addition, classical determinism is questioned, as localization of single photons or electrons is random and unpredictable, while the collective outcome of hits is a regular and predictable interference pattern. (Hobson, 2005)

Both upper secondary school and university students’ explanatory models of quantum phenomena (such as the double-slit experiment) tend to range from purely classical models to successful use of appropriate quantum models (see Ayene et al., 2019; Krijtenburg-Lewerissa et al., 2017). First, explanatory models built solely on classical physics typically contradict with some of the observations and consider objects distinctly either waves or particles (Ayene et al., 2019; Krijtenburg-Lewerissa et al., 2017; see also Hobson, 2005). Second, explanations combining classical and quantum models are often inconsistent and depend much on the context. For example, students may note a disagreement between wave and particle properties of light and mention duality, but they seem to ignore that and jump from one model to another. Third, there are explanations that are beginning to use quantum mechanics' models fluently. Most of students' explanations fall into the first and second category even at university level (Ayene et al., 2019) and even the best students tend to have challenges in explaining the double-slit experiment (Cheong & Song, 2014).

As described above, students have difficulties in explaining quantum phenomena, but quantum theory has also many competing ways of understanding that physicists disagree on. Cheong & Song (2014) suggest that such disagreement depends on one's understanding of physics as a science. Teachers should be aware of different ways to interpret the phenomena and be able to teach them and underline which parts of the subject matter is consensus knowledge and which are subjects under a debate (see Ayene et al., 2019; Cheong & Song, 2014; Krijtenburg-Lewerissa et al., 2017). Cheong and Song (2014) recommend that the double-slit experiment could be used repeatedly at different stages of physics studies because the experiment is concrete enough for qualitative interpretation even at upper secondary school level, and learning more physics allows interpretations that are more advanced as well as a chance for students to notice their progress.

3 Research questions, research design and sample

In this research, we study pre-service physics teachers' vocabularies on photons and electrons, and then, similarities between the vocabularies, investigate how extensively they use scientific terms and words closely related to other scientific terms, and next, what similarities and differences there are in their vocabularies. The specific research questions are:

1. What are the most frequently used physics terms and their connections in pre-service teachers' written reports, where they explain outcomes of double-slit experiment for photons and electrons?

2. How extensively students' vocabularies overlap (i.e., how large are the shared parts of the vocabularies)?

To address the research questions, we analyzed written reports produced by pre-service teachers. The analysis was performed on two levels of accuracy: first, we identified physics terms and classified them into different contextual categories; second, we constructed a description of relationships between words as they appeared in texts. This analysis provided us kinds of lexicons in form of networks of words, called lexical networks in what follows, and allowed us to study the differences between pre-service teachers' vocabularies.

3.1 Participants

The participants of this study were pre-service physics teachers who will obtain a license to teach physics in upper secondary level (N=30; male 18, female 12). The study was carried out at a large research-intensive university in Finland. The participants were in their third or fourth years of university studies and they all had a background of basic physics studies, including quantum physics. The data was collected as part of the physics teacher preparation course (at the intermediate level). The course focused on the organization of introductory quantum physics content knowledge for teaching purposes at the level taught in upper secondary school. The mean age of the participants was 29 years (min–max: 21–46 years). All participants came from a homogeneous cultural background, and all shared the same first language (Finnish).

3.2 Data and its Context

The data came from two tasks, in which pre-service physics teachers were asked to express their understanding of the double-slit experiment with extremely dim light (interpreted as consisting of single photons) and the double-slit experiment with single electrons. The pre-service teachers did not carry out the experiments themselves but were asked to give written explanations of the phenomena in these well-known experiments. These scientific writing tasks included a written report and a chart to illustrate how the presented physics concepts relate to each other. The tasks were designed to enhance pre-service teachers' content knowledge and their skill in using the language of science. The instructions for completing the task were designed so that the pre-service teachers were required to write down an explanation for the basic purpose of the experiment, the findings, and the argumentation to support the findings. The length of the report was usually 1–2 pages. The assignment can be found in Appendix A (see also, Nousiainen, 2017; Mäntylä & Nousiainen, 2014; Nousiainen & Koponen, 2020).

The pre-service teachers were used to such tasks from previous courses, but they might not have had routines for expressing a multitude of interpretations for a single phenomenon. However, the task itself was designed so that it would encourage the pre-service teachers to express a multifaceted view of how the double-slit experiments can be interpreted. Compared to interviews (as a research method), such tasks can give us a more authentic picture of how pre-service teachers express their understanding. The interview situation itself and the expressions the interviewer uses during the interview affect the expressions of the answers (see e.g., Halldén et al., 2007), and it is crucial to minimize such bias when examining terms and expressions.

Data was collected in the form of written reports. Both reports were completed prior to a weekly discussion session about the topic and submitted in advance. As base material, the pre-service teachers read a research article that suggests that both electron and photon can be interpreted as field quanta in the context of the double-slit experiment (Hobson, 2005). We thus assumed that the explanations of the behavior of electrons and photons in the double-slit experiment should contain similarities, especially regarding the quantum terminology. Hence, we hypothesized that there could be vocabularies for photons and electrons which are similar in such a sense.

3.3 Data Handling

Voluntary participation, informed consent, and anonymity of the participants were ensured during the research process. In collecting the data, the pre-service teachers were asked for permission to use their written reports as research data. Consent forms, which explained the purpose of the research, were used to obtain their permission. The pre-service teachers were also given the option not to participate in the research. The pre-service teachers were given the opportunity to ask the researchers about the study and received detailed answers to their questions. All data was stored in encrypted external storage devices and only accessible to the researchers. All researchers had agreed to follow the regulations conforming to the national laws for handling data. The research did not involve intervention in the physical integrity of the participants in any way and thus, according to the National Advisory Board on Research Integrity, did not require an ethics review.

4 Data Analysis

As scientific concepts get their meaning in connections to other concepts and thus form a network, network analysis is a practical way of analyzing contents of scientific texts. This study utilized a network analysis method that was developed in two pilot studies for studying physics vocabularies or lexicons (Koponen & Nousiainen, 2020; Nousiainen & Koponen, 2020). The topic of these studies was also introductory quantum physics, but the pilot studies focused heavily on the development and technical aspects of the method. They used two experts' and four physics students' texts as their small samples, whereas now the method was usable for much bigger samples. The pilot studies focused on the similarities between vocabularies, but in this research, the same method allowed us to also examine differences and identify specific themes within the vocabularies. One advantage of this network analysis is that mostly it is based solely on grammar and syntax and can be done automatically. This reduces the influence of a researcher on the results and enables effective content analysis for larger samples. A comprehensive and well-organized lexical network is a necessity to be able to use the language of science. Still, it alone is not enough. This means that the lexical networks resulting from this analysis tell us about the potential of the vocabularies: with limited lexical networks, not much physics knowledge can be communicated, but comprehensive ones have the potential to do that.

Analysis of the terms used in the reports was carried out in a straightforward way by identifying the relevant physics terms[1] and how many times they appear (see Appendix B Table B2 for analyzed example sentences of the sample). Then the terms were divided into nine *thematic profile categories* P1 to P9 based on each term's role in quantum physics terminology (for detailed description, see Table 1). The thematic profile categories were defined by three expert physicists, along similar lines as in our previous research (see Nousiainen & Koponen, 2020). We condensed this information on physics terms into a nine-dimensional vector. The vector elements are term counts in categories P1-P9. In what follows, we refer to this as *profile weight*.

---

[1] Established physics terms are interpreted as one term even if they consist of several words (e.g., light quantum).

Table 1. The nine thematic profile categories P1–P9 for quantum physics terms with some central example terms describing each profile category.

	Profile category	**Example terms**
P1	Classical field and radiation	Magnetic field, electromagnetic radiation, light
P2	Classical energy and intensity	Conservation of energy, intensity maximum, kinetic energy, power
P3	Classical wave model	Interference, diffraction, wavelength, scattering, frequency
P4	Classical particle model	Elementary charge, mass, trajectory
P5	Quantum mechanics	Elementary particle, electron field, quantum of energy, photon model, state
P6	Stochastics	Predictable, random, statistical, probability distribution
P7	Duality	Wave nature, particle nature, de Broglie hypothesis, dual
P8	Localization and identification	Place of occurrence, observation point, local, hit, single, individual
P9	Double-slit experiment	Diffraction experiment, electron beam, double-slit system, shutter speed

After this, the data was analyzed using so-called stratified lexical networks that were constructed based on grammatical sentence analysis. This analysis focused on nouns and verbs, and the text analysis itself was based only on grammar and syntax. First, the data (reports) was examined sentence by sentence. The sentences were classified into contexts that were defined based on the meaning of the texts. The contexts were established topics of these well-known double-slit experiments discussed in the reports, for example, classical model for light, carrying out the double-slit experiment, observation of single hits on the screen and its qualitative interpretation, and so on. A comprehensive list of contexts can be found in Appendix B (see Table B1) as well as example sentences for the most used contexts from the sample (see table B2). Next, we noted all nouns and root verbs from each clause but paid special attention to the physics terms described above in the profile categories. After this, we transformed the simplified text structure (each clause replaced with its context, root verb, nouns and profile categories P1–P9 of its relevant physics terms) into lexical networks where nouns (physics terms) are connected to root verbs and root verbs are connected to contexts. Lexical networks describe how various terms are connected to each other on the level of sentences and more broadly on a context level.

The lexical network analysis of terms gave us a comparable value, measured by communicability centrality (see Appendix C) which describes each term's role and connectedness in the network of terms. Based on this measure, we defined the total communicability centrality of a given profile P1–P9, which is a nine-dimensional vector with each dimension representing a profile category. In what follows, the total communicability centrality is referred to as *profile communicability*. Profile communicability can be determined both at the sentence and broader context level. The sentence level profile communicability describes how the profiles' terms are connected by shared sentences and the context level respectively by shared contexts. As the contexts can consist of several sentences and broader descriptions than can be expressed in single sentences, profile communicability is expected to be greater at the context level than the sentence level.

The data analysis done here utilized a network approach, but the results are understandable without a detailed explanation concerning the exact analysis method. We offer a brief description of the lexical network method in Appendix B, and it is reported in greater detail in the pilot studies (Koponen & Nousiainen, 2020; Nousiainen & Koponen, 2020). In the following results sections, we give concrete examples on how the calculated measures can be interpreted.

5 Results and their interpretations

We present both the results and their central interpretations in this section to make the results more apprehensible and to tie them to the conclusions more clearly.

5.1 The appearance of physics terms

The data consisted of N=60 reports. The number of physics terms in the reports varied from 36 to 280 (Md=77 and Avg=99). This distribution was wide, and it was skewed to lower values. Since the term count varied greatly between the reports (some being short and others very extensive), it was informative to inspect the relative share of physics terms in each report, i.e., compare the number of physics terms to the number of all nouns expressed in the reports. The percentage of physics terms compared to all nouns in a report varied between 23–65%, with arithmetic average and median both 49%. This relative distribution was more symmetrical than the absolute value distribution. This means that, in general, the longer the report was, the more physics terms there were. Altogether, we found 386 different physics terms that were categorized into the thematic profile categories P1 to P9 (see Table 1).

Table 2. The 30 most-used physics terms, their profile categorization (for description, see Table 1) and appearance in pre-service teachers' reports.

Term	Profile category	Total frequency in the reports	Number of reports using the term
Photon	P5	499	53
Light	P1	425	51
Electron	P5	377	36
Slit	P9	309	49
Particle	P4	305	56
Double-slit experiment	P9	238	56
Wave	P3	194	49
Interference pattern	P9	186	54
Particle nature	P7	137	48
Wave nature	P7	135	48
Single	P8	124	39
Energy	P2	97	38
Wave function	P5	94	25
Intensity	P2	92	34
Field	P1	82	26
Probability	P6	76	40
Screen	P9	69	23
Hit	P8	69	21
Wavelength	P3	64	30
Radiation	P1	63	22
Wave motion	P3	61	24
Double-slit	P9	60	29
Interference	P3	58	27
Hit point	P8	53	24
Wave quality	P7	53	24
Location	P8	52	28
Point	P8	48	5
Momentum	P2	47	24
Wave model	P7	45	18
Particle model	P7	44	22

The 30 most used physics terms included concepts from all categories and the top ten terms were expressed over a hundred times (see Table 2). These most used physics terms described well the physics content of the writing task.

5.2 Physics terms and lexical networks

The term count enabled us to utilize the nine-dimensional *profile weight* that states the number of physics terms in profile categories P1–P9. The number of physics terms was normalized between [0,1] because of the great variance in term count. This normalization helped us to compare the relative existence of thematic profile categories between different reports. The normalized profile weight was directly proportional to the absolute number of physics terms. Due to normalization, there was at least one profile category in each report with value 1, corresponding to the category with the most physics terms. Each report was also described with *profile communicability* (see Appendix C), with nine similar dimensions P1–P9 and normalization [0,1]. The higher the value, the more central the role played by the profile category in the report, i.e., the terms are closely connected to other terms through shared sentences or contexts, or one term is repeated often within nearby sentences. Maximum values showed us which thematic profile categories students presented most comprehensively, minimum values those that they tend to ignore altogether. Minimum value 0 implies that the profile category does not play any significant role in the report, which means that no physics term of that profile category is presented, or terms are used in isolation from the rest of the text, without any connection to the whole. Profile communicability was determined both at the sentence and context level. The values were practically equivalent, so we focused only on the context level profile communicability.

Graphic representations of profile weight and profile communicability for report examples are shown in Figure 1 in a nonagon form. The higher the value of profile weight (the larger the orange-lined nonagon), the more relevant physics terms the report contains (for example, compare Photon_26 and Electron_24 in Fig. 1). The higher the value of profile communicability (or the larger the blue-lined nonagon), the more central are the profile categories in the report (compare cases Photon_26 and Photon_4 in Fig 1). The more symmetrical the nonagon, the more evenly different profile categories are presented in the report. All nonagons of the sample were quite clearly asymmetrical, but interpreting symmetry more loosely, e.g., Photon_4 in Figure 1 had profile communicability positive in all categories P1–P9. There was great variance between the reports, which can be observed qualitatively by visual inspection of the charts. We explore the variance more precisely later in the results. The relationship of profile weight and profile communicability and their meaning is studied more deeply in what follows.

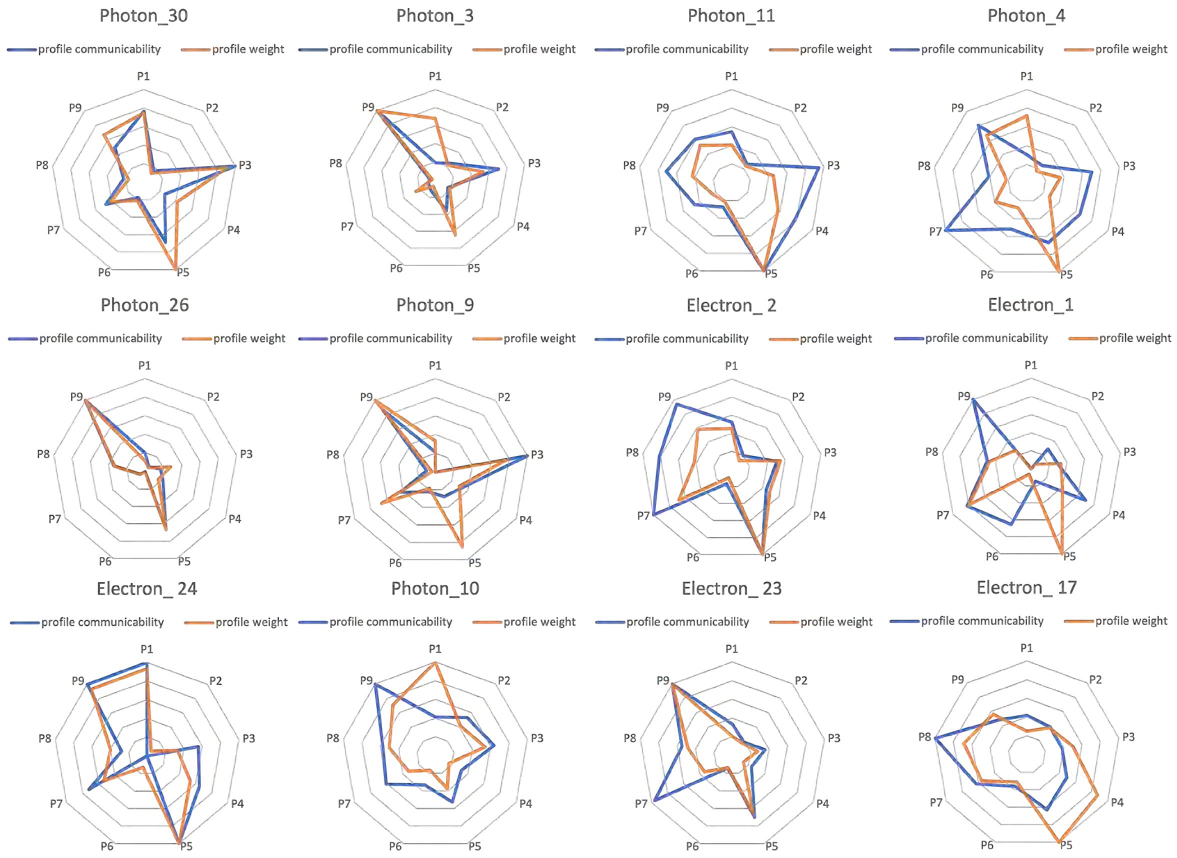


Figure 1. Twelve examples of profile weights and profile communicabilities. Labelled corners P1 to P9 refer to the profile categories listed in Table 1. The blue line represents profile communicability and the orange line profile weight. The center of the polygon corresponds to the minimum value zero and the values increase linearly to the outermost edge corresponding the maximum value one.

5.3 The relationship between profile weight and profile communicability

To examine the relationship between profile weight and profile communicability more precisely, we computed Spearman's and Kendall's rank correlation coefficients. Spearman's ρ was 0.713 using absolute profile weight and 0.812 using relative profile weight. Kendall's τ was 0.544 and 0.643, respectively. This statistical analysis supported the dependence between profile weight and profile communicability values. The correlation was even more significant when we consider relative profile weight instead of the absolute term counts. This correlation was an expected result: to make some viewpoint a central part of a report, we need to use vocabulary related to it.

The correlation explained why the nonagon plots of relative profile weight and profile communicability are essentially similar in form in most reports (e.g., see the plots in the first vertical column from the left in figure 1). Despite this, when we compared the plots, we could find cases in which the normalized profile weight would assume higher profile communicability in some profile categories (see the second vertical column from the left in figure 1), or lower profile weight was combined with higher profile communicability (see the third vertical column from the left in figure 1). There were also some cases where the profile weight plot and the profile communicability plot were more distinctly different in shape as the ratio of the profile weight and profile communicability varied more significantly depending on the profile category (see the rightmost vertical column in figure 1).

We can interpret these differences in terms of how pre-service teachers used physics terms in their reports: First, there were reports with an outright lack of physics terms in several categories. This was the case when a report's profile weight and profile communicability both had a low value in some categories. For example, Photon_26 had both values equal to or below 0.1 in the categories P2 (classical energy and intensity), P6 (stochastics) and P7 (duality); Electron_24 in the category P2. Second, it is possible that the terms were used in a detached manner. In this case, profile weight was notably higher than the corresponding profile communicability in some categories, for example, category P1 (classical field and radiation) in Photon_3 and category P5 (quantum mechanics) in Electron_1. These reports did contain relevant physics terms, but they were used in isolation so that the text was a collection of fragmented pieces of information rather than a connected narrative. In turn, there were also opposite cases where a profile category had few terms compared to higher profile communicability (e.g., reports Electron_23 P7 and Electron_2 P7–P9). In these cases, physics terms were used efficiently: although there were few of them, they formed a central part of the text.

5.4 Comparing pre-service teachers' vocabularies reflected as profile category distributions

Next, we inspected how profile categories were distributed in the whole sample and we considered only those profile categories where profile communicability was above average. The double-slit experiment (P9) could be found in 54 out of 60 reports (28 on photon, 26 on electron) and it was clearly the most prominent profile category. Thus, we focus now on the rest of the categories: classical and modern physics P1–P8 (see Tables 3 and 4, respectively).

Table 3. Classical physics' profile categories in the reports.

Classical physics (P1–P4)	**The number of reports where profile communicability is above average**		
	Photon	Electron	In total
None	4	11	15
Classical field and radiation (P1)	14	8	22
Classical energy, intensity (P2)	6	2	8
Classical wave model (P3)	19	12	31
Classical particle model (P4)	4	12	16
Neither P3 nor P4	9	12	21
Only classical wave model (P3)	17	6	23
Only classical particle model (P4)	2	6	8
Both classical wave and particle model (P3 and P4)	2	6	8

Table 4. Modern physics' profile categories in the reports.

Modern physics (P5–P8)	**The number of reports where profile communicability is above average**		
	Photon	Electron	In total
None	6	2	8
Quantum mechanics (P5)	20	24	44
Stochastics (P6)	2	3	5
Duality (P7)	16	22	38
Localization and identification (P8)	7	15	22

After double-slit experiment (P9), the most prominent profile categories were quantum mechanics (P5, in 44 reports), duality (P7, in 38 reports) and classical wave model (P3, in 31 reports). Profile categories classical field and radiation (P1) and localization and identification (P8) were both found in 22 reports. The least prominent categories were stochastics (P6), classical energy and intensity (P2) and classical particle model (P4) found in only a few reports.

The profile communicability distributions showed similarities between photon and electron: for both, the strongest categories included quantum mechanics (P5) and dualism (P7), while classical energy and intensity (P2) and stochastics (P6) were the weakest. This showed that, as expected, the reports may view photons and electrons symmetrically, but only on some coarse level. Apart from classical particle model (P4), reports on photon emphasized classical physics' categories more than reports on electron and vice versa when it came to modern physics (P5–P8) and classical particle model (P4). Reports on photon relied heavily on vocabulary on classical particle model (P3) ignoring classical particle model (P4) in most cases whereas reports on electron used both equally. Still, about one third of the reports (21 in total) had both classical wave and particle model below average.

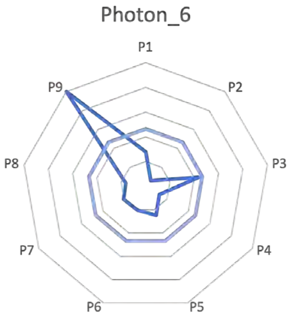
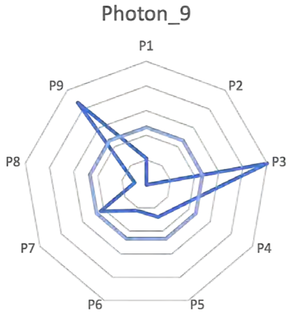
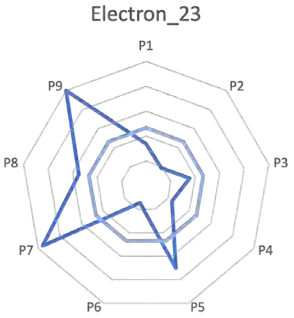
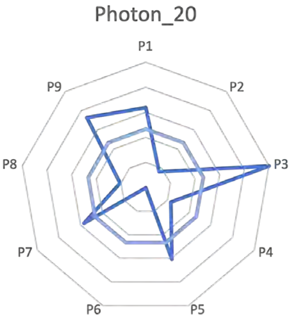
In the case of photons, the profile categories were more clearly divisible into strongest and weakest. In the case of electrons, a similar division would have been more ambiguous. This difference showed that the reports' viewpoints of electrons differed from each other more than those of photons, and that as a group, they described electrons from more diverse perspectives than they did photons.

5.5 Similarities of pre-service teachers' vocabulary reflected by thematic profile categories

We utilized nine-dimensional profile communicability values to study similarities between vocabularies expressed in reports. We analyzed reports pairwise by comparing their profile categories with each other. This very detailed analysis showed that the vocabularies expressed in the reports did not share much similarity: the vocabularies describing photon and electron were scattered and they emphasized different combinations of profile categories.

To produce a more representational and practical classification, we reduced the profile categories into three main classes: *classical physics* (P1–P4 including classical field, energy, wave model and particle model), *modern physics* (P5–P8 including quantum mechanics, stochastics, duality, and localization) and the *double-slit experiment* (P9). In what follows, we considered only those profile categories where profile communicability was above average. The classification is shown in Table 5. Most reports (37 out of 60) had profile categories above average in all three classes and these are shown more closely in Table 6. The whole sample is presented in Appendix D.

Table 5. Classification of the reports based on the representation of the three classes: classical physics, modern physics, and the double-slit experiment. The leftmost column tells which of the three classes have at least one profile category above average 0.47 in profile communicability. These sets are mutually exclusive. The middle column presents the profile communicability of an example report with the average marked in lighter shade. The two rightmost columns tell the number of reports in this class combination.

Represented classes	An example of profile communicability	Reports on photon	Reports on electron
Experiment (P9)		2	0
Experiment (P9) and classical physics (P1–P4)		4	2
Experiment (P9) and modern physics (P5–P8)		2	7
Experiment (P9), classical (P1–P4) and modern physics (P5–P8)		20	17
Classical physics (P1–P4)		0	0

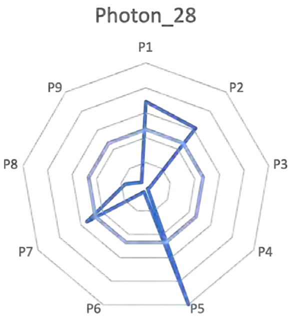
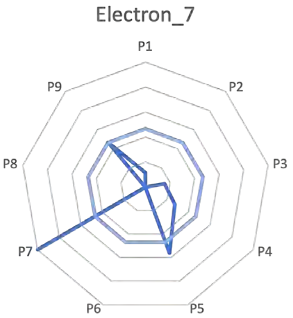
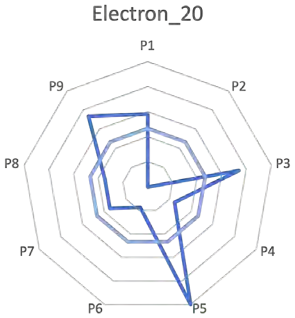
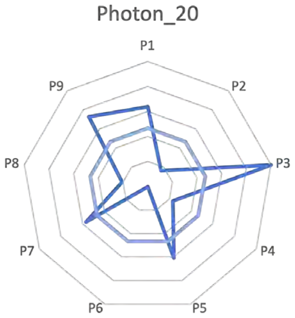
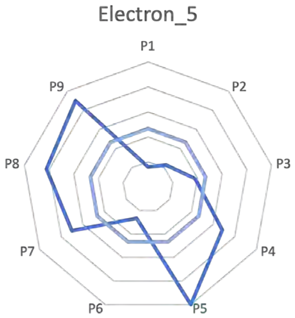
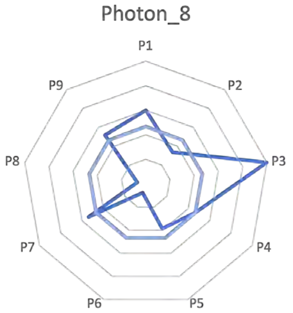
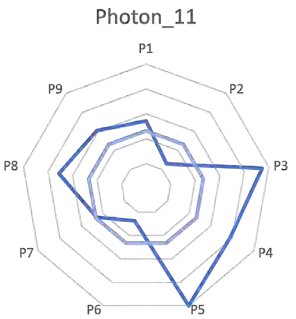
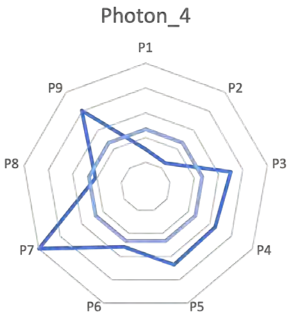
Classical (P1–P4) and modern physics (P5–P8)		2	0
Modern physics (P5–P8)		0	4

Table 6. The reports that contain classical physics, modern physics, and the double-slit experiment. This biggest set (20 reports on photon, 17 on electron) is further divided into subsets based on their use of modern physics (see the leftmost column). The middle column presents the profile communicability of an example report with the average 0.47 marked in lighter shade. The two rightmost columns tell the number of reports in this class combination.

Represented classes	An example of profile communicability	Reports on photon	Reports on electron
Classical physics (P1–P4), quantum mechanics (P5) and experiment (P9)		5	2
Classical physics (P1–P4), quantum mechanics (P5), dualism (P7) and experiment (P9)		5	5
Classical physics (P1–P4), quantum mechanics (P5), dualism (P7), localization (P8) and experiment (P9)		3	7
Classical physics (P1–P4), dualism (P7) and experiment (P9)		2	1

Classical physics (P1–P4), quantum mechanics (P5), localization (P8) and experiment (P9)		2	1
Other modern physics' combinations		3	1

Altogether only nine pre-service teachers had similar profile classification in their both reports (as described in Tables 3 and 4). If both reports belong in the same set, it means that the reports' terminology share similarities in describing photons and electrons. For example, pre-service teacher number 23 used mainly vocabulary on modern physics and double-slit experiment in their reports (see charts for Photon_23 and Electron_23 in Appendix D) while pre-service teacher number 27's vocabulary focused on classical physics, quantum mechanics and double-slit experiment (see charts for Photon_27 and Electron_27 in Appendix D). To sum up, one third of the pre-service teachers used somewhat symmetrical vocabulary describing photon and electron, but this expected symmetry was lacking in other (N=21) pre-service teachers' reports.

The sample contained not a single report that had all nine profile categories above the average (see Tables 3 and 4, and Appendix D). On the contrary, it was usual that a report had several categories below average and even with the minimum value zero. There were five reports where the only category above average was due to the normalization. The most usual number of profile categories above average was four (in 14 reports) and the most comprehensive reports had six categories above average (13 reports). It is to be noticed that the way the data was normalized would suggest that the values were focused on the higher side, since maximum value one could be found in every profile communicability vector and there was no such precondition to minimum value zero. This meant that the method allows reports with high values in nine profile categories, but they did not appear in the sample. At least three profile categories were below average in every report and the use of the vocabularies was not as comprehensive as expected. There was also a difference between reports on photon and electron: there were notably more many-sided portrayals of electron than of photon. Half of reports on electron had five or six profile categories above average while only a third of reports on photon were this comprehensive.

6 Discussion and conclusions

In this study, we investigated pre-service physics teachers' vocabularies from written reports about quantum phenomena. From the vocabularies, we inspected physics terms and categorized them into thematic profile categories P1–P9. Then we studied the interrelations and connectedness of terms, which in this case was measured by communicability centrality utilizing a lexical network approach. We assumed that the vocabularies studied here would be comprehensive and share similarities due to the task design.

The analysis revealed that pre-service physics teachers' vocabularies did not share much similarity. The number of physics terms per report and their relative share of all nouns varied greatly between the reports. A typical report reflected a narrow image of photons and electrons: they considered only some of the thematic profile categories, with the rest of the categories weak or missing. The expected similarity between the vocabularies of electrons and photons was found only on a very coarse level and only in a third of reports.

The lexical network analysis was applied to study how deeply different terms were connected. In most reports, as expected, we could find correlation between profile weight (i.e., number of physics terms used) and profile communicability (the centrality of the thematic profile category). However, we could not find any substantial difference between sentence level and deeper context level, indicating that many expressions were statement-like and semantically shallow, i.e., meaning was not deepened (which would require sentence structures that are more complicated than simple statements, compare with Koponen & Nousiainen, 2020; Nousiainen & Koponen, 2020). We found reports in which high term count was combined with low connection to other terms, which reflected the low internal coherence of the texts, i.e., clauses were not well related. In turn, there were also reports in which a few terms were used effectively to make connections between terms. The latter result suggests that language in these reports shared similar characteristics with the language of science, as Fang (2006) describes. The number of physics terms or word lists do not reveal whether the terms are used sensibly or effectively. This lexical network analysis of physics terms allows us to see how the terms are linked together in the reports and what kind of physics content knowledge they could be able to communicate.

The thematic profile categories were established to identify and analyze how pre-service physics teachers used the terminology related to quantum physics and how they made terminological connections within and between categories. The appearance of profile categories varied between reports. We could find distinct categories that were typically prominent and others that were almost ignored. Categories relating to quantum mechanics (P5) and double-slit experiment (P9) were most prominent for both reports on photons and electrons. Classical energy (P2) and stochastics (P6) categories were the weakest for both. This coarse comparison of the entire data showed that we could identify some of the expected symmetry between the vocabularies on electron and photon. When we looked closer, we saw many differences between the thematic categories. In this sample, the reports on electrons were more term-diverse than those on photons, and there were more many-sided reports (more profile categories represented) on electrons than photons, and we could seldom recognize the terminological symmetry between photon and electron.

For photons and electrons, double-slit experiment (P9) and quantum physics (P5) were the most prominent categories. We can only ponder the reason for this result: it could have been easier for the pre-service teachers to consider a concrete experiment than its more abstract interpretations, or they could have connected the concrete experiment strongly with its results and interpretations. The quantum mechanics category (P5) consisted of many terms referring directly to photon or electron (e.g., photon model, electron field), so it presumably stood out. The quantum mechanics viewpoint could be found in every report, but it was more central in the reports handling electrons than photons. Maybe the observations that were explained by the classical wave model in the case of photons were preferably explained directly with a quantum field model in the case of electrons. Reports on electrons generally used fewer classical physics vocabulary than reports on photons. Still, compared to this, it was contradictory that duality (based on classical wave and particle models) was a more notable viewpoint for electrons than photons.

In the case of photons, classical view of light as waves could be tracked, as the classical wave model (P3) was emphasized in the photon reports. However, classical particle model (P4) did not appear as much with electrons. Moreover, electrons were referred more to classical waves than to the classical particle model. In the case of photons, the terms referring to classical particle model (P4) and localization or identification (P8) were not emphasized. Both these categories link to a particle view that was classically in disharmony with the wave model of light. However, this result was partly unexpected because the report was about photons (a light quantum or light particle) and observed single photon hits on the screen could be interpreted as justification for a particle model for light.

The least discussed viewpoints in the reports for both photon and electron were classical energy (P2) and stochastics (P6). The small number of terms referring to classical energy might be because energy was mostly referred to in the context of quantization and this terminology here belongs to quantum mechanics (category P5). In contrast, the low value of stochastics (category P6) reflected the low incidence of this viewpoint. Randomness and probability distributions associated with the double-slit experiment were usually mentioned as a side note and absent altogether in 11 reports (out of total N=60). Perhaps pre-service teachers considered the stochastics viewpoint as an implicit part of quantum mechanics or that the observed distribution of photons or electrons on the screen in the experiment did not need to be explicitly interpreted with the help of stochastics to justify the quantum mechanical interpretation. Such findings of pre-service teachers’ expressions reflect incoherent use of the language of science.

The reports differed from each other in how they utilized and emphasized certain categories. Even if not all these nine thematic profile categories were equally integral parts of an individual report, our interpretation was that the more categories were well-presented, the more multidimensional a view of photons or electrons the report expressed (cf. connectedness of terms found in expert’s texts on wave-particle dualism in Koponen & Nousiainen, 2020). Pre-service teachers’ depictions of photons and electrons were varied and different from each other, and often, limited, and incoherent. These findings were consistent with our previous study (Nousiainen & Koponen, 2020), which also increases the reliability of the method. The previous study suggests that pre-service teachers’ vocabularies differ from each other, but the vocabularies were even less similar than anticipated in this larger sample. Pre-service teachers use limited vocabularies that intersect less than expected: when two pre-service teachers write about photons or electrons being particles, one might refer to purely classical and the other to quantum mechanical particles.

With limited vocabularies, pre-service teachers can only express simple or one-sided explanations, which are certainly not a hoped-for result of teacher education. One objective in designing such vocabulary tasks was to foster pre-service physics teachers' abilities to use and master the language of science as part of their future teaching at schools. The results show, however, that only a fraction of the pre-service teachers had the multifaceted vocabulary needed to describe quantum phenomena. The results guide attention to demands of more explicit teaching of the language of science to the pre-service teachers. Teachers can be seen as interpreters between science and science learners (students), and therefore pre- and in-service teachers would be better off knowing how to use the language and its vocabulary fluently and comprehensively. Teacher educators need tools for more language-sensitive science education. One possible solution to scaffold pre-service teachers in using the language of science is to encourage them systematically to active reading, talking, and writing science. We suggest that utilizing such complex but structured science writing tasks throughout university studies might help pre-service teachers to build up their science vocabulary. A further research problem is to find out to what extent the science vocabulary of pre-service teachers can develop during their university studies.

References

Ayene, M., Krick, J., Damitie, B., Ingerman, A., & Thacker, B. (2019). A Holistic Picture of Physics Student Conceptions of Energy Quantization, the Photon Concept, and Light Quanta Interference. *International Journal of Science and Mathematics Education*, *17*(6), 1049–1070. https://doi.org/10.1007/s10763-018-9906-y

Bratkovich, M. O. (2018). Shining Light on Language for, in, and as Science Content. *Science & Education*, *27*(7), 769–782. https://doi.org/10.1007/s11191-018-9998-3

Cheong, Y. W., & Song, J. (2014). Different Levels of the Meaning of Wave-Particle Duality and a Suspensive Perspective on the Interpretation of Quantum Theory. *Science & Education*, *23*(5), 1011–1030. https://doi.org/10.1007/s11191-013-9633-2

Estrada, E. (2012). *The Structure of Complex Networks: Theory and Applications*. OUP Oxford.

Fang, Z. (2006). The Language Demands of Science Reading in Middle School. *International Journal of Science Education*, *28*(5), 491–520. https://doi.org/10.1080/09500690500339092

Glynn, S. M., & Muth, K. D. (1994). Reading and writing to learn science: Achieving scientific literacy. *Journal of Research in Science Teaching*, *31*(9), 1057–1073. https://doi.org/10.1002/tea.3660310915

Halldén, O., Haglund, L., & Strömdahl, H. (2007). Conceptions and Contexts: On the Interpretation of Interview and Observational Data. *Educational Psychologist*, *42*(1), 25–40. https://doi.org/10.1080/00461520709336916

Hobson, A. (2005). Electrons as field quanta: A better way to teach quantum physics in introductory general physics courses. *American Journal of Physics*, *73*(7), 630–634. https://doi.org/10.1119/1.1900097

Keys, C. W., Hand, B., Prain, V., & Collins, S. (1999). Using the Science Writing Heuristic as a Tool for Learning from Laboratory Investigations in Secondary Science. *Journal of Research in Science Teaching*, *36*(10), 1065–1084. https://doi.org/10.1002/(SICI)1098-2736(199912)36:10<1065::AID-TEA2>3.0.CO;2-I

Koponen, I. T., & Nousiainen, M. (2020). Lexical Networks and Lexicon Profiles in Didactical Texts for Science Education. In H. Cherifi, S. Gaito, J. F. Mendes, E. Moro, & L. M. Rocha (Eds.), *Complex Networks and Their Applications VIII* (Vol. 882, pp. 15–27). Springer International Publishing. https://doi.org/10.1007/978-3-030-36683-4_2

Krijtenburg-Lewerissa, K., Pol, H. J., Brinkman, A., & van Joolingen, W. R. (2017). Insights into teaching quantum mechanics in secondary and lower undergraduate education. *Physical Review Physics Education Research*, *13*(1), 010109. https://doi.org/10.1103/PhysRevPhysEducRes.13.010109

Lachner, A., Jacob, L., & Hoogerheide, V. (2021). Learning by writing explanations: Is explaining to a fictitious student more effective than self-explaining? *Learning and Instruction*, *74*, 101438. https://doi.org/10.1016/j.learninstruc.2020.101438

Lemke, J. L. (1990). *Talking Science: Language, Learning, and Values*. Ablex Publishing Corporation, 355 Chestnut Street, Norwood, NJ 07648 (hardback: ISBN-0-89391-565-3; paperback: ISBN-0-89391-566-1). https://eric.ed.gov/?id=ED362379

Mäntylä, T., & Nousiainen, M. (2014). Consolidating Pre-service Physics Teachers' Subject Matter Knowledge Using Didactical Reconstructions. *Science & Education*, *23*(8), 1583–1604. https://doi.org/10.1007/s11191-013-9657-7

Mercer, N. (2009). Developing Argumentation: Lessons Learned in the Primary School. In N. Muller Mirza & A.-N. Perret-Clermont (Eds.), *Argumentation and Education: Theoretical Foundations and Practices* (pp. 177–194). Springer US. https://doi.org/10.1007/978-0-387-98125-3_7

Nousiainen, M. (2017). Organization of physics content knowledge for teaching purposes: From knowledge justification schemes to didactical schemes. *European Journal of Science and Mathematics Education*, *5*(2), 210–221. https://doi.org/10.30935/scimath/9507

Nousiainen, M., & Koponen, I. T. (2020). Pre-Service Teachers' Declarative Knowledge of Wave-Particle Dualism of Electrons and Photons: Finding Lexicons by Using Network Analysis. *Education Sciences*, *10*(3), 76. https://doi.org/10.3390/educsci10030076

Schleppegrell, M. J. (2004). *The Language of Schooling: A Functional Linguistics Perspective*. Routledge. https://doi.org/10.4324/9781410610317

Schwarz, B. B. (2009). Argumentation and Learning. In N. Muller Mirza & A.-N. Perret-Clermont (Eds.), *Argumentation and Education: Theoretical Foundations and Practices* (pp. 91–126). Springer US. https://doi.org/10.1007/978-0-387-98125-3_4

Södervik, I., Nousiainen, M., & Koponen, Ismo. T. (2021). First-Year Life Science Students' Understanding of the Role of Plants in the Ecosystem—A Concept Network Analysis. *Education Sciences*, *11*(8), 369. https://doi.org/10.3390/educsci11080369

Stenhouse, D. (1986). Conceptual Change in Science Education: Paradigms and Language-Games. *Science Education*, *70*(4), 413–425.

Wellington, J. J., & Osborne, J. (2001). *Language and literacy in science education*. Open University.

Yun, E. (2020). Language gap between college introductory physics textbooks and high school physics textbooks. *European Journal of Physics*, *41*(6), 065704. https://doi.org/10.1088/1361-6404/abbaae

Yun, E., & Park, Y. (2018). Extraction of scientific semantic networks from science textbooks and comparison with science teachers' spoken language by text network analysis. *International*

Journal of Science Education, *40*(17), 2118–2136. https://doi.org/10.1080/09500693.2018.1521536

Appendix A Assignment

The reports which we analyzed were a part of a broader course task. Pre-service teachers were asked to prepare a concept map reflecting physics knowledge construction and a report (named appendix in the assignment) for more elaborate explanations. This task was repeated in several physics topics: in this course, the topics were photoelectric effect, Compton scattering, double-slit experiment for photons and double-slit experiment for electrons. Students' base material for the latter two topics included Hobson's (2005) article on the quantum field interpretation in the double-slit experiments. The assignment was given in Finnish and translated for the reader of this article.

Assignment for the pre-service teachers

A didactical reconstruction is a knowledge structure that a teacher creates for themselves. It is a frame of reference for planning their teaching. While using a knowledge justification scheme [previous course task] can support working on the didactical reconstruction, it is important to organize and interpret the included content in a way that is suitable for teaching.

As a result, the didactical reconstruction should be a justified and consistent presentation of how the subject matter is justified, which are its most central experiments and laws, and how these are introduced in teaching.

A didactical reconstruction is commonly represented in the form of a graph, which illustrates the relationships between physics concepts. These graphs serve to organize one's understanding of physics knowledge structure–connections between physics concepts and how physics concepts can be formed. The graphs help to form an overall framework for physics knowledge and its structure. The goal is to organize the course content in such a way that new laws or concepts are built upon previous ones through either experiments or theoretical modelling.

A didactical reconstruction shows how the central quantities of a subject are connected to form a cohesive structure. Pay attention to the order of concept formation and the dependencies between different concepts. The reconstruction is a network graph that illustrates the relationships between nodes, which represent physics quantities, laws, experiments and models. These nodes are connected by directional links that reflect the order of progression in the formation of concepts. The graph should show the path and direction of concept formation. Number the links in running order to reflect the progress of the structure.

The nodes in the network belong to one of five concept categories: quantities, laws, principles, experiments, and models. The links connecting the nodes show how the concepts evolve and how they are justified. There is no limit to the number of nodes and links in the graph, but every node must belong to one the five categories.

Overall, the didactical reconstruction includes two parts:

1. Graph
The nodes are categorized by differentiating edge lines (such as squares or ovals). The connections between the nodes are marked with directional links, which are numbered in running order to reflect the order in which the structure is formed.

2. Appendix
The appendix includes more detailed explanations about the content.

Experiments can be quantifying (building or defining new quantities or laws based on measurements) or qualitative (demonstrating the meanings or qualitative dependencies of concepts). Each experiment should be described with its motive or aim, experimental setup, measurements, and resulting dependence.

Models can be categorized as theoretical, data, or explanatory models. For each model, its motive or aim, and how it connects or builds concepts should be described.

The purpose of the appendix is to provide more detailed information that complements the graph. The graph itself should offer a quick overview of the subject matter and the formation and connections between its concepts. When you are finished creating the didactical reconstruction, make sure that it has a sensible structure and that it does not contain any unjustified (theoretical) knowledge.

Appendix B Contexts

Table B1. The quantum physics contexts utilized in analyzing the use of physics terms as described in greater detail in Appendix C.

Context	Description of related sub contexts
Background for studying the phenomenon	Classical physics Classical models of light and electron Young's interference experiment, electron diffraction Photoelectric effect, Compton scattering (contexts for previous course assignments)
Experimental setup and observations	Double-slit experiment One slit closed or particle detection at the slits Both slits open Single hits at the screen
Interpretations and explanations	Qualitative theoretical interpretation of single hits Photon or electron as non-classical particles Rhetorical questions on what happens at the slits Explanations of what happens at the slits Photon or electron localizing on the screen Uncertainty principle Probability interpretation, wave function Wave-particle duality Quantum field interpretation
Not physics related utterances	

Table B2. Example sentences from the reports, their categorization into contexts (see table B1) and identification of relevant physics terms that belong to profile categories P1–P9 (see table 1). The reports were written, and their analysis was done in Finnish. The English translations are done for the readers of this article. Established physics terms were interpreted as one term even if they consisted of several words (e.g., light quantum). Thus, the original Finnish terms and their English translations match well although the two languages are very different.

Context	Sub context	Authentic text sample, physics terms bolded	English translation, physics terms bolded
Background for studying the phenomenon	Classical physics	Klassinen fysiikka perustuu kahteen erilaiseen oliotyyppiin: **hiukkasiin** ja **kenttiin**.	In classical physics, there are two distinct types of objects: **particles** and **fields**.
	Classical models of light and electron	**Elektronia** on pidetty aiemmin **hiukkasena**, koska sillä on havaittu olevan **massa**.	Previously, **electrons** were considered to be **particles** due to their observed **mass**.
	Young's interference experiment or electron diffraction	Young esitti jo vuonna 1802 **kokeen**, jolla pyrki selvittämään **valon** koostumuksen, ts. muodostuuko se **hiukkasista** vai **aalloista**.	Already in 1802, Young presented an **experiment**, with which he attempted to resolve the composition of **light**, that is to say, whether it is made of **particles** or **waves**.
Experimental setup and observations	Double-slit experiment	**Kokeessa** ammutaan **intensiteetiltään** heikko **elektronisuihku kaksoisraon** läpi **pintaan**, jossa **osumat** rekisteröidään.	In the **experiment**, a weak **electron beam** is shot through a **double-slit** onto a **surface**, where the resulting **hits** are detected.
	One slit closed or particle detection at the slits	**Kokeessa** toinen **rako** on kerrallaan kiinni, jolloin **interferenssikuviota** ei muodostu.	During the **experiment**, when the **slits** are closed one at a time, the **interference pattern** does not appear.
	Both slits open	Kun molemmat **raot** ovat auki (ei **ilmaisimia raoissa**), muodostuu **interferenssikuvio**.	When both **slits** are open (no **detectors** at the **slits**), an **interference pattern** emerges.
	Single hits at the screen	**Interferenssikuvio** muodostuu **yksittäisistä osumista pinnalle**.	The **interference pattern** consists of **individual hits** on the **surface**.
Interpretations and explanations	Qualitative theoretical interpretation of single hits	Jos **elektroni** olisi **hiukkanen**, niin tällöin **yksittäiset paikallistuneet**	If the **electron** were a **particle**, the **individual localized hits** would be easily explained.

		osumat olisivat helposti selitettävissä.	
	Explanations of what happens at the slits	Tällöin **elektroni** tulee **aaltona** kummankin **raon** läpi.	Then the **electron** passes through both **slits** as a **wave**.
	Photon or electron localizing on the screen	Jos **elektroni** tulkitaan **kentäksi**, voidaan **yksittäiset osumat** selittää sillä, että ne kertovat vain **elektronien** paikallisista **vuorovaikutustilanteista pinnan atomien** kanssa.	If the **electron** were interpreted as a **field**, the **individual hits** could be explained as evidence of the **electrons' local interactions** with the **atoms** on the **surface**.
	Probability interpretation, wave function	**Interferenssikuvion** muodostumista voidaan kuvata **aaltofunktion** avulla, joka kuvaa **elektronin** esiintymisen **todennäköisyysjakaumaa**.	The emerging of the **interference pattern** can be described with a **wave function**, which describes the **probability distribution** of the **electron**'s appearance.
	Wave-particle duality	**Yksittäiset pisteet** osoittavat **elektronin hiukkasluonteen**, kun taas **interferenssikuvio** osoittaa **elektronin aaltoluonteen**.	The **single spots** indicate the **particle nature** of the **electron** whereas the **interference pattern** indicates the **wave nature** of the **electron**.
	Quantum field interpretation	**Fotoni** ei olekaan **hiukkanen** klassisessa mielessä, vaan pikemminkin **laaja-alainen yksilöitymätön sähkömagneettisen kentän energiakvantti**, joka osoittaa **hiukkasluonteensa** (ts. ilmenee ja yksilöityy) ainoastaan **vuorovaikutustilanteissa**.	As it turns out, the **photon** is not a **classical particle**, but rather a **widespread unidentified energy quantum** of the **electromagnetic field**, which exhibits its **particle nature** (i.e., appears and can be identified) only in **interactions**.

Appendix C Method

The lexical networks utilized here are based on how terms occur in text. We look at the occurrence on different levels, beginning from the syntactic level (the level of clauses) and proceeding to the semantic level (the level of contexts). These connections form a stratified lexical network of the lexical distance of terms in different syntax levels. A more comprehensive description is given in Koponen & Nousiainen (2020) and Nousiainen & Koponen (2020). The method quantifies the connectivity between terms in the lexical network. The method is based on a measure called communicability centrality (Estrada, 2012), which focuses on different contiguous paths found between nodes (terms) in the lexical network. The paths can be weighted according to their lengths. Counting such paths measures lexical proximity. This measure tells us how information can be passed from one node to another through the network.

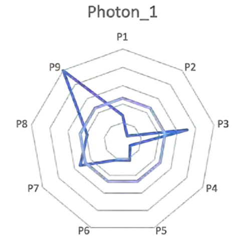
A lexical network that has N nodes can have (at most) $N \times (N - 1)$ different links between the nodes (terms). We describe such a network using adjacency matrix **A,** which elements $[\mathbf{A}]_{pq} = a_{pq}$ have value 1 if there is a connection between nodes p and q, and value 0 if the nodes are not connected. Adjacency matrix **A** can be used to calculate the number of different paths between two nodes in the lexical network. In a well-connected network, the number of long paths increases almost factorially. We are thus interested in the relative weight of these paths and the solution is to divide the number of walks by the factorial. This is called the communicability measure (Estrada, 2012)

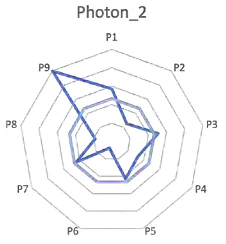
$$G_{pq}(\beta) = \frac{1}{1!}\beta^1[\mathbf{A}^1]_{pq} + \frac{1}{2!}\beta^2[\mathbf{A}^2]_{pq} + \frac{1}{3!}\beta^3[\mathbf{A}^3]_{pq} + \cdots = \left[e^{\beta\mathbf{A}} - \boldsymbol{I}\right]_{pq}$$

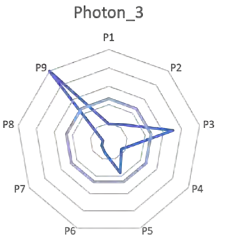
where e[…] is the matrix exponential, **I** the identity matrix and $[\ldots]_{\mathrm{pq}}$ is its element at row p and column q. Note that here a slightly modified version of the standard definition of communicability (Estrada, 2012) is used. The communicability has a free parameter $\beta \geq 1$ that adjusts how wide a part of the network we look at when counting the paths. The optimal value for parameter β offers the best diversity of terms at the lowest possible value of β. The optimal value seems here to be about $\beta = 1$. We can construct a lexical proximity network of key terms, where terms are linked according to their lexical distances. This modified and pruned lexical network is used to inspect only those terms that are connected well enough.

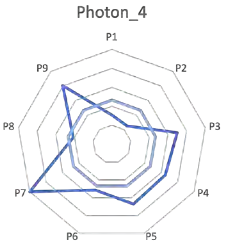
Appendix D Profile communicabilities of the sample

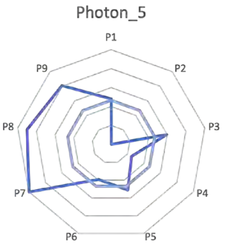
Labelled corners P1 to P9 refer to the profile categories listed in Table 1. The blue line represents profile communicability and lighter line the average 0.47 of all profile communicability values of the sample. The center of the polygon corresponds to the minimum value zero and the values increase linearly to the outermost edge corresponding the maximum value one.

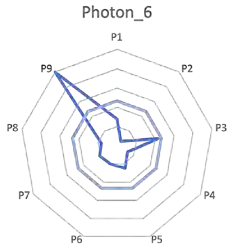


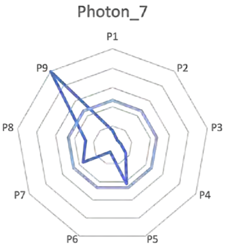


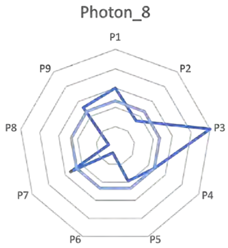


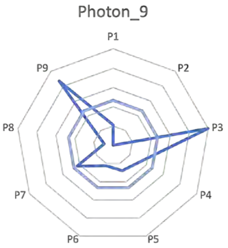


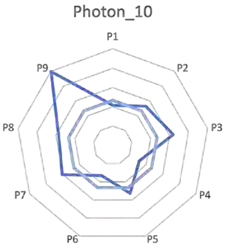


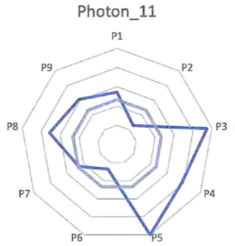


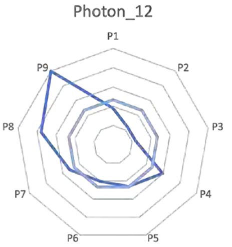


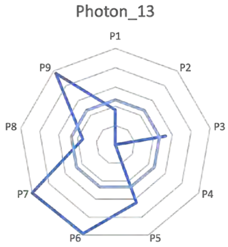


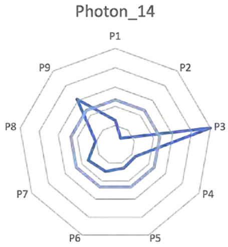


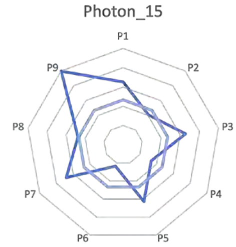


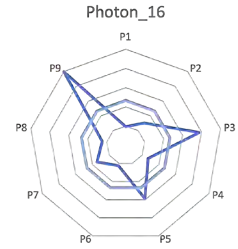


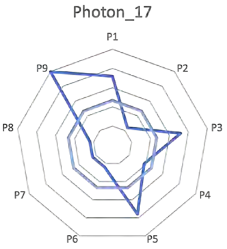


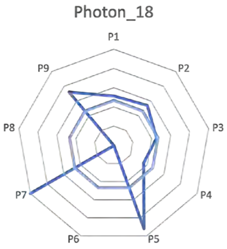


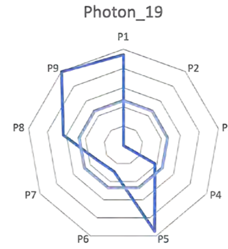


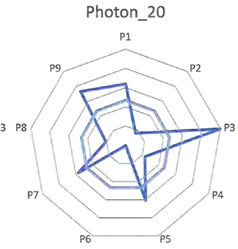


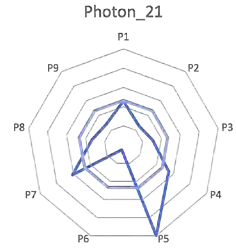


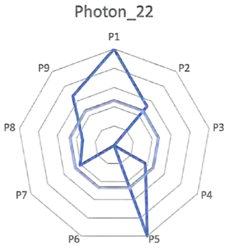


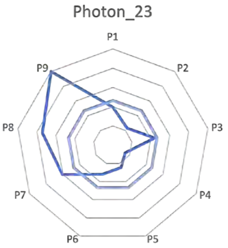


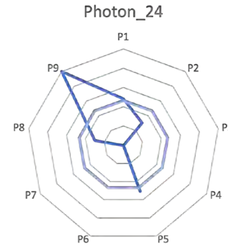


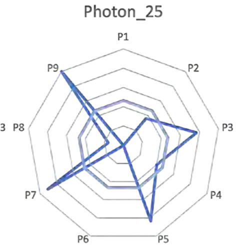


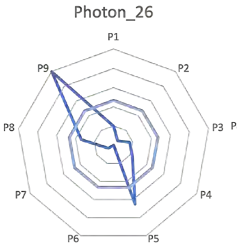


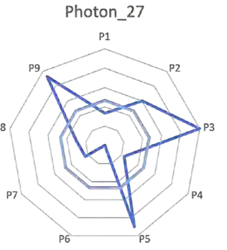


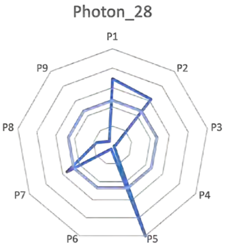


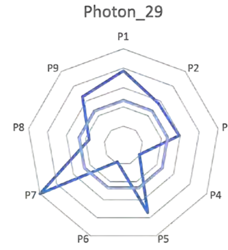


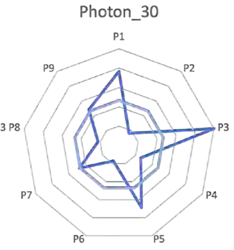


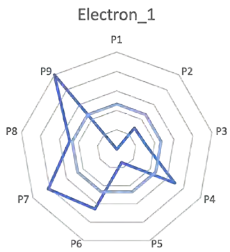


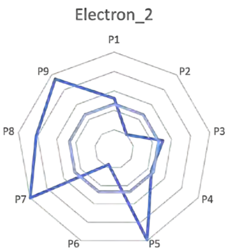


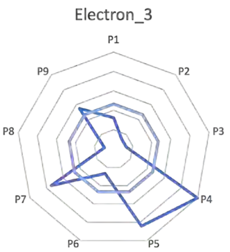


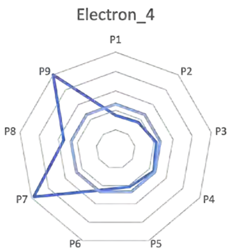


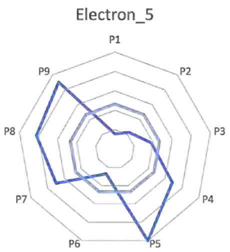


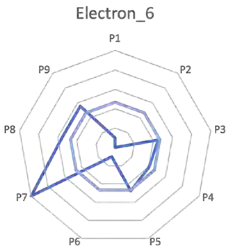

Electron_1
P1
P2
P3
P4
P5
P6
P7
P8
P9

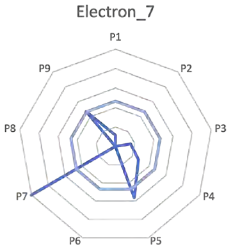

Electron_2
P1
P2
P3
P4
P5
P6
P7
P8
P9

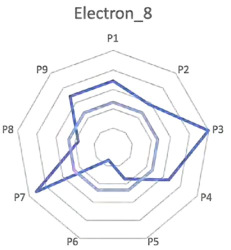

Electron_3
P1
P2
P3
P4
P5
P6
P7
P8
P9

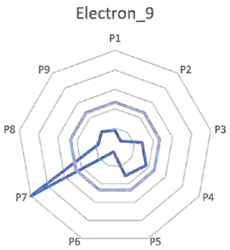

Electron_4
P1
P2
P3
P4
P5
P6
P7
P8
P9

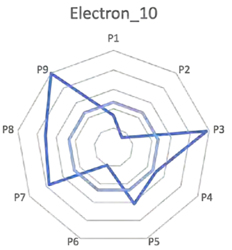

Electron_5
P1
P2
P3
P4
P5
P6
P7
P8
P9

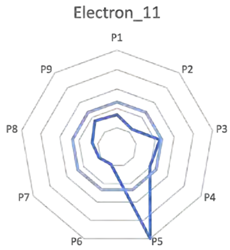

Electron_6
P1
P2
P3
P4
P5
P6
P7
P8
P9

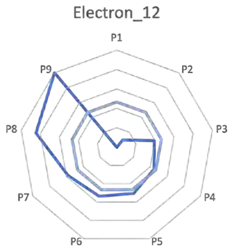

Electron_7
P1
P2
P3
P4
P5
P6
P7
P8
P9

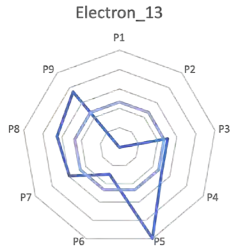

Electron_8
P1
P2
P3
P4
P5
P6
P7
P8
P9

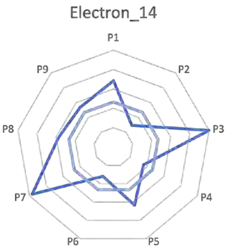

Electron_9
P1
P2
P3
P4
P5
P6
P7
P8
P9

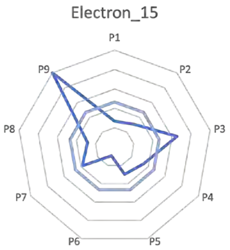

Electron_10
P1
P2
P3
P4
P5
P6
P7
P8
P9

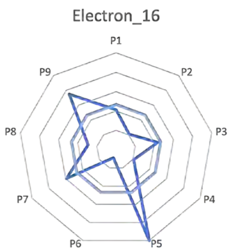

Electron_11
P1
P2
P3
P4
P5
P6
P7
P8
P9

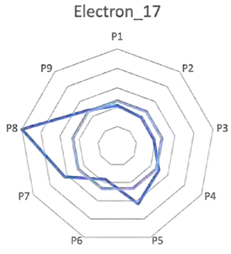

Electron_12
P1
P2
P3
P4
P5
P6
P7
P8
P9

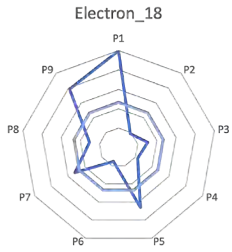

Electron_13
P1
P2
P3
P4
P5
P6
P7
P8
P9

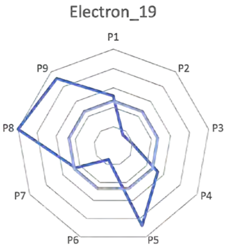

Electron_14
P1
P2
P3
P4
P5
P6
P7
P8
P9

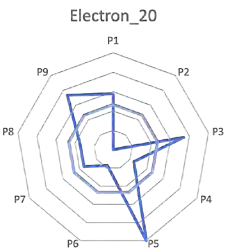

Electron_15
P1
P2
P3
P4
P5
P6
P7
P8
P9

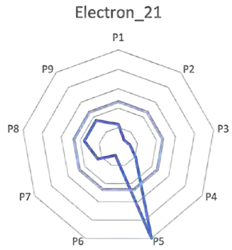

Electron_16
P1
P2
P3
P4
P5
P6
P7
P8
P9

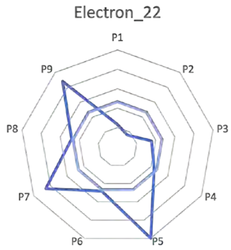

Electron_17
P1
P2
P3
P4
P5
P6
P7
P8
P9

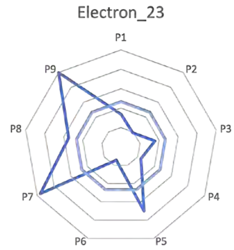

Electron_18
P1
P2
P3
P4
P5
P6
P7
P8
P9

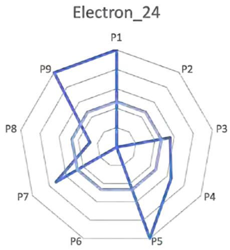

Electron_19
P1
P2
P3
P4
P5
P6
P7
P8
P9

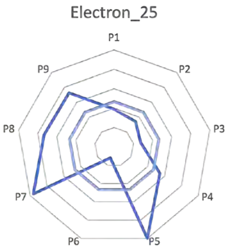

Electron_20
P1
P2
P3
P4
P5
P6
P7
P8
P9

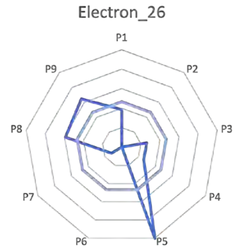

Electron_21
P1
P2
P3
P4
P5
P6
P7
P8
P9

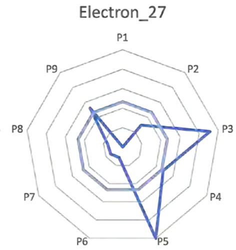

Electron_22
P1
P2
P3
P4
P5
P6
P7
P8
P9

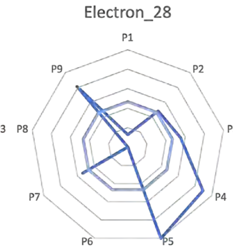

Electron_23
P1
P2
P3
P4
P5
P6
P7
P8
P9

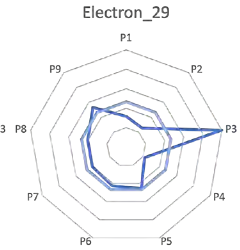

Electron_24
P1
P2
P3
P4
P5
P6
P7
P8
P9

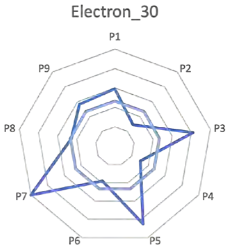

Electron_25
P1
P2
P3
P4
P5
P6
P7
P8
P9


Electron_26
P1
P2
P3
P4
P5
P6
P7
P8
P9


Electron_27
P1
P2
P3
P4
P5
P6
P7
P8
P9


Electron_28
P1
P2
P3
P4
P5
P6
P7
P8
P9


Electron_29
P1
P2
P3
P4
P5
P6
P7
P8
P9


Electron_30
P1
P2
P3
P4
P5
P6
P7
P8
P9