Part 11 / Strand 11

Evaluation and Assessment of Student Learning and Development

***Editors:** Lukas Rokos & Mathias Ropohl*

Part 11. Evaluation and Assessment of Student Learning and Development

Development, validation and use of standardized tests, achievement tests, high stakes tests, and instruments for measuring attitudes, interests, beliefs, self-efficacy, science process skills, conceptual understandings, etc.; authentic assessment, formative assessment, summative assessment; approaches to assessment. Monitoring student learning and implications for teaching.

Part 11: Evaluation and Assessment in Science Education

Editors: *Lukas Rokos & Mathias Ropohl*

Introduction

Evaluation and assessment are still hot topics for researchers (not only) in the field of science education. Evaluation and assessment are important in all contributions presented in Strand 11. In total, there are five contributions from different educational backgrounds in this chapter, namely, Czech Republic, Greece, Portugal, and Singapore. Individual papers present very different perspectives on the issue of evaluation and assessment within various educational systems, and different educational levels (primary, secondary, tertiary), but they also differ in the methodology used (quantitative, qualitative, and mixed methods). Whether we focus on biology, physics, chemistry or mathematics, standardised tests, assessing students' motivation to study science subjects or ways of providing feedback on student performance, you will read interesting insights of the presented research projects that are useful for the entire ESERA community.

We hope that the following contributions will inspire you:

1) National exams and standardised tests

Quality assessment of the Portuguese secondary education biology and geology national exams is a paper focusing on external evaluation at the secondary level, specifically on students' results in biology and geology. It demonstrates the lack of validity and reliability of these exams, as well as cognitively demanding tasks that could be factors for the high number of failures.

Data sensemaking demands of stimulus-based questions in standardized physics assessments presents preliminary findings from the analysis of stimulus-based questions in Singapore. The coding scheme introduced in the paper has the potential to be generally used in stimulus-based questions related to biology and chemistry.

2) Assessment of abstraction skills

Assessing abstraction skills in early primary school amid environmental study describes the development and assessment of computational thinking skills and the use of a respective assessment tool developed by the authors. The statistical analysis of first and second-grade primary school students' abstraction skills provides a background for studies that will attribute ecological essence to computational thinking activities. By the analysis, the correlation between computational thinking and environmental consciousness could be validated.

3) Motivation

The impact of undergraduate students' academic experience on their motivation to learn chemistry and physics uses cluster sampling for students from three different academic majors to investigate motivation in higher education. The authors found that recent academic experience has stronger effects on students' motivation to learn chemistry and physics than more distant experiences.

4) Feedback as a key aspect of assessment

Analysis of teacher written feedback in biology inquiry task with stress on affective connotations introduces findings from distance teaching in the Czech Republic and efforts of in-service teachers to provide formative feedback to students, including an explanation of the assessment criteria and work with their mistakes. Although affective connotations may have a strong motivational effect, they are still rather rare in the feedback provided by teachers.

QUALITY ASSESSMENT OF THE PORTUGUESE SECONDARY EDUCATION BIOLOGY AND GEOLOGY NATIONAL EXAMS

Teresa Lopes[1], *José Precioso*[1]
[1]Universidade do Minho, Braga, Portugal

In Portugal, the external evaluation of students in secondary education takes place through national exams, which grades have a significant impact on their academic future. The students' results in Biology and Geology national exams have revealed a prevalent and persistent scenario of failure, with very low average scores and excessively high failure rates. In the teachers' opinion, among the main causes of this failure is the complexity of the exam. Thus, this study aimed to analyze the validity and technical quality of these tests in order to understand whether these assessment instruments are contributing negatively to the students' failure. To achieve this, a qualitative investigation was carried out, using documentary analysis through content analysis. Two Biology and Geology exams were analyzed, being the selection criterion the extremes of the students' results (the test with the best results and the test with the worst results), with the following objectives: (1) to analyze the validity and the technical quality of Biology and Geology exams; (2) to determine whether these exams assess the purposes of the subject syllabus; (3) to determine whether these exams assess the achievement of the educational objectives proposed by Bloom's Taxonomy.
It is concluded that the exams are cognitively demanding, since most questions assess higher categories of the cognitive process (application and analysis) of conceptual knowledge.
This investigation demonstrates the lack of validity and reliability of Biology and Geology exams, as well as technical quality problems.

Keywords: Science Education; Evaluation; Secondary School

INTRODUCTION

In Portugal, the conclusion and the certification of Secondary Education and the access to Higher Education depend on an evaluation system in which external assessment, in the form of national examinations, has a significant impact.

The students' results in Biology and Geology exam have revealed a serious situation of failure over the years, with very low average scores and excessively high failure rates (Lopes & Precioso, 2018). With regard to the average scores, the maximum score was 10.02 points (in 20) achieved in 2014 and 2016 exams, and the minimum score was 8.21 points, obtained by all students who took the exam in 2013 (Lopes & Precioso, 2018). As for the failure rates, they have been situated between 1/3 and 2/3 of the students who took the exam. The worst record occurred in 2013, in which the failure rate reached 64.4%, which means that 49235 exams revealed a score below 9.5 points, in a total of 76501 exams (Lopes & Precioso, 2018). This numbers show the real dimension of the problem.

In teachers' opinion, the main causes of the students' failure in Biology and Geology external assessment are: the exam's high degree of complexity, the stress and the anxiety caused in the students by the exam; students' difficulties related to reading, interpretation and communication skills; the very penalizing exam's correction and classification criteria; the high degree of complexity of the document analysis; the gap between what is required in the exam and what is required by the subject syllabus; the inadequacy of the exam to the students' maturity and the extensive subject syllabus (Lopes & Precioso, 2019).

In addition, several studies carried out in Portugal (Rosário, 2007; Raposo & Freire, 2008; Madureira, 2011; Sousa, 2011; Lopes, 2013; Marques et al., 2015) show the influence of national exams on pedagogical practices and teachers' evaluations. Teachers guide their practices to "train" students for what is required in the exam, using practices that they do not believe that are of the highest quality for learning; they make their evaluation tests with a similar structure of the exams and with the same classification criteria, even not agreeing with them and considering them penalizing for the students; and show great concern in approaching the entire subject's syllabus, even knowing that the pace they impose for teaching does not provide quality learning (Lopes, 2013).

In this way, exams, due to their existence, induce an inadequate learning model, which, for Black (1998), implies concluding their lack of validity. This is what has been described by research regarding exams in Portugal, which induce teaching to the test, which calls into question the validity of national tests.

Thus, it is important to analyze the validity and the technical quality of these exams to assess whether these assessment instruments are contributing negatively to the students' failure.

OBJECTIVES

Previous studies carried out by us (Lopes & Precioso, 2018) show a prevalent and persistent situation of severe school failure in the Biology and Geology exam, as male and female students have been getting worryingly low average ratings (between 8 and 11 points) and very high failure rates (between 45% and 65%). Therefor it is important to analyze the quality of these exams, in order to ascertain whether the assessment instrument may be contributing to this failure.

The objectives of this study are: (1) to analyze the validity and the technical quality of Biology and Geology exams; (2) to determine whether Biology and Geology exams assess the purposes of the subject syllabus; (3) to determine whether Biology and Geology exams assess the achievement of the educational objectives proposed by Bloom's Taxonomy.

METHOD

This study's population consists of ten Biology and Geology national exams, 1st and 2nd phases, which were carried out in the academic years between 2012/2013, the last year in which changes were applied in the national tests, and 2017/2018, the last year with full data provided by the Júri Nacional de Exames. A non-probabilistic convenience sample was selected, being the criterion the extremes of the students' results, both in terms of average ratings and in terms of test failure rates: Biology and Geology Exam 2014, 1st phase, in which the students achieved the best results; and Biology and Geology Exam 2014, 2nd phase, in which the students achieved the worst results. We opted for a qualitative investigation using document analysis and content analysis. Data collection instruments were constructed, which consisted of analysis grids to analyze the exam questions in order to obtain data on the following dimensions: Science Teaching Dimensions; Bloom's Taxonomy Cognitive Process Dimension. The analysis grids were subjected to prior validation by specialists in Education and Science Education.

In the data treatment, each one of the questions in each one of the exams was analyzed qualitatively, making the respective categorization for each of the dimensions analyzed (presence/absence of the category and points). Finally, it was made a quantitative approach. The results were recorded in a comparative table of the two tests.

RESULTS

The Biology and Geology exams consist of 4 groups of questions, 2 of Biology and 2 of Geology. 4 problems are presented to students and problem-solving questions are asked about them.

Regarding the dimensions of Science Teaching, the main Science Teaching purpose assessed is Learning science, as shown in Table 1. There is consistency from one test to the other, since both tests have 26 questions (175 points) included in the category "Learning science" and 4 questions (25 points) inserted in the category " Learning to do science”. The categories “Learning about science” and “Learning through science” are not represented, although they are presented, in the normative documents of the subject, as important purposes. It is understandable that the exam focuses more on “Learning Science”, that is, on the contents, however, if the syllabus’ subject program considers the other categories as important purposes, it is not understandable that two of them are not even represented, since that the exam, as a subject evaluation instrument, must evaluate the achievement of its purposes.

Table 1. Compared analysis of 2014, 1st phase and 2014, 2nd phase exams.

Quality evaluation of Biology and Geology exams		Exam: 2014 1st phase		Exam: 2014 2nd phase	
		n	Points (in 200)	n	Points (in 200)
Science Teaching Dimensions	Learning science	26	175	26	175
	Doing science	4	25	4	25
	Learning about science	0	0	0	0
	Learning through science	0	0	0	0
Bloom's Taxonomy: Cognitive Process Dimension	Remember	0	0	0	0
	Understand	5	25	3	15
	Apply	14	75	15	80
	Analyze	10	90	11	90
	Evaluate	1	10	1	15
	Create	0	0	0	0

Regarding the dimension of the Cognitive Process of Bloom's Taxonomy, there are some differences that may explain the differences in the students' classifications in the two exams. Although the two exams focus on the application and analysis categories, there are some differences. In the case of the 1st phase, the test in which the students had better results, there are 5 questions in the comprehension category, corresponding to 25 points, while in the 2nd phase, the test in which the students had the worst results, there are only 3 comprehension questions, corresponding to just 15 points. In the "Apply" category, in the 1st phase, there are 14 questions, corresponding to 75 points, while in the 2nd phase, there are 15 questions, corresponding to 80 points. In the "Analysis" category, in the 1st phase there are 10 questions and in the 2nd phase there are 11 questions, but corresponding equally to 90 points. In the "Evaluate" category, the highest category present in the tests, there is only one question in each test, but in the 1st phase, it corresponds to 10 points and, in the 2nd phase, it corresponds to 15 points. In none of the tests there are items that configure the categories of "Remember", the least demanding category, and "Create", the most demanding. But it is important to remember that, within the same category, the questions can have different degrees of difficulty, either because of the content they cover, in terms of quantity and quality, or because of the level and quantity of relationships and inferences that the student has to make to answer the questions.

The exams often assess specific contents and concepts with a higher degree of depth and/or difficulty than what the subject syllabus recommends. Answering to the questions requires the reading and elaborated analysis of documents (9 in the 2014, 1st phase and 8 in the 2014, 2nd phase) which can be texts, images, graphs, maps or tables. In general, these sources of information present new situations which are complex to read and analyze, using elaborate scientific language. There is a clear privilege of multiple-choice questions, although the exams' structure is maintained, with regard to the number and types of questions. It should be noted that, sometimes, the degree of difficulty arises with artificialities and less correct formulations of the questions that cause doubts and insecurities for students (Lopes, 2020).

CONCLUSIONS AND IMPLICATIONS

It is concluded that the exams are cognitively demanding tests, in which most questions assess higher categories of the cognitive process, such as Apply and Analyze, of conceptual knowledge.

It is concluded that the exams require high-level cognitive skills, which is in line with the results of other studies carried out in the same context in our country. Preto (2008) analyzed the degree of difficulty of items related to experimental activities of the 1st and 2nd phases of the 2006 exam and concluded that the questions require a high level of conceptual demand as they require the association of different contents, mainly requesting skills in the domain of reasoning.

The time for doing the exam (2 hours, with 30 minutes of tolerance) is not compatible with the interpretation of so much information in such a short period of time, because answering the questions demands the integration of content that requires complex reasoning.

The privilege that is given to multiple choice questions decreases the validity of the test, because this type of items makes the assessment of the student's ability to develop a reasoning difficult (Black, 1998). The student may be doing at least partially correct reasoning and making the question wrong, or he may be doing the wrong reasoning and getting the question right. It may not let students' problem-solving skills emerge. And therefore, the student is not being given an opportunity to show what he really knows in this domain. This option for multiple choice questions is related to the attempt to increase the reliability of the test, but it decreases its validity. This fact shows the conflict between these two concepts, denounced by Black (1998).

Regarding the science teaching purposes, the exam focuses on the dimension Learning Science; neglecting other purposes, such as Learning to do Science; Learning about the Science; and Learning through Science, which deprives it of validity, because it does not assess what the subject syllabus demands.

It is imperative to produce assessment instruments with better technical quality.

The exam prepared in 2020, due to the changes that the COVID pandemic has imposed during the academic year, may give some guidelines for its reformulation. It consisted of 33 items, of which only 10 were mandatory. 15 items of the remaining 23, contributed to the exam's final grade. Regarding the type of questions, short answer questions and selection completion questions (with multiple choice) were introduced. These changes gave students a greater opportunity to show what they know and what they are able to do and, therefore, we must consider if they should be introduced permanently in exams of the coming years.

A punctual test cannot be considered a rigorous and indisputable instrument that defines what a student knows and is capable of doing and, therefore, cannot so decisively determine the students' academic path and their future. The exam cannot be seen as an unquestionable and infallible assessment tool, but rather it must be rethought and continually improved.

ACKNOWLEGMENTS

This investigation was developed in the Centro de Investigação em Estudos da Criança (CIEC), with the reference UID/CED/00317/2019, financially supported by Fundação para a Ciência e a Tecnologia and by Ministério Português da Ciência, Tecnologia e Ensino Superior, through the award of an individual PhD scholarship (SFRH/BD/123731/2016).

REFERENCES

Black, P. J. (1998). *Testing, friend or foe?: the theory and practice of assessment and testing*. Psychology Press.

Lopes, T. (2013). *Perceções de Professores, Alunos e Encarregados de Educação sobre o (in)sucesso na disciplina de Biologia e Geologia*. Dissertação de Mestrado, Universidade do Minho, Braga, Portugal.

Lopes, T. (2020). *Insucesso Escolar na disciplina e no exame de Biologia e Geologia e fatores associados*. Tese de Doutoramento, Universidade do Minho, Braga, Portugal.

Lopes, T. & Precioso, J. (2018). Evolução do Insucesso Escolar nos Exames Nacionais do Ensino Secundário, por Sexo, em Portugal. *Revista Iberoamericana de Evaluación Educativa*, 11(2), p. 53-69. DOI: https://doi.org/10.15366/riee2018.11.2.003

Lopes, T. & Precioso, J. (2019). O que pensam os Professores portugueses sobre as causas do insucesso dos alunos no exame nacional da disciplina de Biologia e Geologia?. XV Congresso Internacional Galego-Português de Psicopedagogia (4 a 6 de setembro de 2019). Coruña, Espanha: Universidade da Coruña.

Madureira, M. (2011). *A influência dos exames nacionais de Física e Química A e respetivos resultados nas práticas de ensino e de avaliação dos professores*. Dissertação de Mestrado, Universidade do Minho, Braga, Portugal.

Marques, M.; Sousa, J.; Costa, N.; Pacheco, J. (2015). Efeitos da Avaliação Externa das Aprendizagens no Desenvolvimento Profissional de Professores de Matemática do Ensino Básico em Portugal. *Meta: Avaliação*, 7(19), 58-84.

Preto, A. (2008). *Ensino da Biologia e Geologia no Ensino Secundário: Exames e trabalho experimental*. Dissertação de Mestrado, Universidade de Lisboa, Lisboa. Portugal.

Raposo, P. & Freire, A. (2008). Avaliação das Aprendizagens: Perspectivas de Professores de Física e Química. *Revista da Educação*, XVI (1), 97-127.

Rosário, M. (2007). *Influência do exame nacional do 9º ano de escolaridade nas práticas de ensino e de avaliação em Matemática*. Dissertação de Mestrado, Universidade do Minho, Braga, Portugal.

Salgado, R. (2012). *O (in)sucesso em Física e Química A: Um estudo com alunos e professores de uma Escola Secundária de Guimarães*. Dissertação de Mestrado, Universidade do Minho, Braga, Portugal.

DATA SENSEMAKING DEMANDS OF STIMULUS-BASED QUESTIONS IN STANDARDISED PHYSICS ASSESSMENTS

Yann Shiou Ong[1] *and Lisha Mohandas Ragnani*[2]
[1]National Institute of Education, Nanyang Technological University, Singapore
[2] Nanyang Technological University, Singapore

Stimulus-based questions in science assessments typically require students to make sense of data presented as informational text, picture, table, or graph in order to answer the questions. However, there is a lack of a tool for analysing data sensemaking demands of stimulus-based questions based on the type and number of data representations, as well as the data reasoning process involved. To meet this need, this study proposed a data sensemaking demands coding scheme for analysing stimulus-based questions. Preliminary findings from analysing stimulus-based questions in Singapore's national physics examinations for tenth-graders as well as think-aloud interviews with two preservice physics teachers suggest researchers' expected data sensemaking demands (based on coding scheme) agreed fairly well with preservice teachers' enacted data sensemaking when answering a stimulus-based question. Also, moderate agreement was achieved on the coding scheme between two raters. Thus, proposed coding scheme has demonstrated reasonable response process validity and reliability. With further evaluation and revision, proposed data interpretation coding scheme has the capacity for analysing biology and chemistry stimulus-based questions, in addition to physics questions.

Keywords: Assessment, Physics, Representation

INTRODUCTION

Stimulus-based questions in high-stake science assessments are intended to evaluate higher order thinking skills associated with critical thinking. One such example is data-based questions (DbQs) in the biology, chemistry, and physics Singapore-Cambridge GCE Ordinary level (O-level) national examinations taken by secondary four (grade 10) students in Singapore. These questions (eight to 12 marks each) require students to interpret, evaluate, or solve short structured or open-ended items using given information (Singapore Ministry of Education, 2016). That is, students have to make sense of presented data in order to answer the questions. However, there is a lack of a useful tool with ease of use for analysing the data sensemaking demands associated with stimulus-based questions. While various knowledge taxonomies and frameworks for analysing general cognitive demands, (e.g. Biggs SOLO taxonomy, Webb's Depth of Knowledge, and Bloom's revised taxonomy) are widely used by researchers and educators, these taxonomies and frameworks do not adequately reflect demands of making sense of data presented in various representations (e.g. tables, graphs, diagrams). Hence, this study seeks to determine preliminary reliability and validity measures of a proposed data sensemaking demands coding scheme by answering the following research questions (RQs): 1. How reliable is the coding scheme in terms of inter-rater reliability measure? and 2. To what extent is response process validity achieved i.e. agreement between researchers' expected data sensemaking (based on assigned codes) and preservice physics teachers' enacted data sensemaking when answering a stimulus-based question?

DEMANDS OF DATA SENSEMAKING

Learning with multiple external representations is complex and involves considerations of design parameters of the multi-representational system, functions of these representations, and cognitive tasks undertaken by learners when interacting with representations (Ainsworth, 2006). Thus, cognitive demands of working with data are multidimensional. While it is beyond the scope of this study to design a comprehensive analytic framework for data reasoning demands, two aspects—visual representations of data (LaDue et al., 2015) and processes involved in working with presented data (Morris et al., 2015)—provide productive starting points.

Analysis of visual representations in high school New York State Regents (June 2012) examinations in the sciences by LaDue et al. (2015) identified four representation categories in the physics examination paper (in decreasing frequencies): diagrams, graphs, tables, and a network (i.e., chart depicting relationships among components). These visual representations, excluding network, informed data representation types codes in our data sensemaking demands coding scheme. LaDue and colleagues did not consider whether multiple representations were required to answer each item. However, based on challenges associated with learning using multiple external representations (Ainsworth, 2006), it is likely the number of representations and representation type(s) presented would affect item cognitive demand.

Morris et al. (2015) analysed opportunities for working with data in 20 American middle school science textbooks based on four scientific inquiry processes associated with working with data: data recording (recording/displaying/organising data), data analysis (analysing data using mathematical procedure; estimating outcomes), data interpretation (explaining phenomenon; reporting fastest, most, etc.; suggesting alternatives to explanations), and making predictions from data. Aside from data recording, which is not relevant for our analysed questions, other processes informed the data reasoning codes in our coding scheme.

METHODS

Selection of stimulus-based questions and think-aloud interviews

To answer RQ1, 15 stimulus-based items in physics O-Level examination papers between 2009 to 2018 were selected for coding. All question parts were treated as individual items. A total of 145 items were coded by both authors. To answer RQ2, two physics preservice teachers (PST) were interviewed individually by the first author using a think-aloud protocol (van Someren et al., 1994). Participants were recruited from students enrolled in a teacher preparation course at the time of the study. Only two PSTs, Ken and Leo (pseudonyms), volunteered. Both PSTs had physics as a second teaching subject. During interview, PSTs articulated how they answered a seven-part 2015 DbQ (presented in a handout) as they worked through each part. Due to copyright, the actual question is not shown. Parts of the question relevant for discussion of the PSTs' responses are described (with modifications) in Figure 1. Interviews were audio-recorded and transcribed (11 and 18 minutes each). Transcripts and participants' written responses were analysed. Interview transcripts were coded collaboratively by both authors in discussion to reach consensus codes.

Data analysis

The data sensemaking demands coding scheme comprises two dimensions: data representation (type and number) and data reasoning. Answering a DbQ part may require coordination between several data representations. Thus, multiple data representation type codes were allowed per question part (e.g., presented picture and graph may both be required to answer a part). Data representation types (Table 1) include: *information* (I) i.e., text, including equations; *picture* (Pic) i.e. diagram/figure describing aspects of physical object(s) or process(es); *table* (T) where values of two or more quantities are presented in tabulated format; *graph* (G), including line graph, scatter plot, or bar chart, or *no data* (NA). Since multiple data representation type codes were allowed, simple agreement of codes (i.e., all identified data representation types for a part are in agreement) was used to determine inter-rater reliability for this code. Initial iteration of data reasoning codes considered data analysing, interpreting, and predicting/explaining described in Morris et al. (2015). However, difficulties in distinguishing between the three processes during coding practise prompted us to redefine the processes as two codes. Final data reasoning codes (Table 2) considered whether *data was used as information* (DIn) i.e., identify relevant data as answer or selecting relevant data for mathematical manipulation to arrive at answer or whether *data interpretation* (DPr) was involved i.e. determining similarities/differences/ relationships within/between datasets, predicting relationship/outcomes, or relating data to scientific concepts. Question part without data representation is coded as *no data* (NA). Cohen's kappa was used to determine inter-rater reliability for data reasoning codes. Written responses and other inscriptions on interview handouts were considered when deliberating the codes.

Table 1. Data type codes.

Data Type	Description
Information (I)	Information (e.g., values of physical quantities) or chunks of text provided at the start of a question part.
Table (T)	Data i.e. values for two or more variables is presented in table format.
Graph (G)	Line graph, scatter plot, bar chart, etc.
Picture (Pic)	Diagram or figure depicting some aspect(s) of real object(s) or scenes.
No data (NA)	No data or information is provided/required to answer question.

Table 2. Data reasoning codes.

Reasoning Process	Description
Data as information (DIn)	Data is used as information to answer the question. This includes: Identifying relevant data as answer to question without further math operations or reasoning, OR Selecting relevant data for manipulation through math operations (e.g., substitution as a variable into an expression) to arrive at the answer
Data interpretation (DPr)	Data needs to be made sense of in order to answer the question. This includes: Transforming data from one representation to a different representation e.g., plotting a graph based on tabulated data or transformed between different scales, or describing pattern/relationship in words based on given graph. May involve math operations; Determining similarities, differences, or relationships within or between data sets. May involve math operations, OR Predicting relationship beyond given data range or with different conditions. Includes plotting/filling in data point in graph/table beyond given range, suggesting relationship based on partial data sets, and predicting outcomes.
Not applicable (NA)	Answer does not require any given data. Should only apply to item already coded as “NA” for Data Type.

FINDINGS AND DISCUSSION

On coding scheme reliability (RQ1), data reasoning codes achieved Cohen’s kappa of 0.57 ($p<.001$), indicating moderate agreement (Landis & Koch, 1977), while data representation type codes achieved simple agreement of 60.0%. On response process validity (RQ2), expected and enacted codes concurred on all but one item i.e., b(i) for both PSTs (Figure 1). For brevity, only two parts of the question are shown in Figure 1: a(i) provides the context to interpret b(i) which is discussed below. Note that Ken did not provide reasoning for b(ii); Leo did not answer c(ii).

For b(i), the authors opined interpreting (DPr) information (I) about the repeated experiment (“wind blowing over the wires”) as well as tables (T) for the repeated experiment and original experiment were necessary to answer the item. However, both PSTs only considered one data representation type in their written responses. Ken’s utterance “wind, what is the point of this? Oh, reduce overheating”—indicates he connected the information about blowing wind to the concept of heating/cooling (I; DPr), but his written response did not reference the wind. Instead, he said, “the resistance got smaller, for all the wires” and wrote a similar response. This suggests Ken interpreted the table for the repeated experiment (T; DPr) to answer b(i). Leo, however, initially interpreted the table for the repeated experiment, noting “the p.d. drops by 0.01” (T; DPr) compared to the original experiment. He then uttered “wind, could have cooled down the heat generated by the current” (I; DPr). Only the latter idea was included in Leo’s written response.

Part	Question Part Description	Data Interpretation		
		Expected	Enacted ( ) = enacted reasoning not reflected in written response. Discrepant codes are underlined.	
			Ken	Leo
a(i)	[Question stem shows picture of a piece of fuse wire clipped at both ends by metal clips. Length of wires X and Y are stated] Using data [table of current, p.d.s across wire X and wire Y]…describe the relationship between current in X and p.d. across X…for (1) low currents; (2) high currents.	T; DPr	(Pic; DIn) T; DPr G; DPr	(I; DPr) T; DPr G; DPr
b(i)	The experiment is repeated with a wind blowing over the wires. Table shows new readings at low currents [for 0.5A and 1.0A]. Suggest a reason why the values of p.d. at the same current are now lower than previously.	I, T; DPr	(I; DPr) T; DPr	(T; DPr) I; DPr

Figure 1. Expected and enacted codes for parts of 2015 DbQ.

CONCLUSION AND FUTURE WORK

Findings suggest proposed data sensemaking demands coding scheme demonstrated reasonable inter-rater reliability and response process validity when coding scheme was tested with a set of stimulus-based items in standardised physics assessments. To extend coding scheme's usability, it will be used to analyse stimulus-based items in biology and chemistry assessments. Findings will be shared at the conference. This work will be of interest to science education researchers and practitioners seeking a tool to analyse data sensemaking demands of assessment items, in order to help students, meet such assessment demands.

ACKNOWLEGMENTS

This work is funded through a startup grant SUG 09/18 OYS from the from the National Institute of Education, Nanyang Technological University, Singapore (NIE NTU, Singapore)

All views presented are that of the authors and do not reflect the views of NIE NTU, Singapore.

REFERENCES

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. Learning and Instruction, 16(3), 183–198.

LaDue, N. D., Libarkin, J. C., & Thomas, S. R. (2015). Visual representations on high school biology, chemistry, earth science, and physics assessments. Journal of Science Education and Technology, 24(6), 818–834.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159–174.

Morris, B. J., Masnick, A. M., Baker, K., & Junglen, A. (2015). An analysis of data activities and instructional supports in middle school science textbooks. International Journal of Science Education, 37(16), 2708–2720.

Singapore Ministry of Education. (2016). Physics: Singapore-Cambridge General Certificate of Education Ordinary Level (2018) (Syllabus 6091). Retrieved from https://www.seab.gov.sg/content/syllabus/olevel/2018Syllabus/6091_2018.pdf

van Someren, M. W., Barnard, Y. F., & Sanbery, J. A. C. (1994). The think aloud method: A practical guide to modelling cognitive processes. Information Processing & Management (Vol. 31). London: Academic Press.

ASSESSING ABSTRACTION SKILLS IN EARLY PRIMARY SCHOOL AMID ENVIRONMENTAL STUDY

Kalliopi Kanaki[1] *and Michail Kalogiannakis*[1]
[1]University of Crete, School of Education, Department of Preschool Education, Rethymno, Greece

This paper introduces part of a multifarious project on the development and the assessment of essential computational thinking skills, such as algorithmic thinking and abstraction, in early childhood education. More precisely, the part of the project that is presented refers to an assessment tool constructed by the authors, which focuses on evaluating the abstraction skills of first and second-grade primary school students. It employs data collection instruments and analysis techniques of mixed-method research and it is proposed to be applied in the classroom within the context of Environmental Study. Relevant research conducted in Greece with a sample of 435 students, from February to June 2019, is also discussed. The part of the research study we discuss focuses on evaluating students' abstraction skills and testing the relationship between abstraction skills and the levels of the content understanding of the Environmental Study course. The statistical analysis validated the correlation under investigation, supporting the innovative idea of the synergistic cultivation of computational thinking and environmental consciousness, and providing a robust background for relevant studies that will aim for attributing ecological essence to computational thinking activities.

Keywords: Computational thinking, science education, early childhood education

INTRODUCTION

Computational thinking (CT) is seen as a set of important competencies that are required in the modern digital era and, thus, several countries have already adopted priorities for its development from kindergarten to secondary education (Kanaki & Kalogiannakis, 2018; Relkin et al., 2021; Sung et al., 2017). At the same time, the demand for empowering environmental awareness emerges from the need for tackling or preventing contemporary environmental problems, such as pollution (Manisalidis et al., 2020), climate change (Celik, 2020; Lickley et al., 2020), ocean acidification (Doney et al., 2020), loss of biodiversity (Chawla, 2020), and ozone layer depletion (Lickley et al., 2020), to mention but a few. Within this context, we investigate the innovative idea of the synergistic development of environmental consciousness and CT skills during the early primary school years (Ardoin & Bowers, 2020; Kanaki & Kalogiannakis, 2019) providing a sturdy background for forthcoming research studies that will focus on proposing CT activities with ecological reflection.

Our study follows the principles of mixed methods research and focuses on first and second-grade primary school students. Its essence is to provide the educational community with novel means for cultivating CT competencies in the aforementioned age groups, together with a relevant assessment tool. As far as the proposed assessment tool is concerned, the parameters examined were (a) the validity and the reliability of the results obtained when applying the assessment tool, (b) the levels of fundamental CT skills of first and second-grade primary school students, and (c) the correlation of the levels of fundamental CT skills with the content comprehension of the Environmental Study course of first and second-grade primary school students. We also examined the correlation of fundamental CT skills with gender and the class of the students. In this paper, we discuss the part of our research that deals with the evaluation of abstraction, as one of the major CT skills, and its association with the educational achievements in the Environmental Study course. Thus, the research questions that drive this paper are (a) "Which are the levels of students' abstraction skills during the early years of primary school?" and (b) "Is there an association between abstraction skills of first and second-grade primary school students and their performance in Environmental Study course?"

THEORETICAL FRAMEWORK

To support young students to cultivate in-depth basic CT skills, the construction of relevant assessment tools is necessary (Grover, 2017; Román-González et al., 2019). Regarding the fundamental aspects of CT, although there is a variety of opinions (Tang et al., 2020) convergence is observed that abstraction is one of its essential elements (Barr et al., 2011; Shute et al., 2017; Wing, 2006).

Wing, who introduced the term of CT in 2006 (Wing, 2006), claimed that abstraction is the most noteworthy and advanced thought process among CT aspects (Wing, 2006, 2011). It is exercised for determining important properties of objects that are parts of a group and for hiding irrelevant distinctions among them, achieving the goal of an object to stand for many (Wing, 2011). Thus, the two faces of abstraction are (a) the act of not considering exceptional properties of an object in order to become representative for others of its kind, and (b) the process of generalizing by bringing out common properties of instances of an object (Kramer, 2007). For example, let's take the case of the hippopotamus which is a herbivorous mammal. In literature, there are few cases recorded that hippopotami consumed meat, indulged in cannibalism, or exercised predation (Dudley, 1998) although the anatomy of the hippopotamus' stomach does not justify eating meat, which is deviant behaviour and may occur in eating disorders (Eltringham, 1999). These few exceptions do not cause loss of generality. On the contrary, in the context of abstraction, the hippopotamus is characterized as herbivorous. Based on the above, we could say that the ability of an individual to identify exceptions that can be ignored without loss of generality can be a criterion for assessing their abstraction skills.

An important issue that arises is when the individual's abstract ability begins to develop. Due to Piaget's classical theory of children's cognitive development, abstraction begins to develop in the third stage i.e., the concrete operational, from about seven to 12 years old (Piaget, 1928). However, in his more recent work, these allegations seem to be overturned, as he presented in detail tests of the abstract ability of 18-month-old infants (Piaget, 2001).

METHOD

In order to answer the research questions, we conducted a research study adopting a coherent ethics framework, in the school year 2018 – 2019, on the island of Crete, in Heraklion city, employing a research sample of 435 primary school students of first and second grade. The sample was grade balanced – 218 first graders (50.11%) and 217 second graders (49.89%) – and gender balanced – 210 girls (48.28%) and 225 boys (51.72%).

As already mentioned, we employ mixed methodology for collecting data. For quantitative data, we suggest triangulation. In both of the methods proposed, the students have to assign the values of a specific attribute to a group of living organisms or inanimate objects.

The pillar of the first method is the digital platform PhysGramming, which we implemented for the needs of our research study. PhysGramming deploys a hybrid schema of text-based and visual programming techniques, with emphasis on object orientation and provides children the opportunity to create their own digital games (Kanaki & Kalogiannakis, 2018).

The second method is based on filling in worksheets. In both methods, the living organisms/ inanimate objects are depicted on the tool via pictures selected from a pool of pictures provided by the teacher. It is worth mentioning that the teacher who applies the assessment tool chooses its content, based on various criteria, such as the current course's unit, students' place of residence, etc. In other words, the teacher reconfigures the tool based on criteria that deem meaningful to apply.

At a qualitative level, personal semi-structured interviews are conducted, the content of which is relevant to the content of the current unit. The combined results of both quantitative and qualitative data place students into one of the four levels of abstraction: Excellent, Satisfactory, Medium and Basic. We came to these four levels based on the statistical analysis applied to the data of a relevant pilot study we conducted.

To understand the proposed evaluation method, let's see how it was applied in our research. Since the research was conducted in a region of Greece, we examined students' perceptions regarding the fauna of that region. We must mention that, during the presentation of the unit's content, the fauna of the region was not discussed. Thus, the results provided by the assessment tool were not related to the understanding of the unit's content.

We paid extra attention to the case of the crocodile since it has not been long since a crocodile was spotted in a lake of Crete. In addition, a popular family tavern in the region where the research took place hosted a crocodile until recently and for several years. Students should use their abstraction ability and not claim that the crocodile belongs to the region's fauna. If the tool is applied in another region of Greece or abroad, it does not make sense to check the students' abstraction skills by examining their perceptions regarding the fauna of the region where our research was conducted. Instead, the teacher should properly adjust the content of the assessment tool.

To estimate the levels of abstraction skills in quantitative terms, we analyse the data resulting from the indexing of the worksheets, as well as the log files provided by the digital platform that keep track of the students' achievements (Kanaki & Kalogiannakis, 2018). The fewer mistakes students make, the higher their level of abstract ability is. At a qualitative level, we analyse the data that emerge from the interviews.

As far as the assessment of the content understanding is concerned, students are assessed via relevant worksheets. In our research study, we examined students' perceptions about animals' eating habits. According to their performance, they were placed into one of the four levels of content understanding: Excellent, Very Good, Good and Almost Good. We came to these four levels based on the statistical analysis applied to the data of a relevant pilot study we conducted.

RESULTS

For processing, analysing and presenting the data derived from the application of the assessment tool, we employ several methods of statistical analysis. More precisely, we present frequency distributions that provide an introductory essence of the research's results. We construct contingency tables, implement chi-square tests and calculate p-values, in order to bring out possible correlations between independent and dependent variables. In the cases that we detect correlations, we calculate the odds ratio, in order to determine the probability of occurrence of the values of the dependent variable, regarding the independent variable. We finally employ machine learning (Nafea, 2018), as a supplementary method for statistical analysis, that provides clearly interpretable insights and reliable predictions regarding the possibility of new data being placed at each one of the levels of the variables under investigation (Rosé et al., 2019).

Frequency distribution

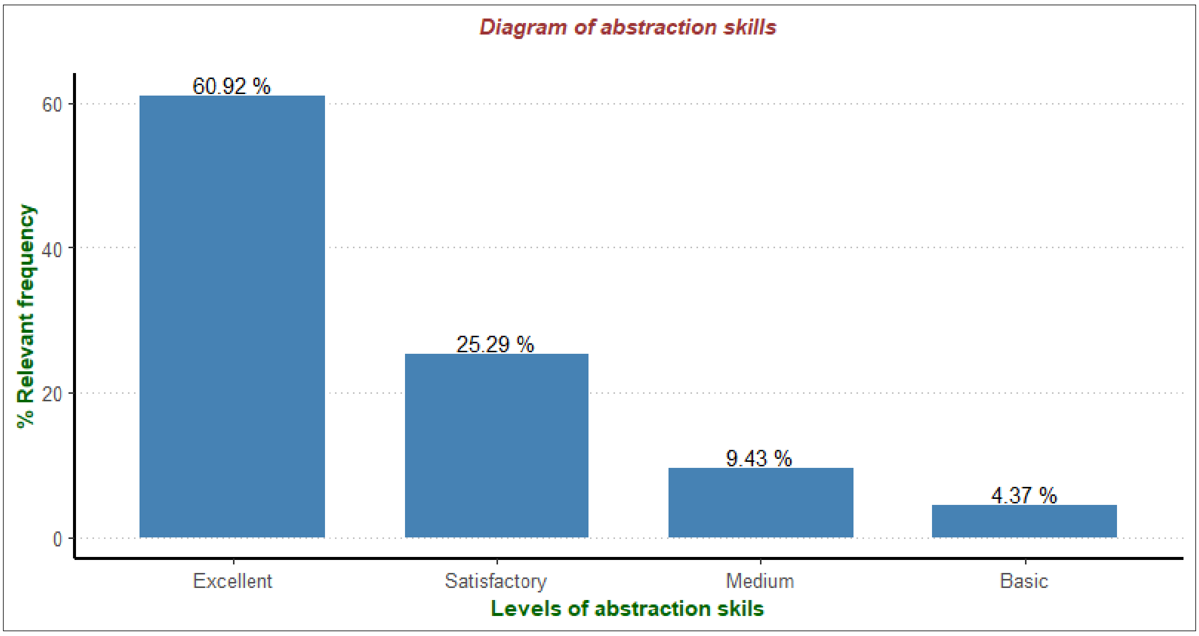

Figure 23. Frequency distribution

Figure 1 depicts the frequency distribution of the levels of abstraction skills detected in the research sample. We observe that 86.21% of the students demonstrated excellent or satisfactory abstraction skills.

Testing the correlation between abstraction skills and content understanding of the Environmental Study course

We set up the hypothesis that abstraction skills are not associated with the content understanding of the Environmental Study course. We test out the hypothesis set at a 5% level of significance and, thus, with the predetermined alpha level of significance to be 0.05. Based on the relevant contingency tables (Table 1, Table 2) chi-square is calculated to be 20.23, degrees of freedom are nine (df = 9) and the p-value is 0.01654. Since $p < 0.05$, we reject the hypothesis set and accept its alternative. Thus, we accept that there is an association between abstraction skills and content understanding.

Table 34. Contingency table of observed frequencies of the association between abstraction skills and content understanding

Content understanding / Abstraction skills	Excellent	Very good	Good	Almost good	Sum
Excellent	77	83	58	47	265
Satisfactory	21	34	29	26	110
Medium	5	12	9	15	41
Basic	4	11	3	11	19
Sum	107	130	99	99	435

Table 35. Contingency table of expected frequencies of the association between abstraction skills and content understanding

Content understanding / Abstraction skills	Excellent	Very good	Good	Almost good	Sum
Excellent	63.7191	83.3708	58.9551	58.9551	265
Satisfactory	26.4494	34.6067	24.4719	24.4719	110
Medium	9.8584	12.8989	9.1213	9.1213	41
Basic	6.9730	9.1236	6.4517	6.4517	19
Sum	107	130	99	99	435

Next, we examine the association of each one of the levels of abstraction skills with each one of the content understanding levels (Table 3). We calculate the odds ratio for every sub-table 2x2 (Table 4). In cases that the value of odds ratio is greater than one, the relevant levels of content understanding and abstraction skills are more likely to occur together than not. Results indicate that abstraction skills increase in relation to content understanding.

Table 36. Association of each one of the abstraction levels with the content understanding levels

			Abstraction skills							
			Excellent		Satisfactory		Medium		Basic	
			Yes	No	Yes	No	Yes	No	Yes	No
Content understanding	Excellent	Yes	77	30	21	86	5	102	4	103
		No	188	140	89	239	36	292	15	313
	Very good	Yes	83	47	34	96	12	118	1	129
		No	182	123	76	229	29	276	18	287
	Good	Yes	58	41	29	70	9	90	3	96
		No	207	129	81	255	32	304	16	320
	Almost good	Yes	47	52	26	73	15	84	11	88
		No	218	118	84	252	26	310	8	328

Table 37. Odds ratio for abstraction levels in relation to content understanding levels

			Abstraction skills							
			Excellent		Satisfactory		Medium		Basic	
			Yes	No	Yes	No	Yes	No	Yes	No
Content understanding	Excellent	Yes / No	1.92		0.65		0.42		0.8	
	Very good	Yes / No	1.20		1.06		0.91		0.17	
	Good	Yes / No	0.88		1.28		0.91		0.60	
	Almost good	Yes / No	0.49		1.09		2.25		5	

We also apply the machine learning method, seeking to predict the possibility of new data being placed at each one of the levels of abstraction skills, in dependence on the content understanding. Thus, the dependent variable is the level of students' abstraction skills, while the independent variable is the understanding of the course content. 80% of the survey data was used to construct the prediction equations and the remaining 20% for testing. The results obtained (Figure 2) verify the relation between the levels of abstraction skills and content understanding.

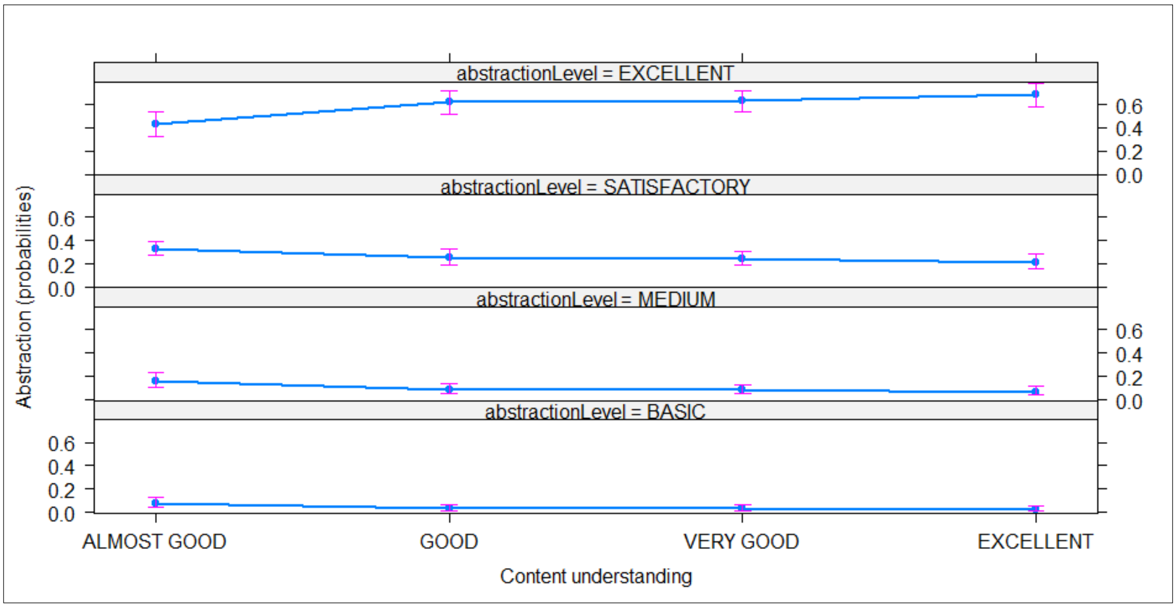

Figure 24. Predicting the probability of new data being placed at each one of the levels of abstraction skills, in dependence on the understanding of the course content

DISCUSSION AND PERSPECTIVES

Assessment is an important element of learning and pedagogy. It is necessary for evaluating students, teachers and educational policies. Our work focuses on evaluating skills related to the core aspects of CT in early primary school amid Environmental Study. We propose a relevant assessment tool, aiming to support the evaluation of CT competencies and, thus, to add to the field of CT development. The proposed tool not only detects the levels of CT skills of young students, but can also be used to facilitate the assessment of the effectiveness of related teaching interventions. It also contributes to the design of targeted teaching interventions for the cultivation of CT.

Furthermore, our study of investigating the potential correlation between fundamental CT skills, such as abstraction, and the content understanding of the Environmental Study course supports the contemporary social and educational demand of cultivating 21st century skills in compulsory education, even from the early years (Ardoin & Bowers, 2020; Kanaki & Kalogiannakis, 2019). Verifying this synergy facilitates the design of pertinent educational policies and the construction of pertinent educational activities, reinforces the concurrent development of the above-mentioned skills and provides a sturdy base for forthcoming studies that will stimulate the design of CT activities with ecological orientation. The innovative nature of our research study stems from the fact that there is no recorded study to investigate the correlation between fundamental CT competencies of first graders, such as abstraction skills, with their performance in the Environmental Study course. Moreover, the majority of pertinent surveys focus on older children (Alsancak, 2020; Kanaki et al., 2020), resulting in the under-exploration of the sensitive first school years.

It is worth mentioning that, trying to ameliorate the impact of the proposed assessment tool on students and teachers, we implemented PhysGramming to run not only on personal computers, but also on smart mobile devices (Kanaki, & Kalogiannakis, 2018). Although we have already tested the proposed assessment tool employing mobile technologies (Kanaki et al., 2022), we plan to further investigate the potential upgraded acceptance of the assessment tool when smart mobile devices take the place of personal computers.

CONCLUSIONS

In conclusion, in this paper, we reemphasize that a basic requirement for the effective cultivation of 21st century skills in K-12 is the construction of assessment tools that are developmentally appropriate for the target groups. Answering the research questions that drive this paper is of great importance to the field of developing CT as a set of skills, practices and concepts, which are relevant not only in studying computer science, but also in tackling multifarious problems computationally (Grizioti & Kynigos, 2021). Since it is of global interest to gear the upcoming generations up for the multifaceted active world problems (United Nations, 2018), it is vital to investigate how students may cultivate and utilize effectively CT competencies to grasp the concepts of contemporary social issues, argue about them and respond to their attendant challenges (Grizioti & Kynigos, 2021).

REFERENCES

Alsancak, D. (2020). Investigating computational thinking skills based on different variables and determining the predictor variables. *Participatory Educational Research*, *7*(2), 102-114. doi: 10.17275/per.20.22.7.2.

Ardoin, N. M., & Bowers, A. W. (2020). Early childhood environmental education: A systematic review of the research literature. *Educational Research Review*, 100353. doi: 10.1016/j.edurev.2020.100353.

Barr, D., Harrison, J., & Conery, L. (2011). Computational thinking: A digital age skill for everyone. *Learning & Leading with Technology*, *38*(6), 20-23.

Celik, S. (2020). The effects of climate change on human behaviors. In *Environment, Climate, Plant and Vegetation Growth* (pp. 577-589). Springer, Cham. doi: 10.1007/978-3-030-49732-3_22.

Chawla, L. (2020). Childhood nature connection and constructive hope: A review of research on connecting with nature and coping with environmental loss. *People and Nature*, *2*(3), 619-642. doi: 10.1002/pan3.10128.

Doney, S. C., Busch, D. S., Cooley, S. R., & Kroeker, K. J. (2020). The impacts of ocean acidification on marine ecosystems and reliant human communities. *Annual Review of Environment and Resources*, *45*, 83-112. doi: 10.1146/annurev-environ-012320-083019.

Dudley, J. P. (1998). Reports of carnivory by the common hippo Hippopotamus Amphibius. *South African Journal of Wildlife Research-24-month delayed open access*, *28*(2), 58-59.

Eltringham, S. K. (1999). *The hippos: natural history and conservation*. Princeton University Press.

Grizioti, M., & Kynigos, C. (2021). Children as players, modders, and creators of simulation games: A design for making sense of complex real-world problems: Children as players, modders and creators of simulation games. In *Interaction Design and Children* (pp. 363-374). doi: 10.1145/3459990.3460706.

Grover, S. (2017). Assessing algorithmic and computational thinking in K-12: Lessons from a middle school classroom. In *Emerging research, practice, and policy on computational thinking* (pp. 269-288). Springer, Cham. doi: 10.1007/978-3-319-52691-1_17.

Kanaki, K., & Kalogiannakis, M. (2019, November). Assessing Computational Thinking Skills at First Stages of Schooling. In *Proceedings of the 2019 3rd International Conference on Education and E-Learning* (pp. 135-139). doi: 10.1145/3371647.3371651.

Kanaki, K., & Kalogiannakis, M. (2018). Introducing fundamental object-oriented programming concepts in preschool education within the context of physical science courses. *Education and Information Technologies*, *23*(6), 2673-2698. doi: 10.1007/s10639-018-9736-0.

Kanaki, K., Kalogiannakis, M., Poulakis, E., & Politis, P. (2022). Employing mobile technologies to investigate the association between abstraction skills and performance in environmental studies in early primary school. *International Journal of Interactive Mobile Technologies (iJIM).* (under publication).

Kanaki, K., Kalogiannakis, M., & Stamovlasis, D. (2020). Assessing Algorithmic Thinking Skills in Early Childhood Education: Evaluation in Physical and Natural Science Courses. In *Handbook of Research on Tools for Teaching Computational Thinking in P-12 Education* (pp. 104-139). IGI Global. doi: 10.4018/978-1-6684-2411-7.ch024.

Kramer, J. (2007). Is abstraction the key to computing?. *Communications of the ACM*, *50*(4), 36-42. doi: 10.1145/1232743.1232745.

Lickley, M., Solomon, S., Fletcher, S., Velders, G. J., Daniel, J., Rigby, M., ... & Stone, K. (2020). Quantifying contributions of chlorofluorocarbon banks to emissions and impacts on the ozone layer and climate. *Nature communications*, *11*(1), 1-11. doi: 10.1038/s41467-020-15162-7.

Manisalidis, I., Stavropoulou, E., Stavropoulos, A., & Bezirtzoglou, E. (2020). Environmental and health impacts of air pollution: a review. *Frontiers in public health*, *8*, 14. doi: 10.3389/fpubh.2020.00014.

Nafea, I. T. (2018). Machine learning in educational technology. In *Machine learning-advanced techniques and emerging applications* (pp. 175-183). Hamed Farhadi, IntechOpen. doi: 10.5772/intechopen.72906.

Piaget, J. (2001). *Studies in reflecting abstraction*. United Kingdom: Psychology Press.

Piaget, J. (1928). *Child's conceptions of the world.* Routledge & K. Paul, London. Retrieved from https://openlibrary.org/. Accessed January 18, 2022.

Relkin, E., de Ruiter, L. E., & Bers, M. U. (2021). Learning to code and the acquisition of computational thinking by young children. *Computers & Education*, *169*, 104222. doi: 10.1016/j.compedu.2021.104222.

Román-González, M., Moreno-León, J., & Robles, G. (2019). Combining assessment tools for a comprehensive evaluation of computational thinking interventions. In *Computational thinking education* (pp. 79-98). Singapore: Springer. doi: 10.1007/978-981-13-6528-7_6.

Rosé, C. P., McLaughlin, E. A., Liu, R., & Koedinger, K. R. (2019). Explanatory learner models: Why machine learning (alone) is not the answer. *British Journal of Educational Technology*, *50*(6), 2943-2958. doi: 10.1111/bjet.12858.

Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review*, *22*, 142-158. doi: 10.1016/j.edurev.2017.09.003.

Sung, W., Ahn, J., & Black, J.B. (2017). Introducing Computational Thinking to Young Learners: Practicing Computational Perspectives Through Embodiment in Mathematics Education, *Technology, Knowledge and Learning, 22*(3), 443-463. doi: 10.1007/s10758-017-9328-x.

Tang, X., Yin, Y., Lin, Q., Hadad, R., & Zhai, X. (2020). Assessing computational thinking: A systematic review of empirical studies. *Computers & Education*, *148*, Article 103798. doi: 10.1016/j.compedu.2019.103798.

United Nations (2018). Sustainable Development Goals. Retrieved from https://unstats.un.org/sdgs/files/report/2018/TheSustainableDevelopmentGoalsRseport2018-EN.pdf. Accessed January 25, 2022.

Wing, J. (2011). Research notebook: Computational thinking—What and why. *The Link Magazine*.

Wing, J. M. (2006). Computational thinking. *Communications of the ACM, 49*(3), 33-35. doi: 10.1145/1118178.1118215.

THE IMPACT OF UNDERGRADUATE STUDENTS' ACADEMIC EXPERIENCE ON THEIR MOTIVATION TO LEARN CHEMISTRY AND PHYSICS

Katerina Salta[1] *and Dionysios Koulougliotis*[2]
[1]National and Kapodistrian University of Athens, Athens, Greece
[2]Ionian University, Zakynthos, Greece

This study aimed at investigating the impact of students' academic experience on their motivation to learn chemistry and physics in combination with the influencing factor related to students' academic major. A clustered sampling was used to select 281 participants who were undergraduate students attending three different academic majors (Environmental Technology, Conservation of Cultural Heritage, Food Technology) at a Greek tertiary education institution. Statistical analyses showed that students' motivation to learn chemistry and physics depends strongly on their academic major. Mediation analysis provided strong evidence that the influence of the academic major on motivation is mediated by the students' experience from the attendance of physics and chemistry courses both at secondary and tertiary education. The most recent academic experience was shown to exert a significantly more intense effect in shaping students' motivation to learn chemistry and physics relative to the one from the more distant past.

Keywords: Motivation, Higher Education, Quantitative methods

INTRODUCTION

The role of motivation in science learning is considered crucial by science educators (Koballa & Glynn, 2007) and the influence of several factors on students' motivation to learn various science disciplines (e.g., age, gender, school factors, parental support) is a field of active investigation (e.g., Britner & Pajares, 2006; Pintrich, 2003; Salta & Koulougliotis, 2015, 2020; Vedder-Weiss & Fortus, 2013). In an earlier study (Salta & Koulougliotis, 2020), undergraduate students' motivation to learn chemistry and physics were found to strongly depend on their academic major. Considering the role of several school factors, which establish a unique experience for each student, in influencing students' motivation, this study is an investigation of possible relations of students' motivation with their perceptions about academic experience. More specifically, the aim of this study is the statistical testing of the hypothesis that students' perceptions about their academic experience in both high-school and university science (physics and chemistry) courses could mediate the effect of academic major on students' motivation to learn chemistry or physics.

THEORETICAL BACKGROUND

The theoretical background of our studies is the social-cognitive theory developed by Bandura (2001) and extended by others (e.g., Pintrich, 2003). Within this theoretical perspective, motivation is defined as "an internal state that arouses, directs, and sustains science-learning behavior" (Glynn et al., 2011, p. 1160), and is conceptualized as a multidimensional construct.

Several motivational constructs associated with science learning, such as self-efficacy, self-determination, intrinsic motivation, career motivation and grade motivation, have been identified, thoroughly studied, and reviewed (Bryan et al., 2011; Glynn et al., 2011; Koballa & Glynn, 2007; Pintrich, 2003). Self-efficacy (SE) refers to students' beliefs with regard to their ability to do well in science tasks; self-determination (SD) is students' beliefs with regard to the control they have over their learning science; intrinsic motivation (IM) refers to the inherent satisfaction derived from learning science; and extrinsic motivation, which may be delineated to career motivation (CM) which refers to students' expectations for a science related career and grade motivation (GM) which is related to students' goals for high academic performance (Glynn et al., 2011; Koballa & Glynn, 2007; Pintrich, 2003).

Social cognitive perspectives also propose that students form their motivational beliefs by interpreting experience from their school environment, which includes, among others, the influence of the teacher, the school's specific goals, purposes and values, and students' perceptions of them (e.g., Britner & Pajares, 2006; Vedder-Weiss & Fortus, 2013). In this respect, there is a need for a deeper understanding of the role of students' academic experience on their motivation to learn various science-related subjects (Ardura et al., 2021; Pintrich, 2003).

METHOD

Participants

The participants in the current study were 281 full-time undergraduate students drawn from three different academic departments of a small-sized regional tertiary education institution (TEI) located in Western Greece, offering Bachelors' degrees in the following academic majors: Environmental Technology (ET, N=95), Conservation of Cultural Heritage (CCH, N=93) and Food Technology (FT, N=93).

Measures

Motivation to learn physics and chemistry: Students' motivation to learn physics and chemistry was measured by the Greek version (Salta & Koulougliotis, 2015) of the discipline-specific original instrument known as SMQ II (Glynn et al., 2011). The reliabilities (internal consistencies) of all scales, assessed by Cronbach's α, have values larger than 0.81, indicating a good overall reliability according to DeVellis (2003).

Experience from the attendance of physics and chemistry courses: Students reported their experiences from i) the textbook, ii) the teacher, iii) the laboratory and iv) the course in total, with 4 items referring to high school and 4 referring to TEI experience for each discipline course (physics or chemistry), using a Likert-type scale ranging from 1 (very negative) to 4 (very positive). Therefore, a total of sixteen experience items was made up.

A principal components exploratory factor analysis was independently conducted on each set of 8 items (one for the physics and one for the chemistry course) with varimax rotation in order to examine the number of dimensions existing among the items-variables related to each discipline course. Each analysis resulted in two factors with eigenvalues over Kaiser's criterion of 1 for each discipline course. These two factors, in combination, explained 66.23% and 68.19% of the variance for the chemistry and physics course, respectively. (Table 1) The items that cluster on the same factor suggest that factor 1 represents experience from high school courses while factor 2 represents experience from TEI course. The experiences from TEI physics, high school physics, TEI chemistry and high school chemistry courses all had high reliabilities; all Cronbach's α have values larger than 0.81, indicating a good overall reliability according to DeVellis (2003).

TABLE 1: Exploratory factor analysis for students' experiences

Items	**Rotated factor loadings**	
How would you characterize your experience …	Experience from TEI physics courses	Experience from high school physics courses
… from the TEI physics labs?	**.826**	.116
… from TEI physics courses overall?	**.819**	.178
… from the teachers who taught you physics at TEI?	**.792**	.190
… with TEI physics textbooks?	**.791**	.211
… high school physics courses overall?	.206	**.835**
… from the teachers who taught you physics at high school?	.053	**.834**
… with high school physics textbooks?	.256	**.780**
… from the high school physics labs?	.201	**.756**
Eigenvalues (Variance)	3.854 (48.179)	1.601 (20.006)
Cronbach's α	.845	.838
How would you characterize your experience …	Experience from high school chemistry courses	Experience from T.E.I. chemistry courses
… from high school chemistry courses overall?	**.856**	.178
… from the teachers who taught you chemistry at high school?	**.805**	.041
… with high school chemistry textbooks?	**.799**	.175
… from the high school chemistry labs?	**.780**	.099
… from the teachers who taught you chemistry at TEI?	.017	**.834**
… from TEI chemistry courses overall?	.237	**.827**
… from the TEI chemistry labs?	.103	**.775**
… with TEI chemistry textbooks?	.135	**.729**
Eigenvalues (Variance)	3.443 (43.039)	1.855 (23.186)
Cronbach's α	.836	.813

Note: Factor loadings over .40 appear in bold

RESULTS

Paired-samples t-tests showed that the mean score for students' experience from high school was lower than their experience from TEI for both physics and chemistry courses. Moreover, students reported more positive experience from the attendance of chemistry relative to physics courses in either high school or TEI. The earlier study conducted on the same sample (Salta & Koulougliotis, 2020) had showed that students majoring in CCH are differentiated from those in the other two majors by possessing lower motivational scores in all five measured motivational scales which were related to physics learning (namely IM. SE, SD, CM and GM) and three scales (namely IM, SE and GM) related to chemistry learning. That study had also showed that students majoring in FT are differentiated from those of the other two majors by exhibiting higher motivation in the scales of CM and SD related to chemistry learning. Our hypothesis that students' experience could be mediating the effect of the academic major on their motivation to learn either chemistry or physics was tested via the use of the PROCESS variable path analysis modeling tool of SPSS (Hayes, 2013). In this analysis, the antecedent variable X (academic major) is modeled as influencing the consequent variable Y (motivation to learn chemistry or physics) directly as well as indirectly through the following two mediators: students' experience from high school physics or chemistry courses (mediator M_1) and students' experience from TEI physics or chemistry courses (mediator M_2). The conceptual diagram of this model is shown in Figure 1. Statistical inference for estimating the difference of the evaluated statistical coefficients from a zero value is conducted via examination of *p*-values (for the coefficient c' of the direct effect) and via bootstrap confidence intervals (for coefficients a_1b_1, a_2b_2 and $a_1b_1+a_2b_2$ of the specific and total indirect effects) at the 95% confidence level.

The analysis showed a statistically significant direct effect of the academic major on eight (out of 10 in total) motivational components concerning chemistry and physics learning (all five related with chemistry and three related with physics namely IM, SE and GM). In addition, the total indirect effect of the academic major on motivation (combination of M_1 and M_2), is statistically significant for all 10 motivational components. An examination of the specific indirect effects shows that the one concerning the experience from high school courses (mediator M_1) is statistically significant for six (out of 10 in total) motivational components (physics and chemistry IM, SE and GM), while the one concerning the experience from TEI courses (mediator M_2) is statistically significant for all 10 motivational components. It is thus deduced that experience from high school and TEI courses collectively mediate the effect of the academic major on students' intrinsic motivation, self-efficacy, and grade motivation to learn physics and chemistry. On the other hand, the effect of the academic major on students' self-determination and career motivation (for both physics and chemistry) is mediated solely by the experience from TEI courses.

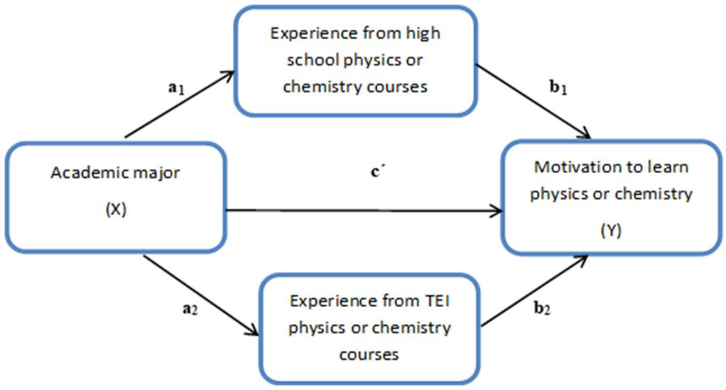

Figure 1. The conceptual diagram of the parallel multiple mediator model with two mediators.

CONCLUSIONS

The current study provided strong evidence that the previously documented domain specificity of motivation – exhibited via an increased motivation for chemistry vs physics learning (Salta & Koulougliotis, 2020) – was accompanied by a similar domain specificity of students' experiences as well. Thus, the undergraduate students report more positive experiences from the attendance of chemistry-related (vs physics-related) courses both in secondary as well as in tertiary education. This trend was stronger in the more recent (i.e., from TEI) relative to the more distant (i.e. from high school) experiences. The domain specificity of motivation depends not only on the specific motivational construct but on the academic major of the students as well (Salta & Koulougliotis, 2020). In fact, pursuing studies in an academic major is related to a distinct set of academic characteristics related with the curriculum content and often with the corresponding student. More specifically, the curricula of all three majors include courses which are related with both physics and chemistry but to a different extent. Thus, the majors CCH and FT contain more courses related to chemistry relative to physics with the ratio of chemistry-related vs physics-related ECTS units being ca. 2.4 and 3.8 for CCH and FT, respectively. In contrast, the curriculum of the ET major is almost equilibrated between chemistry and physics-related courses with the latter being slightly more by ca. 15%. In addition, similarly to the motivation measurements, the domain specificity of students' experiences was shown to be differentiated among the three academic majors as well, with it being more intense among the students of the CCH majors. Thus, CCH majors report more positive experiences in chemistry (vs physics) courses both in TEI and in high school, while the FT and ET majors report statistically similar high school experiences between the two courses and more positive TEI experiences in chemistry relative to physics.

Our hypothesis that the relationship between students' academic major and their motivation to learn either chemistry or physics is mediated by their experiences was confirmed. Taking into account that school influence on students' motivation is multidimensional and involves, among others, social and instructional dimensions (Wigfield et al., 2015), the tested factor has been the experiences of the students from the attendance of chemistry and physics-related courses either in high school or during their studies at TEI. This finding provides evidence that that the measured students' motivations for chemistry and physics learning are not simply a result of the differential student composition in each major. In addition, the documented more intense mediation effect of the most recent experience (TEI vs high school courses), supports the notion that, irrespective of the students' background, the provision of higher quality academic experiences in the present time could enhance their motivation for chemistry and physics learning.

The main limitation of the present study is related to the fact that the inference of the results is based solely on quantitative self-report data. However, its findings may provide the basis for continuing research in the field of motivation to learn science. The observed difference between undergraduate students' motivation to learn chemistry and physics and its connection with students' experiences from the attendance of the respective courses could be tested among students in different age groups starting from lower secondary school preferably in a study that follows a longitudinal research design. Examination of the temporal evolution of the domain specificity of motivation and its influencing factors would provide new knowledge useful to researchers, education practitioners and educational policy makers.

REFERENCES

Ardura, D., Zamora, Á., & Pérez-Bitrián, A. (2021). The role of motivation on secondary school students' causal attributions to choose or abandon chemistry. *Chemistry Education Research and Practice, 22*, 43-61.

Bandura, A. (2001). Social cognitive theory: An agentive perspective. *Annual Review of Psychology*, 52, 1-26.

Britner, S. L., & Pajares, F. (2006). Sources of science self-efficacy beliefs of middle school students. *Journal of Research in Science Teaching, 43*, 485-499.

Bryan, R. R., Glynn, S. M., & Kittleson, J. M. (2011). Motivation, achievement, and advanced placement intent of high school students learning science. *Science Education*, 95, 1049-1065.

DeVellis, R. F. (2003). *Scale development: Theory and applications*. (2nd ed.), Thousand Oaks, CA: Sage.

Glynn, S. M., Brickman, P., Armstrong, N., & Taasoobshirazi, G. (2011). Science motivation questionnaire II: Validation with science majors and nonscience majors. *Journal of Research in Science Teaching, 48*, 1159-1176.

Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. New York: The Guilford Press.

Koballa, T. R., Jr., & Glynn, S. M. (2007). Attitudinal and motivational constructs in science learning. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education*. (pp. 75–102). Mahwah, NJ: Lawrence Erlbaum Associates.

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology, 95*, 667-686.

Salta, K., & Koulougliotis, D. (2015). Assessing motivation to learn chemistry: adaptation and validation of Science Motivation Questionnaire II with Greek secondary school students. *Chemistry Education Research and Practice, 16,* 237-250.

Salta, K., & Koulougliotis, D. (2020). Domain specificity of motivation: chemistry and physics learning among undergraduate students of three academic majors. *International Journal of Science Education*, *42*, 253-270.

Vedder-Weiss, D., & Fortus, D. (2013). School, teacher, peers, and parents' goal emphases and adolescents' motivation to learn science in and out of school. *Journal of Research in Science Teaching, 50*, 952-988.

Wigfield, A., Eccles, J. S., Fredricks, J. A., Simpkins, S., Roeser, R. W., & Schiefele, U. (2015). Development of achievement motivation and engagement. In M. E. Lamb & R. M. Lerner (Eds.), *Handbook of child psychology and developmental science: Socioemotional processes* (pp. 657-700). Hoboken, NJ, US: John Wiley & Sons Inc.

ANALYSIS OF TEACHER'S WRITTEN FEEDBACK IN BIOLOGY INQUIRY TASK WITH STRESS ON AFFECTIVE CONNOTATIONS

Lukas Rokos[1]

[1] University of South Bohemia in České Budějovice, Faculty of Education, Department of Biology, České Budějovice, Czech Republic

This article focuses on the analysis of written feedback provided by teachers on the student's solution to a human biology inquiry task during distance teaching in Covid-19 lockdown. The analysis monitored whether the written feedback was formative, contained affective connotations, and an explanation of the evaluation criteria. In total, four classes of students (N = 60) and four teachers from three selected schools participated in this study. Teachers' feedback was analyzed using the coding tool developed and verified within the ASSIST-ME project. The analysis showed that the teachers tried to make their comments formative, explained the evaluation criteria repeatedly to the students, and justified why the student's solution is incorrect. The affective connotations were rather exceptional in the feedback provided. Teachers stated that it was difficult to write a comment that was easy to understand for the students and that they would prefer to provide oral feedback where they could clarify everything with the students immediately. Teachers who provided feedback that contained affective connotations also added that they would like to motivate students with these positive comments.

Keywords: formative assessment, emotion, feedback

THEORETICAL BACKGROUND

Assessment and Formative Feedback

Assessment is a key competency of a teacher, and the importance of formative assessment has increased step by step since the end of the twentieth century (Black & Wiliam, 1998). Especially in light of the situation associated with distance teaching, it is often necessary to use continuous assessment of student performance and provide them with adequate feedback. The level and processing of the feedback provided are a key aspect for the further development of the student's learning. Teachers are encouraged to provide formative feedback to students, but the question is whether this feedback really carries aspects of formative assessment. It means that it gives students guidance on how to continue their learning in order to achieve the set goals and eliminate the identified imperfections. Appropriate formative feedback allows students to focus on learning difficulties and thus facilitate their future correction (Wiliam & Leahy, 2015).

Feedback is not in itself a formative assessment, and for which it is often mistaken for confusion. The aim of feedback as a basis for formative assessment is to overcome the gap between the current level of student knowledge or skills and the level that the teacher has set as a desirable goal (Hattie & Timperley, 2007). Feedback provides teachers with information that helps guide their further teaching. Students receive information on how their previous learning is progressing (Nicol & Mcfarlane-Dick, 2006).

The effectiveness of feedback is influenced by a number of factors, such as the student's learning style, the teacher's person, the classroom climate, the teaching methods and forms used, as well as the student's language level and knowledge and skills (Evans et al., 2016). In order to be able to use formative assessment effectively in teaching, the teacher must have a clear idea of why he / she uses this assessment and what its main purpose is. The teacher's actions in a given situation depend not only on his pedagogical and professional knowledge (Magnusson, Krajcik, & Borko, 1999) but also on the specific situation that occurred during teaching (Cross & Lepareur, 2015).

Emotions and Affective Connotations

The role of emotions in assessment and feedback is less explored than the cognitive and motivational components, although affective connotations could be a crucial part of feedback because they can affect the way students use this feedback (Rowe, 2017). If students feel negative emotions from the feedback of the assessor, there could be a higher rejection rate of the offered help (Ryan & Henderson, 2018). In contrast, the feeling of positive emotions leads to a sense of success and can also lead to greater student effort in other tasks.

Research has identified several emotions that can affect the learning process, such as joy, hope, pride, anger, anxiety, shame, hopelessness, boredom, or interest (Rowe, Fitness & Wood, 2015; White, 2013). Furthermore, positive emotions have been shown to generally improve learning and achievement, especially in the areas of metacognition, focus on performance, or motivation (Pekrun & Stephens, 2010). Negative emotions, while in many cases impairing motivation and performance, can also be beneficial (Pekrun & Stephens, 2010). Psychological research shows that different words can trigger specific emotional responses (Abbasi et al., 2015) and that emotions and language expressions are strongly linked (Lindquist et al., 2015). When communicating, nonverbal channels can be used to emphasize emotions, facial expressions, postures, or body language (Niedenthal et al., 2003), which can be used to provide oral feedback. However, in written feedback, care must be taken with abstract words that tend to have more emotional characteristics than specific words (Meteyard et al., 2012). Talmi (2013) presents the connection between remembering a given moment and its emotional aspect. Huntsinger (2013) adds that attention can be increased by positive emotions, such as happiness, but reduced by a feeling of anxiety.

METHODOLOGY

The presented study is part of larger project (reg. No. TL02000368) aimed at promoting the use of formative assessment in inquiry-based teaching of science and mathematics. This research was carried out during the first wave of covid-19 lockdown in the Czech Republic (in spring 2020), when the way of teaching had changed from face-to-face teaching to distance teaching and the evaluation of student results was shifted to the level of written feedback mainly.

The study focuses on the following research questions:

1) To what extent does the feedback provided contain formative aspects?

2) To what extent do teachers use subjective sentences and emotional expressions in the feedback provided?

3) How does the feedback from an individual teacher differ?

Data Collection and Sample

Students solved the human biology inquiry task and carried out their own experiment. In total, 4 eighth grader classes from 3 selected schools (N = 60) and 4 teachers participated in this study (they will be more characterized in the results). As mentioned, the study took place during distance learning in spring 2020, and students completed a protocol related to the inquiry task, teachers provided them with structured written feedback on their performance.

The study also included semi-structured interviews with the four teachers, in which they reviewed the assessment process and discussed the quality of the student's protocols, provided feedback, but also described how the students received their feedback and subsequently used it to resubmit their protocols. The duration of the interview ranged from 30 to 50 minutes and was preceded by obtaining oral consent to record their responses to the recorder and their use in the research study.

Data Analysis

Teacher feedback was analyzed with a coding tool that was developed and validated within the ASSIST-ME project (Tab. I; Constantinou & Papadouris, 2016). The tool evaluated the following dimensions of feedback: emotional side, completeness and weight of comments from the teacher, dependence on the specified criteria, justification of the given comments, and level of guidance for the next steps in the learning process of the students.

Table 1. Illustrative examples from the coding tool (Constantinou & Papadouris, 2016).

<table>
<tr><td colspan="5">To what extent does the feedback carry connotations related to affective aspects? Specify the frequency of statements (e.g., sentences or segments of sentences) within the feedback comments that carry affective connotations.</td></tr>
<tr><td colspan="5">a) instances that were likely to serve as encouraging feedback but were restricted to vague phrases (e.g., excellent/bravo, etc.).</td></tr>
<tr><td colspan="5">b) instances that were likely to serve as encouraging feedback by explicitly identifying (and providing credit for) what the student has achieved</td></tr>
<tr><td colspan="5">c) instances that were likely to serve as discouraging feedback and were formulated in a vague manner (e.g., you have not put much effort, this is substandard, this was not a good response, etc.).</td></tr>
<tr><td colspan="5">d) instances that were likely to serve as discouraging feedback and were formulated in a specific manner that identifies what the student has failed to do.</td></tr>
<tr><td colspan="5">Overall, what is the likely affective impact of the feedback comments on the student taking into consideration other data sources, e.g., post-instructional interviews, observations? Select the level of the Likert scale that you deem more appropriate.</td></tr>
<tr><td>1
It is likely to have a very negative impact</td><td>2
It is likely to have a rather negative impact</td><td>3
Neutral
(any affective impact)</td><td>4
It is likely to have a rather positive impact</td><td>5
It is likely to have a very positive impact</td></tr>
</table>

The audio recordings taken from the interviews were transcribed into a text editor. Two researchers then independently identified interesting and important statements in the interviews (one of the coders was not involved in the research, so his view could not be affected). Their selections were compared, and after the discussion, the most important messages were included in the analysis.

RESULTS

The results will be presented in relation to individual research questions.

RQ1: To what extent does the provided feedback contain formative aspects?

Teachers in depth (84.6 %) or carefully (10.4%) related their assessment to set criteria for evaluation of student work. They mostly mentioned that students should correct or add more information to their protocols to better achieve the goal. In 64.4 % of the cases, the items contained such advice, in 25.4 % of the remaining items the protocol was correct, and therefore there was no need to provide the advice. However, it was very interesting how the teacher gave this advice. In some cases, the teacher directly wrote what should be corrected or added, so it can be said that this feedback was not very formative because the student only followed the instructions from the teacher and wrote the information in his / her protocol. In other cases, the teacher tried to guide the students to the new solution through instructional questions or statements that lead the students to a new solution. This way could be seen as the formative one.

At the same time, it was investigated on which other aspects teachers focused their assessment. Most often (63.4 %) they mentioned the level of the detail of information in the student's protocols (the need to expand, supplement, etc.). Furthermore, they focused their comments on the correct description of the procedure or the use of the correct terms (23.8 %). They were less focused on the text quality in terms of grammar (8.7 %) and product-related aspects, e.g., graph size, text structure, and way of recording data (4.4 %).

RQ2: To what extent do teachers use subjective sentences and emotional expressions in the feedback provided?

In total, 82 subjective statements appeared in all feedbacks analyzed (N = 480). Affective connotations were found in 77 cases. 42.1 % of these statements represented the feedback which supported and accurately described and highlighted what the students had already achieved. 26.3 % of the statements were vague phrases such as 'excellent', 'bravo', etc. Affective connotations with negative meaning were found in 25 cases (5.2 %). 13 of them were again vague phrases (e.g., 'you slacked it off', 'this is not correct', etc.) and 12 gave information where the students made a mistake only.

The impact of teacher's comments on students was neutral in most cases (68.8%), 13.3% of the comments were rather positive and 8.8 % were very positive. There were only four comments with a very negative estimated impact and 8.3% of the comments with a rather negative impact.

RQ3: How does the feedback from an individual teacher differ?

Teacher A has 13 years of experience and is also familiar with formative assessment. She participated in several projects that focused on the use of formative assessment, namely, the use of written feedback and peer assessment. She tries to use formative assessment in her teaching often, most often in oral form. She creates various materials and plans when and in what contexts she will use formative assessment, e.g., self-assessment or peer-assessment. However, she adds that she most often performs on-the-fly assessments. She most often used subjective sentences (46.7 %) and 25.8 % of all comments included affective connotations. Therefore, the feedback was very personal, and it also contains emotional charges.

Teacher B has 1.5 years of experience and is therefore a beginning teacher. She does not have much experience with formative assessment, she knows all the information from her training at the university. She states that she is trying to incorporate formative assessment, but for the beginner, it is relatively difficult for now because she has to focus on many other aspects related to teaching. She finds it easier to use the on-the-fly assessment, so she tries to incorporate it into her teaching. She also adds that the information gained from the students helps her better plan the next lessons. Teacher B used subjective sentences in 21.9 % of cases and in almost a third of all comments, this teacher wrote affective connotations (31.3 %).

Teacher C has 21 years of experience; we can call her an experienced teacher. She encountered formative assessment in the context of professional development courses, but personally this approach did not interest her much. She states that she is used to providing support in her style, which she is not going to change. She believes that she simply provides the necessary information to the students in her comments. Teacher C used subjective sentences very rarely (2.7 %), but there were emotional-themed statements in 15.2 % of the comments, which, unlike previous teachers, were mostly negative.

Teacher D has 4 years of experience and can also be described as a beginning teacher. She met with a formative assessment while studying at the university. Her bachelor and diploma theses dealt with this issue. Even as a beginning teacher, she tries to use as much formative assessment as possible, creates various materials, and searches for other resources. Teacher D barely used even subjective sentences (1.3 %) or affective connotations (0.7 %).

The character of the statements provided by the individual teachers is summarized in Figure 1.

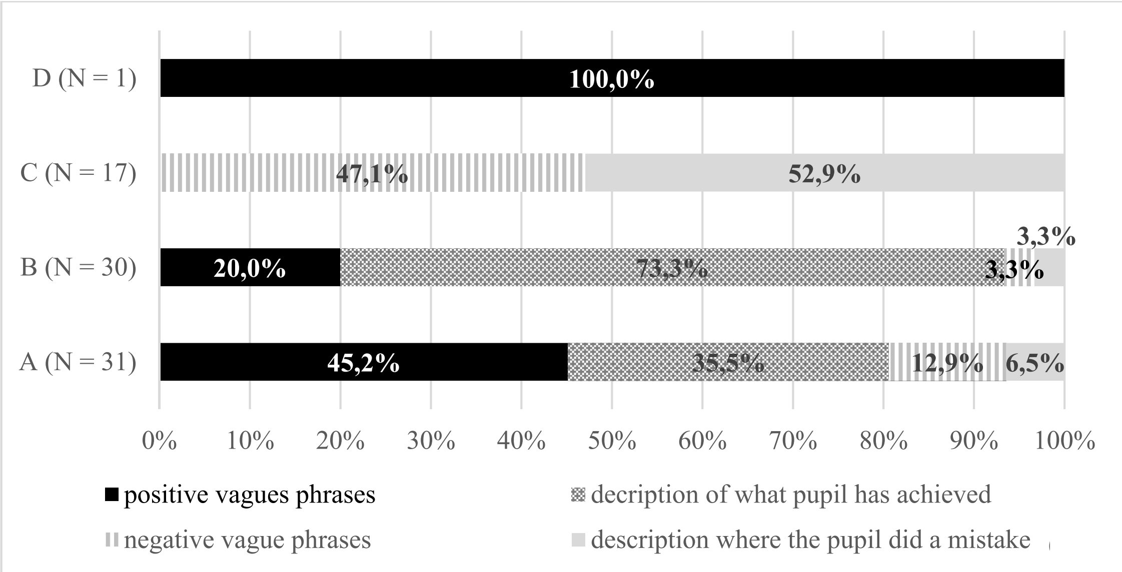

Figure 1. The character of affective connotations of the individual teacher.

The aspects on which the teachers focused their feedback were also identified. As can be seen in Figure 2, the teachers focused mainly on the level of detail provided. Teacher A also very often evaluated the visual aspect in a positive way. She praised the students for how they presented the results of the experiment, how they recorded the data, etc. Teacher C also focused on the visual aspect, but she looked for errors. She criticized the arrangement of the protocol. Teacher C also focused on the quality of the text, blaming students for grammatical errors or missing commas in sentences in the procedure description. In two cases, teacher D warned the student to read the text and correct some of the grammatical errors, but without the intention of harming them. Teacher B focused explicitly on content related to the field of human biology and did not care about grammar or visual aspects of the protocol.

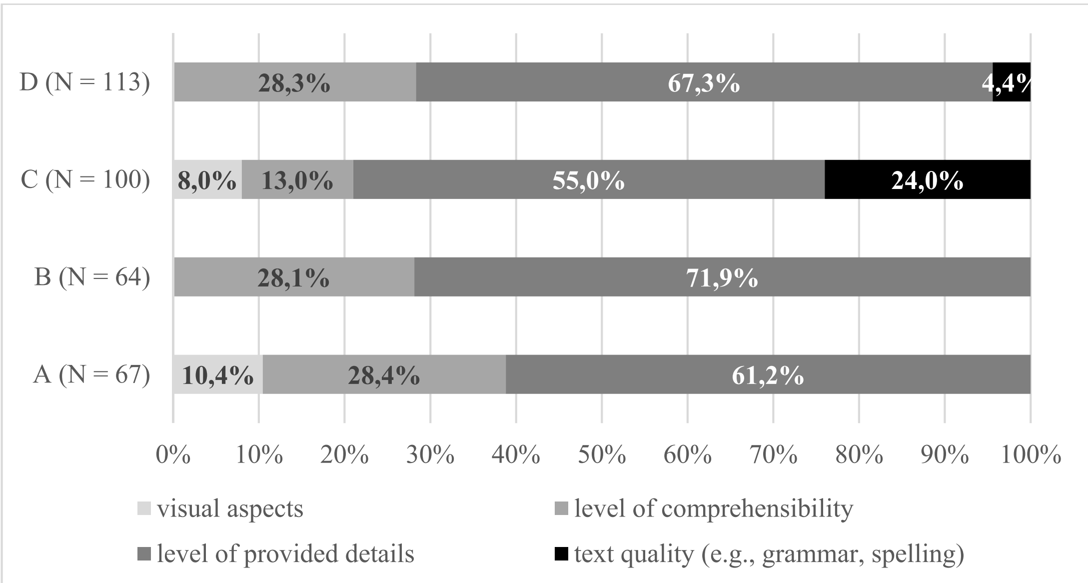

Figure 2. Aspects mentioned in teachers' feedback.

For individual comments, the presumable emotional impact of the feedback provided on the student was also estimated (Fig. 3). Of course, in this sense, these are rather illustrative conclusions which, however, were discussed with teachers in the interviews.

Most of the comments were written in such a way as to have a neutral impact on the students. As teachers A and B subsequently explained in the interviews, they tried to motivate the students with positive feedback, which could have an encouraging effect (e.g., *I like that you specifically describe the process of breathing and specify how breathing should take place).* The analysis showed that they tried not only to praise the students who had the protocol in order, but also the students who had shortcomings in it (e.g., *You did a piece of work in the protocol, but I recommend to proceed step by step according to the assignment*.). Some comments from Teacher C would have a rather negative impact on the student (e.g., *You have terrible nonsense in that protocol.*; *Is this how you write the results?*). Teacher C added that she always assesses students this way and warns them about mistakes. She sees no reason why she should praise someone for having the protocol in order. According to her, the student will receive a good grade for this. Teacher D mainly tried to use neutral commentary.

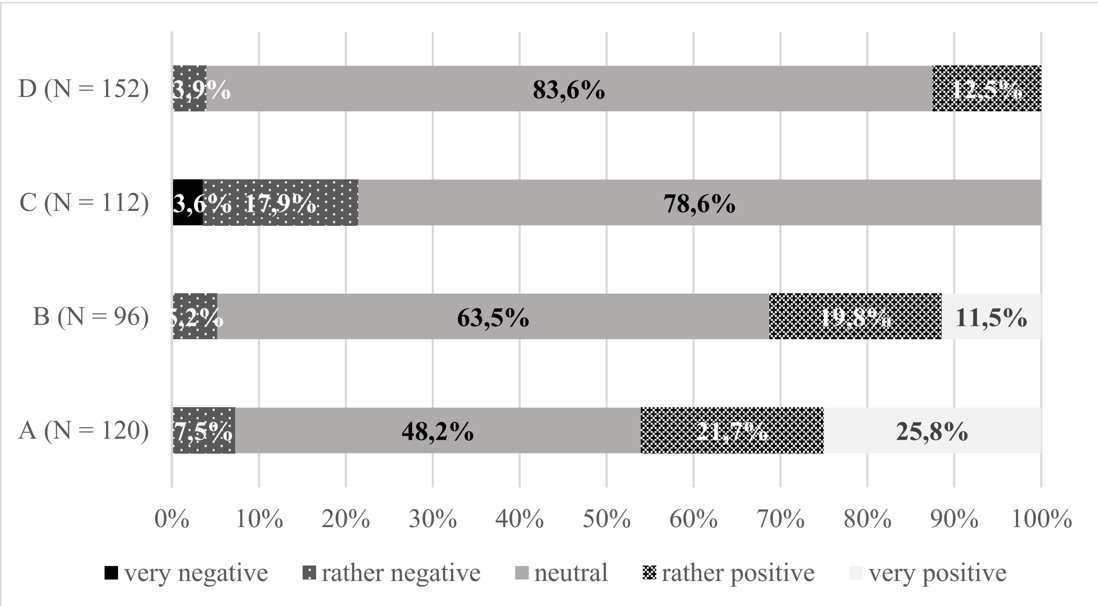

Figure 3. Presumable affective impact of the feedback provided on the student.

DISCUSSION AND CONCLUSION

The analysis of the feedback showed that the teachers tried to make their comments mostly formative. The exception is teacher C, who told the students where they had made mistakes and what to add to the protocol to make it correct. However, this was a mere statement, not a guide on how to do it on their own. Teachers A, B and D repeatedly explained the evaluation criteria to the students, justified why the student's solution was incorrect, and guided them to a different way of dealing with suggestions through stimulating questions. Introducing students to assessment criteria and explaining them are key steps in teaching. It is ideal to explain all the criteria before the activity (Sadler, 2015), which was done here, but emphasizing the criteria in the feedback is also beneficial because individual comments may relate to them.

Two teachers focused on finding out what the students had already achieved and tried to praise them. The affective connotations were rather exceptional in the feedback provided. To a greater extent, these connotations were rather vague formulations, especially of positive meaning, which were supposed to have a motivating effect on the students.

Assessment is deeply personal and can create strong feelings in the assessed individual (Crossman, 2007), but there is still little evidence of how emotions affect the way students are able to receive feedback and then work with it (Dowden et al., 2013, Rowe et al., 2014). Unfortunately, due to distance learning, we do not have exact data on how students responded to feedback. From the teachers' answers, it was only possible to assume how the students worked with the feedback, but the relevance of this information was also influenced by the distance form of teaching.

Even negative emotions can lead a student to talk to the teacher, which in the case of a constructive discussion can help the student to understand his own mistakes (Nicol et al., 2014). However, it must be acknowledged that affective connotations can lead to a strong subjectivity of such an assessment. The subjectivity of the assessment has often been seen as an obstacle, while in formative assessment it can be a tool that reflects personal needs, identities, interests, or attitudes. And emotions in that case can give such feedback the right charge. For example, Gates (2016) supports the use of subjective evaluation in creative areas of human activity, for example, in art education. However, more scientific evidence is needed to determine the precise mechanisms of the processes involved and contribute to the development of new frameworks that better explain the relationship of emotions to the cognitive, motivational, neurological, and social dimensions of feedback and assessment. (Rowe, 2017, p. 169)

Reflections from teachers themselves have shown that it is generally difficult for them to write a comment to make sure that students understand it. In this case, the teachers preferred to provide oral feedback, where they could explain everything to students immediately. On the other hand, teachers A and B added that they tried more to assess subjectively and give at least some emotions in the feedback, because they realized that distance learning is very difficult for students. And they felt the need to work better with their mistakes, to explain everything to them more, but most of all to encourage them to continue working.

It is certainly not possible to generalize the results, as the data were obtained at a specific time and on a small sample of respondents, but the following study could focus on, for example, how students receive feedback with and without affective connotations.

ACKNOWLEGMENTS

This study was supported by Technology Agency of the Czech Republic *Learning Hyperspace for Formative Assessment and Inquiry Based Science Teaching* (Reg. No. TL02000368).

REFERENCES

Abbassi, E., Blanchette, I., Ansaldo, A. I., Ghassemzadeh, H., & Joanette, Y. (2015). Emotional words can be embodied or disembodied: the role of superficial vs. deep types of processing. *Frontiers in Psychology*, *6*, 1–13. https://doi.org/10.3389/fpsyg.2015.00975

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappa, 80*(2), 139–148.

Constantinou, C. P., & Papadouris, N. (2016). *Synthesis of factors determining the uptake of various assessment methods in different educational settings*. Deliverable 5.11 in ASSIST-ME project.

Cross, D., & Lepareur, C. (2015). PCK at stake in teacher-student interaction in relation to students' difficulties. In M. Grangeat (Ed.), *Understanding science teachers' professional knowledge growth* (pp. 47–61). Sense.

Crossman, J. (2007). The role of relationships and emotions in student perceptions of learning and assessment. *Higher Education Research & Development, 26*(3), 313–327.

Dowden, T., Pittaway, S., Yost, H., & McCarthy, R. (2013). Students' perceptions of written feedback in teacher education: Ideally feedback is a continuing two-way communication that encourages progress. *Assessment & Evaluation in Higher Education, 38*(3), 349–362.

Evans, R. H., Ropohl, M., Nielsen, J. A., & Papadouris, N. (2016). *Affordances and challenges of written feedback as formative assessment in inquiry-based STEM education*. Abstract from NARST Conference, Baltimore, USA.

Gates, G. (2017). Embracing Subjective Assessment Practices: Recommendations for Art Educators. *Art Education*, *70*(1), 23–28. https://doi.org/10.1080/00043125.2017.1247565

Hattie, J., & Timperlay, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112.

Huntsinger, J. R. (2013). Does emotion directly tune the scope of attention? *Current Directions in Psychological Science*, *22*, 265–270

Lindquist, K. A., MacCormack, J. K., & Shablack, H. (2015). The role of language in emotion: predictions from psychological constructionism. *Frontiers in Psychology*, *6*, 1–17. https://doi.org/10.3389/fpsyg.2015.00444

Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources, and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome, & N. G. Lederman (Eds.), *Examining pedagogical content knowledge: The construct and its implications for science education* (pp. 95–132). Kluwer Academic Publishers.

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: a review of embodiment and the neuroscience of semantics. *Cortex*, *48*, 788–804. https://doi.org/10.1016/j.cortex.2010.11.002

Nicol, D., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, *31*, 199–218.

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment & Evaluation in Higher Education, 39*(1), 102–122.

Pekrun, R., & Stephens, S. J. (2010). Achievement emotions in higher education. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (pp. 257–306). Springer.

Rowe, A. D. (2017). Feelings About Feedback: The Role of Emotions in Assessment for Learning. In D. Carless, S. M. Bridges, C. K. Y. Chan, & R. Glofcheski (Eds.), *Scaling up Assessment for Learning in Higher Education* (pp. 159–172). Springer.

Rowe, A. D., Fitness, J., & Wood, L. (2015). University student and lecturer perceptions of positive emotions in learning. *International Journal of Qualitative Studies in Education*, *28*(1), 1–20.

Rowe, A., Fitness, J., & Wood, L. (2014). The role and functionality of emotions in feedback at university: A qualitative study. *The Australian Educational Researcher, 41*(3), 283–309.

Ryan, T., & Henderson, M. (2018). Feeling feedback: students' emotional responses to educator feedback. *Assessment & Evaluation in Higher Education*, *43*(6), 880–892. https://doi.org/10.1080/02602938.2017.1416456

Sadler, D. R. (2015). Backwards assessment explanations: Implications for teaching and assessment practice. In D. Lebler, S. Harrison, & G. Carey (Eds,). *Assessment in Music: From Policy to Practice* (pp. 9–19). Springer.

Talmi, D. (2013). Enhanced emotional memory: Cognitive and neural mechanisms. *Current Directions in Psychological Science*, *22*(6), 430–436

White, C. J. (2013). Higher education emotions: A scale development exercise. *Higher Education Research and Development*, *32*(2), 287–299.

Wiliam, D., & Leahy, S. (2015). *Embedding formative assessment: Practical techniques for K-12 classrooms.* Learning Sciences International.

RISK PERCEPTION CONCERNING COVID-19 PANDEMIC, GLOBAL WARMING AND FOOD AND NUTRITION SECURITY BY PRESERVICE SCIENCE TEACHERS

Leandro Silva[1], Pina Sollero[1], Júlia Bernardes[1] and Maurício Pietrocola[1]
[1]University of São Paulo, São Paulo, Brazil

Manufactured risks, generated by scientific and technological progress, are increasingly present in modern social life. Dealing with those risks require the ability to understand complex contexts and make decisions based on judgments. To understand risk perception among students a survey was developed and applied using the Amplified Risk Perception (ARP) theory based on three different manufactured risks: Global Warming, COVID-19 Pandemic and Food and Nutrition Security. The students were asked to answer the survey before and after the course and to develop a Teaching and Learning Sequence (TLS) about one of these three themes. Preliminary results shows that pre and post-tests had statistically significant differences between each other, resulting in an increased risk perception. Both the course and the TLS' development were significant to increase risk perception as well as the ARP's diagram, which made possible to create an assessment tool to evaluate teaching-learning situations.

Keywords: Socioscientific Issues; Diagnostic Tools; Scientific Literacy

INTRODUCTION

The pandemic caused by SArs-CoV-2 virus is spread all over the word. After more than two years, people are still fearing death and lacking the ability to deal with uncertainties that came in simple tasks of daily life. A diffuse feeling of insecurity pervades people's minds when they need to face risky situations, i.e., problems with no clear boundaries that are involved in unpredictable consequences. Otherwise, science was seen as a source of certainty that humankind has used to overcome fear and feeling of insecurity. Notwithstanding the benefits developed by science and technology all over the last 300-hundred-year, people feel more vulnerable and exposed to more risks (Douglas 1994).

Nowadays, risk situations are characteristic of today's industrial society and it emerges from complex situations, where many dimensions of social life like health, economy, science, and others are combined. (Beck et al. 2013). For Ulrich Beck (1992), we are living in a Risk Society, characterized as an age of Reflexive Modernity (Beck, Giddens and Lash 1994). In this kind of Modernity, central concerns of society have changed from developing and implementing recent technologies to managing risks associated with already existing technologies. In Risk Society, it is possible to distinguish between two types of risks: *external* – generated from outside of modern social life – and *manufactured* – generated by progress of humankind social and technological development. The external risks are more accessible and easier to be evaluated with basic knowledge and scientific support. But manufactured risks are normally hidden behind the complexity of Contemporary problems, as for example in the case of Covid-2019 pandemic. Search solutions in this scenario of Risk Society and complex problems require to be prepared to deal with dilemmas: health x economy, or safety and earning money. Kolstø (2001), as expatiates on socioscientific dilemmas, shows that students question risk assessment sources and their trusty relation on scientist intentions. In the Risk Society, trust does not arise from precision and authority. Instead, it comes from the ability to perceive multidimensionality in different contexts and producing adjusts on the go (Christensen 2009). Christensen reinforces the relevance of themes related to risk situations in school training for citizenship as it can take off from science the full ability to explain, predict and control all kinds of problems.

With this investigation, we aim to understand how manufactured risk may be an educational issue for science education. Specifically, we would like to develop teaching and learning sequences (TLS) to prepare students to perceive manufactured risk associated with Contemporary problems. Also, we may want to develop a tool to evaluate risk perception to be used in science classes/courses.

RATIONALE

Christensen (2009) argues that students' analysis is weakly grounded in their science knowledge or their understanding about the problem in place when they make risk assessment. The perception about risk is important in its management. It presumes to be able to discriminate distinct aspects presented in the situations, to evaluate consequences using reliable sources, and make a decision based on judgment. Simplistic causal relations, readily taking few available elements in the account must be avoided (Hansen and Hammann 2017). Thus, we adopt the diagram of amplified risk perception (ARP) (Pietrocola at. al, 2020), shown in Figure 1, to follow the way people's perception about manufactured risk may evolve. The ARP is a three-dimensional risk perception space indicating a group's ability to perceive risks related to a particular situation. *Access* is associated with rational thinking supported by scientific cognition. *Urgency* defines a hierarchy of risk connected with values and practices in each culture. *Range* relates to the ability to make impact assessments either in a close or far perspective. This three-dimensional risk perception space indicates one's (or group's) ability to perceive risks related to a particular situation.

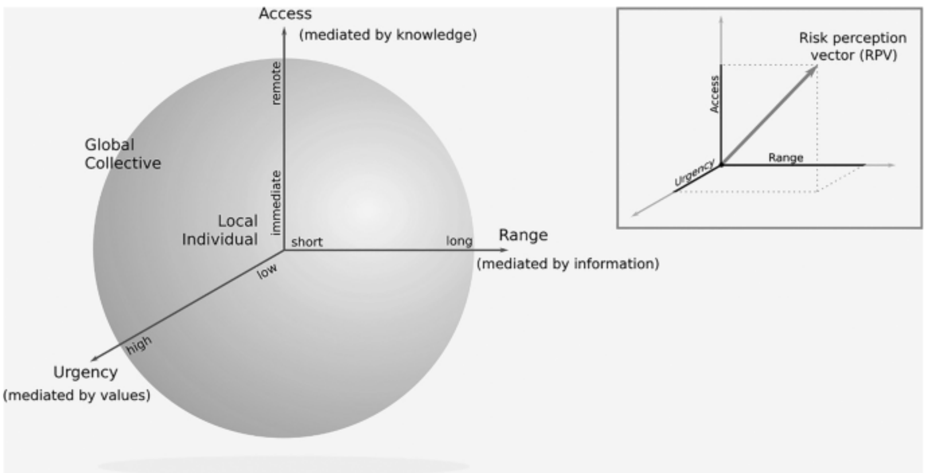

Figure 1. The Amplified Risk Perception (ARP) diagram.

In a course dedicated to prepare Pedagogical interventions in science classes, pre-service science teachers prepare Teaching and Learning Sequences (TLS) about the themes: (i) Global warming (GW), (ii) Covid-19 Pandemic 2019 (PCOV) and (iii) Food and Nutrition Security (FNS). They all three are defined as a manufactured risk (Giddens, 1990), and representing risk situations faced in the present moment (PCOV), immediate past (GW), and in immediate future (FNS). During the course, students were asked to read articles about risk perception and management, to choose a risk situation related to one of the three main themes and to make their own risk management's matrix, based on their appropriation of the subject. Later, each student chose one of the 3 themes to develop a TLS for science classes at High School level. In groups of 10 to 12 they were able to discuss which subjects were more significant to their TLS.

RESULTS

In this first analysis, we used the ARP diagram to identify and locate students within the Access, Range and Urgency axes through the risk perception survey. To calculate these indexes, we considered the simple average responses, according to the degree of agreement. Figure 2 shows the distribution in both pre and post-tests.

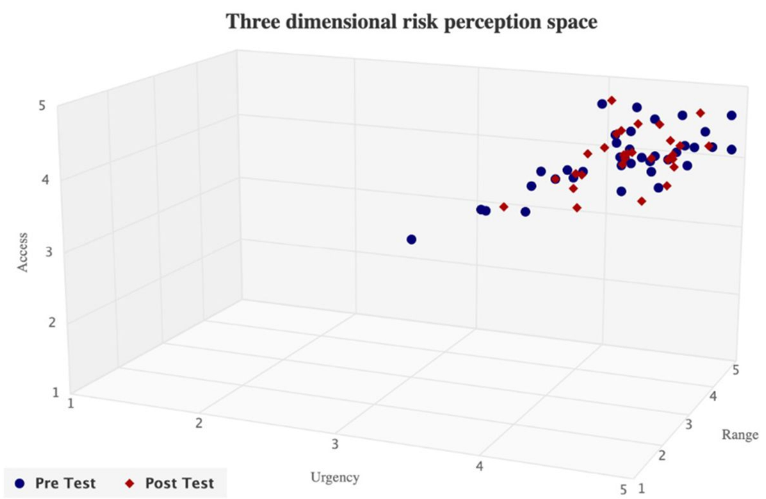

Figure 2. Students' distributions according to the ARP diagram.

Visually, we were able to notice a greater dispersion of students in the parts closer to the axes in the pre-test, compared to the post test result, indicating that there was an increase in the risk perception throughout the course. Additionally, we used the Wilcoxon signed-rank test to determine statistically whether there were significant changes in risk perception, comparing the results before and after the course, using the computer software Jamovi, assuming the hypothesis that the post-test means (M2) are greater than pre-test means (M1), which indicates an increase of risk perception. Table 1 illustrates this data.

Table 1 – Paired Wilcoxon result comparing risk perception before and after the course.

	Pre-test	Post-test				
	Mean (M1)	**Mean (M2)**	**N**	**Statistics**	**p**	**Effect Size**
Urgency[1]	4.26	4.32	29	56.50	0.107	-0.3392
Range[2]	4.05	4.09	29	80.00	**0.040**	-0.4203
Access[3]	4.11	4.18	29	53.50	0.143	-0.3007

1 – 11 pairs of value are tied; 2 – 6 pairs of value are tied; 3 – 12 pairs of value are tied

Note: $H_a\mu_{M1-M2} < 0$; $H_0\mu_{M1-M2} = 0$

It was observed there was a meaningful change in all axes of the ARP diagram, with a greater variation in the Range axis. The Effect Size values indicates moderate (between 0.3 and 0.5) change directed to M2 which denotes greater risk perception, assuming there is a notable change between the pre and post-test means. The p-value for the Range axis represents statistical significance (< 0.05) against the null hypothesis and strong evidences that the students had their risk perception affected by the course. Although, the same result was not observed for the Access and Urgency axes.

Individually, each questionnaire had equivalent results, with meaningful p-value and Effect Size in one axis. Tables 2, 3 and 4 show Covid-19 Pandemics, Global Warming and Food and Nutrition Security results, respectively.

Table 2 – Paired Wilcoxon result comparing Covid-19 pandemic's risk perception before and after the course.

	PCOV Pre-test	PCOV Post-test				
	Mean (M1)	**Mean (M2)**	**N**	**Statistics**	**p**	**Effect Size**
Urgency[1]	4.37	4.39	29	36.50	0.473	-0.0641
Range[2]	4.19	4.14	29	83.00	0.789	+0.2206
Access[3]	4.34	4.41	29	02.50	**0.052**	-0.7619

1 – 17 pairs of value are tied; 2 – 13 pairs of value are tied; 3 – 23 pairs of value are tied

Note: $H_a\mu_{M1-M2} < 0$; $H_0\mu_{M1-M2} = 0$

Table 3 – Paired Wilcoxon result comparing Global Warming's risk perception before and after the course.

	GW Pre-test	GW Post-test				
	Mean (M1)	**Mean (M2)**	**N**	**Statistics**	**p**	**Effect Size**
Urgency[1]	4.25	4.35	29	09.50	0.066	-0.5778
Range[2]	3.79	3.92	29	00.00	**0.004**	-1.0000
Access[3]	3.87	3.90	29	49.00	0.275	-0.1833

1 – 20 pairs of value are tied; 2 – 14 pairs of value are tied; 3 – 22 pairs of value are tied

Note: $H_a\mu_{M1-M2} < 0$; $H_0\mu_{M1-M2} = 0$

Table 4– Paired Wilcoxon result comparing Food and Nutrition Security's risk perception before and after the course.

	FNS Pre-test Mean (M1)	FNS Post-test Mean (M2)	N	Statistics	p	Effect Size
Urgency[1]	4.17	4.22	29	05.00	0.074	-0.6429
Range[2]	4.36	4.50	29	06.00	**0.002**	-0.8857
Access[3]	3.94	3.97	29	14.50	0.100	-0.4727

1 – 22pairs of value are tied; 2 – 15 pairs of value are tied; 3 – 19 pairs of value are tied

Note: $H_a \mu_{M1 - M2} < 0$; $H_0 \mu_{M1 - M2} = 0$

The results show that the increase of risk perception is not homogeneous between the themes either the axes of the ARP diagram. Students' risk perception in the PCOV theme had a meaningful change in the Access axis whereas GW and FNS themes had significant change in the Range axis; which shows that the course improved their knowledge about the COVID-19 Pandemics, whereas expanded their information to Global Warming and Food and Nutrition Security. For the other two axes of the ARP diagram in each theme we cannot assure statistically that the course has affected students' risk perception since there is not significant p-value (>0.05) or have low Effect Size value (below 0.3).

Also, in all three groups, it was observed that students changed their perceptions both in the theme in which they worked, as in the others. It can be noted that the justifications addressed aspects of the discussions held during the course activities. The FNS group showed the greatest increase in risk perception associated with its own theme. But this result was not repeated for the other two themes, since the GW group showed the greatest increase in the risk perception associated with COVID-19 and the increase in perception regarding global warming was not homogeneous between the axes, with GW group showing a greater increase in the axes of access and urgency, but in the axis of range was the FNS group.

CONCLUSION

The results obtained allowed us to conclude that the ARP diagram can be used as an assessment tool in teaching-learning situations. The technique we used to associate dimensions of the diagram with assertions and degrees of agreement allowed us to obtain a semi-quantitative indicator that represents the amplitude of students' risk perception. We were also able to verify that the course had a decisive effect in amplifying students' risk perception, and the production of the TLS allowed them to apprehend complexity of risk situations. Contrary to what common sense might indicate, there was no direct correlation between the topic studied and the increase in the perception of risk about it. Although it is only a preliminary result, risk perception should not be seen as a local skill. In other words, it does not work as in the solution of traditional problems where there is a delimited epistemological profile. The result obtained reinforces the idea that risk situations should be treated as complex problems and it is important to develop global skills of analyses and decision making.

ACKNOWLEGMENTS

Maurício Pietrocola thanks Brazilian Scientific Research Council (CNPq) for the partial support for this research. Leandro Silva thanks the Coordination of Superior Level Staff Improvement (CAPES) for partial support for this research.

REFERENCES

Beck, U. (1992). Risk society: towards a new modernity. Sage Publications.

Beck, U., Giddens, A., & Lash, S. (1994). Reflexive Modernization Politics, Tradition and Aesthetics in the Modern Social Order. Stanford: Stanford University Press.

Beck, U., Blok, A., Tyfield, D., & Zhang, J. Y. (2013). Cosmopolitan communities of climate risk: Conceptual and empirical suggestions for a new research agenda. Global Networks, 13(1), 1–21. doi:10.1111/glob.12001

Christensen, C. (2009). Risk and school science education. Studies in Science Education, 45(2), 205–223. doi:10.1080/03057260903142293

Douglas, M. (1994). Risk and Blame: Essays in Cultural Theory. London: Routledge.

Giddens, A. (1990). The consequences of modernity. Stanford, CA: Stanford University Press.

Hansen, J., & Hammann, M. (2017). Risk in Science Instruction: The Realist and Constructivist Paradigms of Risk. Science and Education, 26(7–9), 749–775. doi:10.1007/s11191-017-9923-1

Dan M. Kahan, Hank Jenkins-Smith & Donald Braman (2011) Cultural cognition of scientific consensus, Journal of Risk Research, 14:2, 147-174, DOI: 10.1080/13669877.2010.511246

Kerby, D. S. (2014). The simple difference formula: An approach to teaching nonparametric correlation. Comprehensive Psychology, 3, 2165–2228.

Kolstø, S. D. (2001). Scientific literacy for citizenship: Tools for dealing with the science dimension of controversial socioscientific issues. Science Education, 85(3), 291–310. https://doi.org/10.1002/sce.1011

Levinson, R., Kent, P., Pratt, D., Kapadia, R., & Yogui, C. (2012). Risk-based decision making in a scientific issue: A study of teachers discussing a dilemma through a microworld. Science Education, 96(2), 212–233. doi:10.1002/sce.21003

Pietrocola, M., Rodrigues, E., Bercot, F., & Schnorr, S. (2020, online first). Science education in pandemic times: what can we learn from COVID-19 on science technology and risk society. https://doi.org/10.35542/osf.io/chtgv

R Core Team (2021). R: A Language and environment for statistical computing. (Version 4.0) [Computer software]. Retrieved from https://cran.r-project.org. (R packages retrieved from MRAN snapshot 2021-04-01).

The jamovi project (2021). jamovi. (Version 2.2) [Computer Software]. Retrieved from https://www.jamovi.org.