U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

PROFILES – WP3:
Stakeholders Involvement and Interaction

PROFILES

Curricular Delphi Study on Science Education

Interim Report on the Third Round of the UP Working Group

João Paiva, Carla Morais, José Barros

Department of Chemistry and Biochemistry,
Faculty of Sciences, University of Porto, Portugal

April 2013

PROFILES

Table of Contents

List of Tables

List of Figures

1 Introduction

The main focus of the "International PROFILES Curricular Delphi Study on Science Education" is to involve different stakeholders as experts in reflecting on contents and aims of science education. In particular, the PROFILES Curricular Delphi Study on Science Education seeks to collect in three consecutive rounds different stakeholders' views and opinions about those aspects of scientific literacy that they consider relevant and pedagogically desirable for the individual in the society of today and in the near future (see Figure 1) (Bolte & Schulte, 2012; Schulte & Bolte, 2012).

The outcomes of this study will serve the development of innovative learning environments (WP4) and the preparation of continuous teacher training courses (WP5) "aiding the implementation and dissemination of PROFILES ideas, intentions and objectives to facilitate the uptake of innovative science teaching and the enhancement of scientific literacy" (PROFILES Consortium, 2010, p. 20).

The stakeholder sample was specified with four groups related to sciences and science education: students with basic or advanced science courses, science teachers (education students, trainee teachers, in-service teachers and teacher educators), science education researchers and scientists. As it will be further explained, UP sample is divided in two groups only: students and teachers (see section 4).

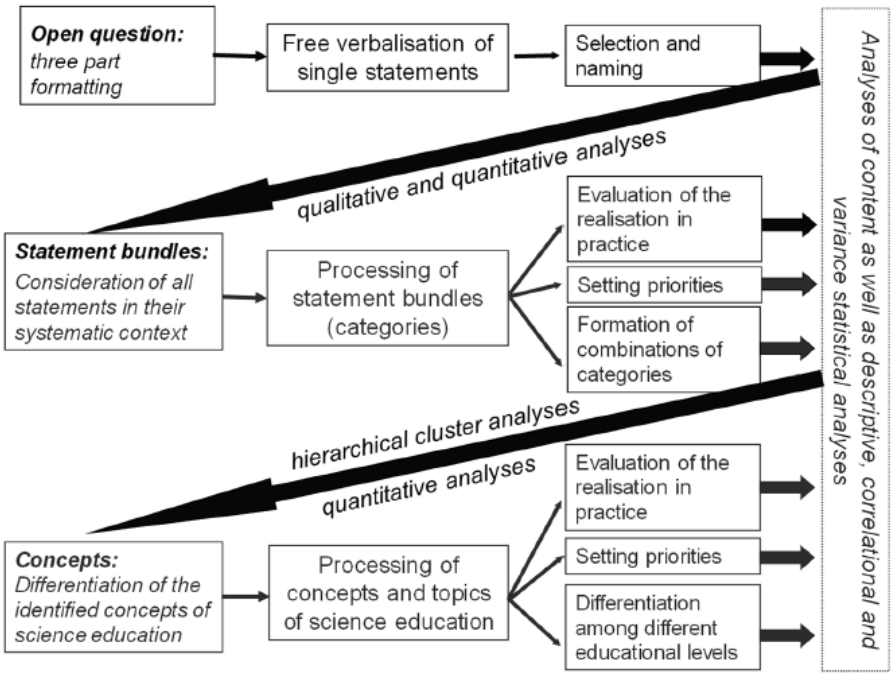


Figure 1: Structure of the PROFILES Curricular Delphi Study on Science Education (Bolte, 2008)

In the first interim report on the UP PROFILES Curricular Delphi Study on Science Education (see also Deliverable 3.1), the framework, aims, structure, concepts and methods of this study as well as the results from the first round of this study were introduced (Paiva, Morais & Barros, 2011).

The second interim report on the UP PROFILES Curricular Delphi Study on Science Education (see also Deliverable 3.2), following up on the first report, took up questions that resulted from the analyses in the first round and introduced the procedure and the outcomes of the second round (Paiva, Morais & Barros, 2012).

As scientific literacy is a complex construct (Bybee, 1997; DeBoer, 2000; Eckebrecht & Schneeweiß, 2003; Gräber & Bolte, 1997; Gräber & Nentwig, 2002), its enhancement is not possible by referring to the different aspects only individually. Following Gräber & Bolte (1997), promoting scientific literacy is only possible if the complexity of the scientific literacy construct is taken account of in content, method and conception. Therefore, the empirically identified aspects of desirable science education from round 1 were in the second part of round 2 considered in relation to content, method and conception. In order to identify empirically sound concepts regarding science education that are considered important, the participants were in part II of round 2 asked to combine from the given set of 80 categories those categories that seem especially important to them in their combination. The combinations which the participants made in part II of the second round questionnaire were analyzed by means of hierarchical cluster analyses, using the Ward method and squared Euclidian distance (Bortz, 2005).

As the UP hierarchical cluster analyses did not yield a coherent outcome, the three concepts of desirable science education of FUB analyses were adopted (Bolte & Schulte, 2012). The concepts were termed as follow:

Concept A: Awareness of the sciences in current, social, globally relevant and occupational contexts relevant in both educational and out-of-school settings.

Concept B: Intellectual education in interdisciplinary scientific contexts.

Concept C: General science-related education and facilitation of interest in contexts of nature, everyday life and living environment.

Table 1 presents a detailed description of each concept.

Concept A: Awareness of the sciences in current, social, globally relevant and occupational contexts relevant in both educational and out-of-school settings *Concept A (Awareness of the sciences in current, social, globally relevant and occupational contexts relevant in both educational and out-of-school settings*) refers to an engagement with the sciences within the frame of current, social, globally relevant, occupational and both educational and out-of-school contexts, enhancing emotional personality development and basic skills. The impressions a person gets through engaging with topics and associated science-related questions from his or her environment influence both the person's sensibility and his or her attitudes towards the sciences. Dealing with scientific issues or phenomena in out-of-school or social and public contexts respectively also facilitates conscious experiences of scientific phenomena, scientifically precise observation and cognitive ability. Moreover, basic and professionally relevant skills such as finding, interpreting and communicating information can be enhanced in this way. Suggestions for this kind of engagement and education are amongst others provided e.g. by current issues or media coverage. Dealing with the history of the sciences especially reveals how findings and methods of the sciences enable, enhance and bring forward research in the natural sciences. This shows moreover how historical science-related developments are still linked to applications in industry and technology, how these applications changed the world and how they influence our professional and everyday life.
Concept B: Intellectual education in interdisciplinary scientific contexts Concept B (*Intellectual education in interdisciplinary scientific contexts)* refers to an engagement with the sciences, their terminology, their methods, their basic concepts, their interdisciplinary relations, their findings and their perspectives, which enhance individual intellectual personality development. Dealing with the sciences serves in this course not only the acquisition of science-related basic knowledge but also helps to understand fundamental findings and the process of gaining knowledge in the sciences. Moreover, dealing with questions and topics of the sciences helps to comprehend and follow (empirical and experimental) scientific research methods, facilitates analytical abilities and fosters the ability to take differentiated perspectives. In addition, an engagement with current scientific research reveals not only how findings and methods of the sciences enable, enhance and support both scientific research and its applications, but also how scientific research is interconnected interdisciplinary.
Concept C: General science-related education and facilitation of interest in contexts of nature, everyday life and living environment *Concept C (General science-related education and facilitation of interest in contexts of nature, everyday life and living environment)* refers to a science-related engagement with everyday life and living environment issues that takes up and promotes students' interests, enhancing general personality development and education. In this way, aspects such as opinion-forming and acting reflectedly and responsibly are particularly important. Dealing with topics from the natural and technological living environment shows how scientific research, scientific applications and scientific phenomena influence both public and personal life. Another important aspect of this concept is engaging with different values and perspectives as well as reflecting on both personal and public deliberations and course of action. Moreover, this concept refers to facilitating the motivation for scientific inquiry beyond school, including aspects such as realizing and shaping one's own interests. Dealing with scientific issues and phenomena within the contexts of social and public fields such as technological developments, their consequences and issues about safety and risks enhances in particular the students' own abilities to judge and both critically reflect and rationally account for their own actions.

Table 1: Concepts description

2 Leading questions of the third round

Following the procedure of curricular Delphi study in general, a central aspect is in the third round of the International PROFILES Curricular Delphi Study on Science Education to address the findings and corresponding questions from previous – here: the second – round (Bolte, 2003, 2008; Häußler, Frey, Hoffmann, Rost, & Spada, 1980; Linstone & Turoff, 1975; Mayer, 1992).

In line with the general and overarching question of the PROFILES Curricular Delphi Study on Science Education[1], the third round focuses especially on the following questions:

1.1 Which priorities regarding concepts of desirable science education can be identified in the participants' assessments?

1.2 To what extent are the respective concepts of desirable science education according to the participants' assessments realized in current science educational practice?

1.3 What kind of priority-practice differences can be identified in the participants' assessments?

2.1 Which priorities regarding concepts of desirable science education can be identified in the participants' assessments with regard to different educational levels?

2.2 To what extent are the respective concepts of desirable science education according to the participants' assessments realized in current science educational practice regarding different educational levels?

2.3 What kind of priority-practice differences can be identified in the participants' assessments regarding the different educational levels?

3 What differences or similarities appear in the general assessments between the two different sub-sample groups?[2]

[1] The Curricular Delphi Study on Science Education addresses the following general question: *What aspects regarding science education are in general considered desirable and pedagogically meaningful for the individual in the society today and in the near future?*

[2] UP sub-samples groups in PROFILES Curricular Delphi Study on Science Education are two, comprehending students and teachers. Trainee-teachers and scientists were collapsed in the teachers sub-sample group, however the International sample include education researchers. Some PROFILES partners also included a group with other stakeholders.

3 Design of the questionnaire and method of data analysis

The concepts of desirable science education (see Table 1) that were identified in the second part of the second round through hierarchical cluster analyses were in the third round reported back to all the participants from the second round for their assessment.

The questionnaire, part III**,** that UP PROFILES used in the Second Round of Delphi Study on Science Education is available online in https://docs.google.com/spreadsheet/viewform?formkey=dDlqdHhWQlB5SmVWcnBmSW9femRWSVE6MQ.

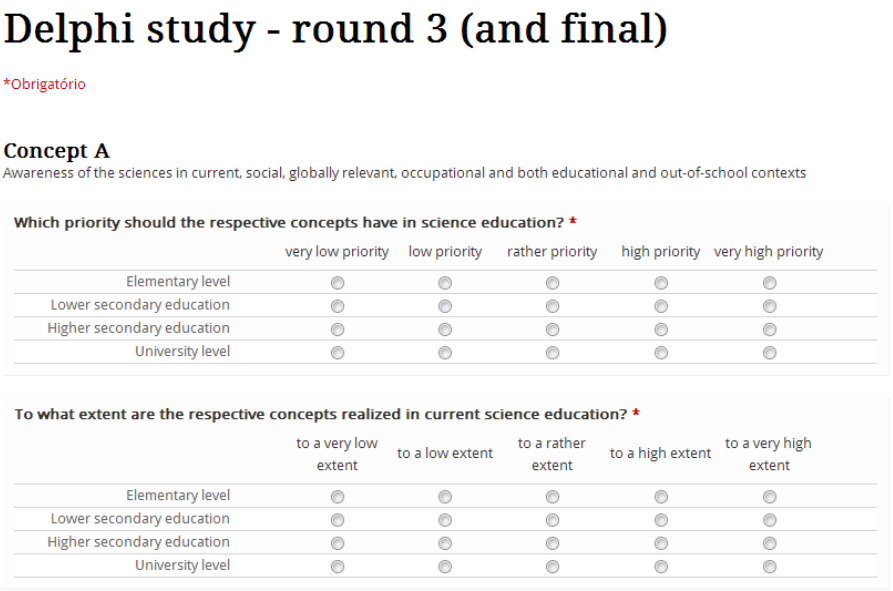
The precise wording of the task and the questions in part III of the questionnaire are shown in Figure 2.

The concepts were to be assessed from two points of view ("priority" and "practice") on a five-tier scale from "very low" to "very high" levels and were also differentiated according to different educational levels (Elementary level, Lower secondary education, Higher secondary education and University).

The two points of view were specified by the following questions:

- Which priority should the respective concepts have in science education?
- To what extent are the respective concepts realized in current science education?

The data were coded with the numbers 1-5 according to the five-tier scale. The data were analyzed by means of descriptive and variance analytical methods.



Delphi study - round 3 (and final)

*Obrigatório

Concept A

Awareness of the sciences in current, social, globally relevant, occupational and both educational and out-of-school contexts

Which priority should the respective concepts have in science education? *

	very low priority	low priority	rather priority	high priority	very high priority
Elementary level					
Lower secondary education					
Higher secondary education					
University level					

To what extent are the respective concepts realized in current science education? *

	to a very low extent	to a low extent	to a rather extent	to a high extent	to a very high extent
Elementary level					
Lower secondary education					
Higher secondary education					
University level					

Figure 2: Design of part III of the UP questionnaire of the third round available online.

4 Data collection and sample of the second round of the UP PROFILES Curricular Delphi Study on Science Education

In this third round of the PROFILES Curricular Delphi Study on Science Education, a total number of 39 participants took part (see Table 2). The sample of participants in the third round includes, according to the Delphi method, only participants who have taken part in the first and second rounds. There was only 1 subject identified as belonging to the group "Scientists" and another belonging to the group "Teachers undergoing education", who were collapsed into the "Teachers" group.

<table>
<tr><th colspan="2" rowspan="2">UP sample sub-groups</th><th colspan="4">Number of responses</th><th colspan="2" rowspan="2">Participation rate</th></tr>
<tr><th colspan="2">Round 1</th><th colspan="2">Round 2</th><th>Round 3</th></tr>
<tr><td colspan="2">Students</td><td colspan="2">20</td><td colspan="2">20</td><td>12</td><td colspan="2">60%</td></tr>
<tr><td rowspan="4">Teachers</td><td>Education Students</td><td>9</td><td rowspan="4">38</td><td>4</td><td rowspan="4">32</td><td>0</td><td>0%</td><td rowspan="4">68%</td></tr>
<tr><td>Trainee teachers</td><td>1</td><td>1</td><td>1</td><td>100%</td></tr>
<tr><td>Science Teachers</td><td>28</td><td>25</td><td>25</td><td>89%</td></tr>
<tr><td>Teacher Educators</td><td>4</td><td>2</td><td>0</td><td>0%</td></tr>
<tr><td colspan="2">Education Researchers</td><td colspan="2">2</td><td colspan="2">1</td><td>0</td><td colspan="2">0%</td></tr>
<tr><td colspan="2">Scientists</td><td colspan="2">2</td><td colspan="2">1</td><td>1</td><td colspan="2">50%</td></tr>
<tr><td colspan="2">People who are not directly involved with sciences</td><td colspan="2">20</td><td colspan="2">0</td><td>0</td><td colspan="2">0%</td></tr>
<tr><td colspan="2">Total</td><td colspan="2">86</td><td colspan="2">54</td><td>39</td><td colspan="2">45%</td></tr>
</table>

Table 2: Sample structure and response rate of the third round of the UP PROFILES Curricular Delphi Study

The group of *teachers* makes up a total of 89,8%, with a number of 27 participants (see Table 3). The group of *students* consists of 12 participants altogether (30,8%).

UP sample sub-groups	Total N	Percent
Students	12	30,8%
Teachers	27	69,2%
Total	39	100%

Table 3: Sample descriptionof UP PROFILES Curricular Delphi Study

5 Results – descriptive and variance statistical analyses

In the following sections, the results of the third round of the UP PROFILES Curricular Delphi Study on Science Education are presented. The results include descriptive-statistical analyses with regard to the priority and practice assessments as well as to the identified priority-practice differences.

The analyses and descriptions are made on the basis of both the total sample and the two different sample groups (students and teachers).

According to the structure of the questionnaire, the description of the results is divided into three parts. The first part (5.1) refers to the general assessments of the three given concepts of desirable science education by the total sample (see Table 1 above); the second part (5.2) considers the assessments of the concepts by the total sample differentiated according to different educational levels; finally, the third part (5.3) refers to the general assessments of the concepts by the two sub-sample groups.

As a test to identify statistically significant differences between the assessments of the three concepts, the Wilcoxon signed-rank test was applied (Bortz, 2005). This significance test was applied for the following three possible pair comparisons: Concept A / Concept B, Concept A / Concept C, and Concept B / Concept C. Furthermore, in order to identify statistically significant differences between the assessments of the two sub-sample groups, the Mann-Whitney-U test was applied. This significance test was applied for the following one possible pair comparison: students/teachers.

Data analysis was conducted using the Statistical Package for Social Sciences (SPSS), version 19.

Statistically significant differences taking into account a confidence level of 95% are displayed in **bold letters**.

5.1. Concepts of desirable science education – general assessment by the total sample

This section will address the general assessments of the three concepts by the total sample with regard to priority (5.1.1), realization in practice (5.1.2), the calculated priority-practice differences (5.1.3) and the comparison between priority and practice means (5.1.4). A brief overview is present at the end of the section (5.1.5).

5.1.1 Priority assessments

Concept A		Concept B		Concept C		Significance values		
						A/B	A/C	B/C
Mean value	Standard deviation	Mean value	Standard deviation	Mean value	Standard deviation	0,508	0,665	0,504
3,99	0,744	3,92	0,665	3,93	0,821			

Table 4: Mean values and standard deviation of the general priority assessments by the total sample and significance test values (Wilcoxon signed-rank test)

As we can see in Table 4, the assessment values of all three concepts are above the median scale point (3), suggesting that our sample clearly recognizes their importance. Concept A has the highest mean (3.99) closely followed by Concepts C and B (3.93 and 3.92, respectively). The values do not differ in a statistically significant way.

5.1.2 Practice assessments

As one may see in Table 5, the assessment values of the practice realization are below the median point of the scale (3). Once again, the highest mean is to be found in Concept A (2,39), followed by Concepts C and B (2,31 and 2,22, respectively). No statistically significant differences were found (although in the comparison pair A/B, *p* is close to the statistical threshold of 0,05).

Concept A		Concept B		Concept C		Significance values		
						A/B	A/C	B/C
Mean value	Standard deviation	Mean value	Standard deviation	Mean value	Standard deviation	0,06	0,511	0,29
2,39	0,808	2,22	0,685	2,31	0,660			

Table 5: Mean values and standard deviation of the general practice assessments by the total sample and significance test values (Wilcoxon signed-rank test)

5.1.3 Priority-practice differences

This sub-section addresses the priority-practice differences (PPD) in the assessments of the total sample. The calculated priority-practice differences show the gap that exists according to the assessments of the sample between the priority they assign to a concept and its perceived realization in educational practice. The priority-practice were

determined on the basis of all data by subtracting the practice values from the priority values ($\Delta PPD = X_{Priority} - Y_{Practice}$).

Table 6 displays the PPD for the total sample as well as the results from the significance test (Wilcoxon signed-rank test) with respect to the pair comparisons of the concepts' assessments.

Concept A		Concept B		Concept C		Significance values		
						A/B	A/C	B/C
Mean value	Standard deviation	Mean value	Standard deviation	Mean value	Standard deviation	0,406	0,389	0,828
1,59	0,851	1,69	0,799	1,62	0,886			

Table 6: Mean values and standard deviation of the priority-practice differences of the total sample and significance test values (Wilcoxon signed-rank test)

As one may observe in Table 6, the three concepts present positive values. As previously referred, the mean values of priority assessments are systematically higher than the mean values of the practice assessments. Thus, the participants' perception of realization of all three concepts in science education is below the importance given to the concepts. The mean PPD values of the total sample range between 1,59 (Concept A) to 1,69 (Concept B). No statistically significant differences were found.

In the next section, we will explore the gap between given importance and perceived realization in practice.

5.1.4 Priority and practice assessments

This section addresses the discrepancy observed between given importance and perceived realization in practice considering all three concepts in desirable science education. Table 7 presents the mean values for each concept as well as the total mean both for priority assessments and practice assessments.

Concepts	Mean values		Significance values
	Priority	Practice	
Concept A	3,99	2,39	**0,0**
Concept B	3,92	2,22	**0,0**
Concept C	3,93	2,31	**0,0**
Total	3,94	2,31	**0,0**

Table 7: Mean values and standard deviation of the priority-practice differences of the total sample and significance test values (Wilcoxon signed-rank test)

As observed before, priority assessments means are systematically higher and above the median point of the five-tier scale. On the contrary, practice assessments means are systematically lesser and below the median point of the scale. All four comparisons present significant statistically differences (p=0,0).

5.1.5 Summary

In brief, according to the results presented in this section, it is possible to extract two major observations:

- The UP sample participants attribute high (and positive) importance to all three concepts in science education. This is, it seems that all three concepts are equivalently important according to UP sample participants' point of view.
- The UP sample participants perceptions of the practice realization of all three concepts is lesser (and negative) than the given importance. This is, according to participants' point of view, even if they are important, these concepts are not being properly implemented.

5.2 Concepts of desirable science education – assessment by total sample regarding educational levels

After having considered the results from the general assessment of the three concepts by the total sample with regard to priority, realization in practice, the calculated priority-practice differences and the comparison of priority and practice means, the following sections will address the results from the assessments by the total sample differentiated according to the following different educational levels:

- Elementary level
- Lower secondary education
- Higher secondary education
- University

The descriptions are structured again into priority assessments (5.2.1), practice assessments (5.2.2) and the calculated priority-practice differences (5.2.3). A summary is presented at the end of the section (5.3.4).

5.2.1 Priority assessments

As shown in Table 8, the assessments means regarding educational level are very similar across concepts. Concept A presents the highest means, ranging from 3,61 in Elementary level to 4,29 in Higher Education level, followed by Concept C, except for University level. For this level, Concept B presents the highest mean (4,23), followed by Concept A (4,16).

Higher Secondary Education presents the highest average of all three concepts (4,24), followed by University (4,17), Lower Secondary Education (3,84). Elementary level has the lowest mean value (3,53).

No statistically significant differences are present when comparing concepts in science education regarding educational levels.

	Mean values				Significance values		
Educational level	Concept A	Concept B	Concept C	Average of all three concepts	A/B	A/C	B/C
Elementary level	3,61	3,49	3,51	3,53	0,499	0,544	0,827
lower secondary education	3,89	3,77	3,87	3,84	0,384	0,788	0,474
higher secondary education	4,29	4,18	4,23	4,24	0,419	0,499	0,627
University	4,16	4,23	4,10	4,17	0,635	0,635	0,248

Table 8: Mean values of the priority assessments by the total sample regarding different educational levels and significance test values (Wilcoxon signed-rank test)

5.2.2 Practice assessments

As it is possible to see in Table 9, all practice assessment values are negative (below the median point of the scale). The lowest values relate to Elementary level (average of all three concepts = 1,97). At this educational level, Concept B registers the lowest mean (1,79). Contrarily to what has been observed in priority assessments (see section 5.2.1, above), practice assessment values differ in a statistically significant way in the comparison pairs A/B (p= 0,033) and B/C (p=0,014) at the Elementary level.

	Mean values				Significance values		
Educational level	Concept A	Concept B	Concept C	Average of all three concepts	A/B	A/C	B/C
Elementary level	2,05	1,79	2,08	1,97	**0,033**	0,783	**0,014**
Lower secondary education	2,34	2,23	2,05	2,21	0,47	0,66	0,26
Higher secondary education	2,37	2,26	2,36	2,32	0,18	1	0,48
University	2,82	2,62	2,74	2,72	0,357	0,782	0,510

Table 9: Mean values of the practice assessments by the total sample regarding different educational levels and significance test values (Wilcoxon signed-rank test)

According to the UP sample perceptions, it is at the University level that the highest implementation of science education concepts is to be found (average of all three concepts = 2,72), followed by Higher Secondary Education (2,32) and Lower Secondary Education (2,21).

5.2.3 Priority-practice differences

In Table 10, one may observe that the slightest differences between priority and practice assessments in science education concepts occur at the University level (average of all three concepts= 1,45) and the highest at the Higher Secondary Education level (average of all three concepts= 1,91).

	Mean values				Significance values		
Educational level	Concept A	Concept B	Concept C	Average of all three concepts	A/B	A/C	B/C
Elementary	1,55	1.69	1.44	1,55	.197	.651	.144
Lower secondary education	1.55	1.54	1,82	1,64	.891	.119	.102
Higher secondary education	1.92	1.92	1.87	1,91	.888	.947	.823
University	1.34	1.62	1.36	1,45	.229	.655	.323

Table 10: Mean values of the priority-practice differences of the total sample regarding different educational levels and significance test values (Wilcoxon signed-rank test)

No statistically significant differences are present when comparing PPD in science education regarding educational levels.

5.2.4 Summary

In brief, it is possible to outline three major observations from the results presented in this sub-section:

- The UP sample participants perceive the concepts in science education as being more important at Higher Secondary Education and at University level. Simultaneously, it is at the University level that the UP sample participants perceive more realization in practice. The gap between priority and practice realization at Higher Secondary Education.
- The UP sample participants perceive the concepts in science education as being less important at Elementary level (when compared with other education levels). Simultaneously, they perceive less implementation of these concepts at Elementary level. This point deserves more attention if one considers that Elementary level is often perceived as a fundamental ground for learning.
- Finally, at Elementary level, Concept B seems to be even less implemented.

5.3 Concepts of desirable science education – general assessment by the sub-sample groups

In this section, the concepts of desirable science education are presented by sub-sample groups (i.e., students and teachers). The section follows the structure of the previous one. Fist, priority assessments are presented (5.3.1); second, practice assessments (5.3.2); third, priority-practice differences (5.3.3). A summary is given at the end of the section (5.3.4).

5.3.1 Priority assessments

As one may see in Table 11, the average of all three concepts is higher for the sub-sample group of students (4,23) than for the sub-sample group of teachers (3,82). In both sub-sample groups, Concept A presents the highest mean (students = 4,33; teachers = 3,83). The mean values are high and positive. Nonetheless, sub-sample groups do not differentiate the concepts of desirable science education. No statistically significant differences are observed.

UP sub-sample groups	Mean values				Significance values		
	Concept A	Concept B	Concept C	Average of all three concepts	A/B	A/C	B/C
Students	4,33	4,17	4,19	4,23	0,170	0,481	1
Teachers	3,83	3,81	3,81	3,82	0,979	0,937	0,517

Table 11: Mean values of the general priority assessments by the sub-sample groups and significance test values (Wilcoxon signed-rank test)

In Table 12, the mean values of priority assessment of each concept of desirable science education are compared regarding the sub-samples groups. No statistically significant differences are found, however Concept A is close to statistically significant threshold of 0,05 (p=0,053).

Concepts	Significance values	Mean values		
	Students/Teachers	Students	Teachers	Total
Concept A	0,053	4,33	3,83	3,97
Concept B	0,126	4,17	3,81	3,92
Concept C	0,233	4,19	3,81	3,93
Number of statistical significant differences	0			

Table 12: Mean values of the general priority assessments by the sub-sample groups and significance test values (Mann-Whitney-U-Test)

5.3.2 Practice assessments

As one may observe in Table 13, practice assessments of all three concepts of desirable science education are negative and bellow the median scale point. Students perceptions of practice realization (average of all three concepts = 2,74) are systematically higher than teachers' (average of all three concepts = 2,11). In the sub-sample group of students, Concept A and concept C have the same mean (2,75), while Concept B is slightest lesser (2,71). In the sub-sample group of teachers, Concept A presents the highest mean (2,23) and concept B the lesser (2,01) and the difference observed is statistically significant (p=0,023).

UP sub-sample groups	Mean values				Significance values		
	Concept A	Concept B	Concept C	Average of all three concepts	A/B	A/C	B/C
Students	2,75	2,71	2,75	2,74	0,892	0,891	0,713
Teachers	2,23	2,01	2,11	2,11	0,**023**	0,165	0,472

Table 13: Mean values of the general practice assessments by the sub-sample groups and significance test values (Wilcoxon signed-rank test)

In Table 14, one may observe that the differences between the sub-sample groups regarding concept B and concept C are statistically significant (p= 0,011 and p=0,014, respectively).

Concepts	Significance values (Students/Teachers)	Mean values		
		Students	Teachers	Total
Concept A	0,207	2,75	2,23	2,39
Concept B	**0,011**	2,71	2,01	2,22
Concept C	**0,014**	2,75	2,11	2,31
Number of statistical significant differences	2			

Table 14: Mean values of the general practice assessments by the sub-sample groups and significance test values (Mann-Whitney-U-Test)

5.3.3 Priority-practice differences

According to the results presented in Table 15, the sub-sample group of teachers (1,7) has a higher priority-practice difference than students' (1,49). No statistically significant differences are observed.

	Mean values				Significance values		
UP sub-sample groups	Concept A	Concept B	Concept C	Average of all three concepts	A/B	A/C	B/C
Students	1,58	1,46	1,44	1,49	0,670	0,526	0,917
Teachers	1,59	1,79	1,7	1,7	0,2	0,094	0,735

Table 15: Mean values of the general priority-practice differences of the sub-sample groups and significance test values (Wilcoxon signed-rank test)

As one may observe in Table 16, no statistically significant differences are observed between students and teachers sub-sample groups.

Concepts	Significance values	Mean values		
	Students/Teachers	Students	Teachers	Total
Concept A	0,792	1,58	1,59	1,59
Concept B	0,271	1,46	1,79	1,69
Concept C	0,126	1,44	1,7	1,62
Number of statistical significant differences	0			

Table 16: Mean values of the general priority-practice differences of the sub-sample groups and significance test values (Mann-Whitney-U-Test)

5.3.4 Summary

In brief, the results presented in this section may be organized in two major observations:

- UP sub-sample group of students perceive the concepts of desirable science education as being more important and, actually, being more implemented in practice.
- UP sub-sample group of teachers perceive the concepts of desirable science education as less than students do. Teachers' perceptions of practice realization are also lesser. The gap between priority and practice is greater among teachers. This point must be stressed as it suggests actor-observation differences.

6 Discussion

In this section, a detailed discussion of the results is presented. The section is structured in following sub-sections: general assessment by the total sample (6.1), Assessment by the total sample regarding different educational levels (6.2); and General assessment by the sub-sample groups (6.3).

6.1 General assessment by the total sample

The results of the analyses presented in sub-section 5.1 show that – regarding the *priority and practice assessments* – the sample considers all three concepts categories as relevant. Regarding *practice assessments*, it is possible to conclude that the sample consider that practice usually falls short considering given priority.

As explained in UP Interim Report on the Second Round (Paiva, Morais e Barros, 2012), the clusters in Portuguese dendrogram could not be interpreted to identify concepts of desirable science education. The third round draw on FUB results (Schulte & Bolte, 2012). This may possible help to understand why the Portuguese sample participants' assessment failed to distinguish among concepts.

6.2 Assessment by the total sample regarding different educational levels

The results of the analyses presented in sub-section 5.2 suggest that the priority and practice perceptions of the concepts of desirable science education vary according to educational levels.

It seems that the more advanced the educational level is, the higher the priorities attributed to the concepts. A specific gap between priority and practice has been identified in Higher Secondary Education.

It is particularly interesting that Elementary level seems to be an outlier. The data do not let us know if the perception associated to Elementary level, where the concepts of desirable science education are considered less important than elsewhere and where practice realization is perceived to be lesser than elsewhere, is related to sample constraints or if represents a relevant insight on how Portuguese teachers and students represent the specific opportunities to science education in the educational system.

6.3 General assessment by the sub-sample groups

As previously explained, UP sample is divided in two groups: students and teachers. It is important to keep present the fact that students and teachers perform different roles in the educational system and in the classroom.

Therefore, that teachers and students have different perception regarding the priority and realization in practice of the concepts of desirable science education is not surprising, but it is important to discuss the direction of the difference.

Students seem at once more demanding, as their assessment on the priority of the concepts in desirable science education is higher. Nonetheless, students are simultaneously less critic, as their assessments on practice realization are also higher.

Contrarily, teachers are, perhaps, less ambitious regarding the priority of the concepts but more critical considering the implementation. Thus, it would be important to know the locus of control of the two groups on how these concepts should be integrated in educational settings.

7 Summary and outlook

In this UP Interim report on the third round of the Curricular Delphi Study on Science Education, the results of the final questionnaire answered by the UP sample were presented and discussed.

Considering the second round (Paiva, Morais & Barros, 2012), UP sample registered a drop-out of 15 participants. Thus, the third round registered a response rate of 45% attending to the first round (n=86). Several constraints are connected to UP sample. One must stress, specifically, two limitations. On the one hand, it is a relatively small sample (n = 39). On the other hand, it was only possible to divide the sample in two groups: students (n=12) and teachers (n=27). Nonetheless, one may expect that the perceptions and evaluations of the UP sample participants may be useful to gain insights on the concepts of desirable science education in Portuguese context.

The three concepts were considered relevant to science education even if UP sample participants were unable to differentiate them. This may be explained by the difficulties to identify the concepts of desirable science education in UP second round. As a matter of fact, It is not possible to say if the concepts developed in the course of the hierarchical cluster analyses in the second round by FUB team (Bolte & Schulte, 2012; Schulte & Bolte, 2012), which were adopted for the UP third round, apprehend Portuguese reality. The results suggest that other structure would be more adequate.

Although this may be the case, some interesting findings deserve our attention.

- There is a serious gap between the importance given to the concepts in desirable science education and practice realization perceptions extensive to all three concepts, according to the two sub-sample groups, at every educational level.
- It seems that the more advanced the educational level, the higher the priorities attributed to the concepts are. Some isomorphism it is found in practice realization: the more advanced the level the highest practice realization is observed with the exception of Secondary education that present the largest gap between priority and practice. This may suggest that more attention should be given to this level. If science education is to be reformed, serious attention should be given to the elementary level, as all sub-sample groups seem to give it less importance.
- There are clear sub-sample group differences. This may refer to actor-observation differences, but if this is the case or if it not we cannot tell.

It seem important to compare the Portuguese panorama outlined here with the PROFILES country partners in order to have a global view on science education in the European setting.

8 References

Bolte, C. (2003). Chemiebezogene Bildung zwischen Wunsch und Wirklichkeit - Ausgewählte Ergebnisse aus dem zweiten Untersuchungsabschnitt der curricularen Delphi-Studie Chemie. *ZfDN*, *9*, 27–42.

Bolte, C. (2008). A Conceptual Framework for the Enhancement of Popularity and Relevance of Science Education for Scientific Literacy, based on Stakeholders' Views by Means of a Curricular Delphi Study in Chemistry. *Science Education International*, *19*(3), 331–350.

Bolte, C., & Schulte, T. (2011). PROFILES Curricular Delphi Study on Science Education. Interim Report on the First Round of the FUB Working Group. *Polyscript (Status July 2011). Unpublished*.

Bolte, C., & Schulte, T. (2012). PROFILES Curricular Delphi Study on Science Education. Interim Report on the Second Round of the FUB Working Group. *Polyscript (Status May 2012). Unpublished.*

Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaften*. Heidelberg: Springer.

Bybee. (1997). Toward an Understanding of Scientific Literacy. In W. Gräber & C. Bolte (Eds.), *Scientific Literacy - An International Symposium*. Kiel: IPN.

DeBoer, G. (2000). Scientific Literacy. Another Look at its Historical and Contemporary Meanings and its Relationship to Science Education Reform. *Journal of Research in Science Teaching*, *37*(6), 582–601.

Eckebrecht, D., & Schneeweiß, H. (2003). *Naturwissenschaftliche Bildung: Gedanken und Beispiele zur Umsetzung von scientific literacy* (1. Aufl.). Stuttgart [u.a.]: Klett.

Gräber, W., & Bolte, C. (1997). *Scientific Literacy. An International Symposium*. Kiel: IPN.

Gräber, W., & Nentwig, P. (2002). Scientific Literacy - Naturwissenschaftliche Grundbildung in der Diskussion. In *Scientific Literacy. Der Beitrag der Naturwissenschaften zur allgemeinen Bildung* (pp. 7–20). Opladen: Leske und Budrich.

Häußler, P., Frey, K., Hoffmann, L., Rost, J., & Spada, H. (1980). *Physikalische Bildung: Eine curriculare Delphi-Studie. Teil I: Verfahren und Ergebnisse. IPN-Arbeitsbericht 41*. Kiel: IPN.

Linstone, H. A., & Turoff, M. (1975). *The Delphi Method: Techniques and Applications*. Reading, Mass. u.a: Addison-Wesley.

Mayer, J. (1992). *Formenvielfalt im Biologieunterricht: Ein Vorschlag zur Neubewertung der Formenkunde*. Kiel: IPN.

Paiva, J., Morais, C. & Barros, J. (2011). PROFILES Curricular Delphi Study on Science Education. Interim Report on the First Round of the UP Working Group. *Unpublished*.

Paiva, J., Morais, C. & Barros, J. (2012). PROFILES Curricular Delphi Study on Science Education. Interim Report on the Second Round of the UP Working Group. *Unpublished*.

PROFILES Consortium. (2010). FP7 Negotiation Guidance Notes - Coordination and Support Actions. Annex I - Description of Work. *unpublished*.

Schulte, T., & Bolte, C. (2012). European Stakeholders Views on Inquiry Based Science Education – Method of and Results from the International PROFILES Curricular Delphi Study on Science Education Round. In C. Bolte, J. Holbrook, & F. Rauch (Eds.), *Inquiry based Science Education in Europe - First Examples and Reflections from the PROFILES Project* (pp. 42–51). Berlin: Freie Universität Berlin. Bolte, C. (2003). Chemiebezogene Bildung zwischen Wunsch und Wirklichkeit - Ausgewählte Ergebnisse aus dem zweiten Untersuchungsabschnitt der curricularen Delphi-Studie Chemie. *ZfDN*, *9*, 27–42.

9 Appendix

9.1. Tables

Priority assessments of the UP total sample			Total sample		
			Valid N	Mean values	Standard deviation
General assessment	Concept A		38	3.99	.744
	Concept B		39	3.92	.665
	Concept C		39	9.93	.821
Assessment differentiated according to different educational levels	Concept A	Elementary level	38	3.61	1.028
		Lower Secondary Education	38	3.89	.863
		Higher Secondary Education	38	4.29	.835
		University	38	4.16	.945
	Concept B	Elementary level	39	3.49	.914
		Lower Secondary Education	39	3.77	.742
		Higher Secondary Education	39	4.18	.823
		University	39	4.23	.902
	Concept C	Elementary level	39	3.51	.914
		Lower Secondary Education	39	3.87	.951
		Higher Secondary Education	39	4.23	.872
		University	39	4.1	1.021

Table 17: Priority assessment of the UP total sample – general assessment and assessment regarding different educational levels

Priority assessments of the UP sub-sample group students		Students		
		Valid N	Mean values	Standard deviation
General assessment	Concept A	12	4.33	.404
	Concept B	12	4.17	.222
	Concept C	12	4.19	.614

Table 18: Priority assessment of the UP sub-sample group students – general assessment

Priority assessments of the UP sub-sample group students		Teachers		
		Valid N	Mean values	Standard deviation
General assessment	Concept A	26	3.82	.815
	Concept B	27	3.81	.764
	Concept C	27	3.81	.884

Table 19: Priority assessment of the UP sub-sample group teachers – general assessment

Practice assessments of the UP total sample			Total sample		
			Valid N	Mean values	Standard deviation
General assessment	Concept A		38	2.39	.807
	Concept B		39	2.22	.685
	Concept C		39	2.31	.66
Assessment differentiated according to different educational levels	Concept A	Elementary level	38	2.05	.769
		Lower Secondary Education	38	2.34	1.072
		Higher Secondary Education	38	2.37	1.051
		University	38	2.82	1.43
	Concept B	Elementary level	39	1.79	.409

		Lower Secondary Education	39	2.23	.986
		Higher Secondary Education	39	2.26	.966
		University	39	2.62	1.350
	Concept C	Elementary level	39	2.08	.739
		Lower Secondary Education	39	2.05	.510
		Higher Secondary Education	39	2.36	1.038
		University	39	2.74	1.352

Table 20: Practice assessments of the UP total sample – general assessment and assessment regarding different educational levels

Practice assessments of the UP sub-sample group students		Students		
		Valid N	Mean values	Standard deviation
General assessment	Concept A	12	2.75	1.108
	Concept B	12	2.71	.838
	Concept C	12	2.75	.783

Table 21: Practice assessment of the UP sub-sample group students – general assessment

Practice assessments of the UP sub-sample group students		Teachers		
		Valid N	Mean values	Standard deviation
General assessment	Concept A	26	2.23	.578
	Concept B	27	2.01	.483
	Concept C	27	2.11	.497

Table 22: Practice assessment of the UP sub-sample group teachers – general assessment

Priority-Practice differences of the UP total sample			Total sample		
			Valid N	Mean values	Standard deviation
General assessment	Concept A		38	1.59	.851
	Concept B		39	1.69	.799
	Concept C		39	1.62	.886
Assessment differentiated according to different educational levels	Concept A	Elementary level	38	1.55	1.083
		Lower Secondary Education	38	1.55	1.179
		Higher Secondary Education	38	1.92	1.239
		University	38	1.34	1.582
	Concept B	Elementary level	39	1.69	.922
		Lower Secondary Education	39	1.54	.1072
		Higher Secondary Education	39	1.92	1.085
		University	39	1.62	1.369
	Concept C	Elementary level	39	1.44	1.021
		Lower Secondary Education	39	1.82	.914
		Higher Secondary Education	39	1.87	1.174
		University	39	1.36	1.597

Table 23: Priority-practice differences of the UP total sample – general assessment and assessment regarding different educational levels

Priority-Practice differences of the UP sub-sample group students		Students		
		Valid N	Mean values	Standard deviation
General assessment	Concept A	12	1.58	1.024
	Concept B	12	1.46	.838
	Concept C	12	1.44	.739

Table 24:Priority-Practice differences of the UP sub-sample group students – general assessment

Priority-Practice differences of the UP sub-sample group students		Teachers		
		Valid N	Mean values	Standard deviation
General assessment	Concept A	26	1.59	.781
	Concept B	27	1.79	.775
	Concept C	27	1.7	.946

Table 25: Priority-Practice differences of the UP sub-sample group teachers – general assessment