JOURNAL ON EFFICIENCY AND RESPONSIBILITY IN EDUCATION AND SCIENCE

The ERIES Journal is being managed by an international editorial board as a regular scientific journal. A rigorous process of papers' reviews (double-blind peer review) is fully supported by a web-based submission system. The journal is published electronically four times a year, on March 31, June 30, September 30 and December 31 of the current year.

The journal is indexed in:

- BASE - Bielefeld Academic Search Engine
- Directory of Open Access Journals
- EBSCO database
- Emerging Sources Citation Index - Web of Science™ Core Collection
- ERIC - Education Resources Information Center
- ERIH PLUS
- Google Scholar
- Open Academic Journals Index
- ResearchBib
- SCOPUS
- The list of reviewed periodicals in the Czech Republic

www.eriesjournal.com

Scientific journal of the Czech University of Life Sciences Prague JOURNAL ON EFFICIENCY AND RESPONSIBILITY IN EDUCATION AND SCIENCE, distributed by the Faculty of Economics and Management. Published quarterly. Executive editors: Ing. Martin Flégl, Ph.D., Ing. Igor Krejčí, Ph.D. and Ing. Tereza Sedlářová Nehézová. Editorial Office: ERIES Journal, Czech University of Life Sciences Prague, CZ 165 21 Prague 6 - Suchdol, Czech Republic, email: editor@eriesjournal.com, tel: +420 224 382 355.

volume 14
issue 4

2021

An international peer-reviewed journal published by
Faculty of Economics and Management
Czech University of Life Sciences Prague

editor@eriesjournal.com
www.eriesjournal.com
Online ISSN: 1803-1617
Printed ISSN: 2336-2375

EDITORIAL BOARD

EDITORIAL OFFICE

ERIES Journal, Czech University of Life Sciences Prague, CZ 165 21 Prague 6 - Suchdol, Czech Republic
email: editor@eriesjournal.com * tel: +420 224 382 355
Registration number: MK ČR E 21414 © Czech University of Life Sciences Prague, Czech Republic, 2021

INSTRUCTIONS FOR AUTHORS

The Journal on Efficiency and Responsibility in Education and Science publishes papers of the following categories: full research papers, short communications, review studies and book reviews (on invitation only).

- FULL RESEARCH PAPERS
- SHORT COMMUNICATION
- REVIEW STUDY

Papers are published in English. A paper may comprise an empirical study using an acceptable research strategy, such as survey, case study, experiment, archival analysis, etc. It may contain a theoretical study aimed at advancing current theory or adapting theory to local conditions or it may arise from theoretical studies aimed at reviewing and/or synthesizing existing theory. Concepts and underlying principles should be emphasized, with enough background information to orient any reader who is not a specialist in the particular subject area.

Submission checklist

The paper. The paper is carefully formatted according to the template of the journal (see bellow). Special attention is paid to the exact application of the Harvard referencing convention to both continuous citations and list of references. If an electronic source has the DOI number assigned, also it will be provided in the list of references. Manuscripts are submitted via the editorial system in the DOC.

Research highlights. The core results, findings or conclusions of the paper are emphasized in 2-4 bullet points (max. 150 characters per bullet point including spaces). The highlights are submitted as a text into the submission form in the editorial system.

Copyright form. The submission of a paper will imply that, if accepted for publication, it will not be published elsewhere in the same form, in any language, without the consent of the Publisher. The manuscript submitted is accompanied by the copyright form signed by the corresponding author who declares the agreement of all authors with the conditions in the Form. The Form is submitted into the editorial system in the PDF format.

Suggested reviewers. It is required to suggest two experts appropriate to evaluation of the paper. The experts should be out of the affiliation of the author(s), Czech University of Life Sciences Prague, and also both experts should be from different affiliations. The reviewers are submitted into the text fields in the submission form of the editorial system.

Preparation of the manuscript (technical notes)

Authors are responsible for applying all requirements that are specified in the journal's paper template in individual sections. Especially, the paper must provide a short review of current state in the area of the paper's aim in Introduction. The paper should refer significant sources, particularly scientific journals or monographs.
Papers must be closely scrutinized for typographical and grammatical errors. If English is not author's first language then the paper should be proof-read by a native English-speaking person, preferably one with experience of writing for academic use. Spelling should follow the Oxford English Dictionary.
Tables, graphs and illustrations should be drawn using a suitable drawing package. Colour may be used. Place all diagrams and tables where you wish them to appear in the paper. Ensure your diagrams fit within the margins and are resizable without distortion.

Review procedure

Following Editorial recommendation, papers are submitted to a double-blind peer review process before publication. Commentary by reviewers will be summarized and sent by email to authors, who can choose to revise their papers in line with these remarks. Re-submitted papers should be accompanied by the description of the changes and other responses to reviewers' comments (see above), so that the desk-editor can easily see where changes have been made.

Copyright

Authors are fully responsible for the paper's originality and for correctness of its subject-matter, language and formal attributes. Author's statement should be enclosed declaring that the paper has not been published anywhere else.

The submission of a paper will imply that, if accepted for publication, it will not be published elsewhere in the same form, in any language, without the consent of the Publisher. Before publication, authors will be asked to complete a copyright release, giving the publisher permission to publish the paper in a specific issue of this Journal. Overall copyright ownership of the paper, however, remains with the author/s. It is the authors' responsibility to obtain written permission to quote material that has appeared in another publication.

With this last issue (Vol. 14, No. 4) of 2021, it is time to recapitulate a little bit the ending year. It has been the second year in which ERIES Journal has obtained the Scimago Journal & Country ranking SJR score and the third year since the inclusion in Scopus database. All metrics show significant growth, indicating an increasing journal quality and growing interest of readers. For example, the number of citations per document almost doubled during the last two years, so the external cites per document. As a result, ERIES Journal attracted more international authors from around the globe. For example, during this year, the authors from Argentina, Colombia, the Czech Republic, Ethiopia, Mexico, Norway, Russia, Turkey, among others, have published their research in ERIES Journal. Although we received significantly more articles in comparison to the situation couple of years ago, we are committed to accept for publication only the research of the highest quality with a potential to advance the research dealing with issues of efficiency and/or responsibility in education and related scientific disciplines.

Further, to continue to strengthen the journal's position, to increase the scientific reach and impact, we continued to extend the Editorial board of ERIES Journal. Recently, we welcomed two new members: Dr. David Pérez Jorge from the Department of Didactics and Educational Research, University of La Laguna in Spain and Dr. Martin Daumiller from the Department of Psychology, University of Augsburg in Germany. Both researchers have a strong publication history in top quality international journals, which will help during discussion about the future goals of the journal, as well as to cover more areas in educational research.

Finally, ERIES Journal has achieved a new chapter in its history as, for the first time, ERIES Journal prepared the first special issue call for papers. The special issue on Practitioner Inquiry: Towards Efficiency and Responsibility in Teaching and Teacher Education is covered by the Guest editors Assoc. Prof. Dr. Irem Çomoglu from Dokuz Eylul University in Turkey and Dr. Ozgehan Ustuk from Balikesir University in Turkey. ERIES Journal welcomes research that demonstrates critical and effective pedagogies, approaches, strategies, or implementations of bottom-up teacher education and/or professional development through practitioner inquiry. The deadline for abstract submissions is on January 30, 2022. For more information, please visit the ERIES Journal website or follow the updates on our LinkedIn page.

The last volume of 2021 includes six articles. First, Zülal Ayar in the article "Investigating Professional Teacher Identity through ESP Courses: Voices from Three EFL Instructors via Case Studies" investigates the perceptions of EFL instructors about ESP practices and their professional identities. Whether instructors' sense of self-efficacies is interrelated with their teaching practices and professional identities. And, if instructors' professional competencies overlap with their professional identities. Second, Gustavo Ferro and Carlos Romero in the article "The Productive Efficiency of Science and Technology Worldwide: A Frontier Analysis" measure the knowledge production of publications and patents obtained from human and non-human resources at the country level. Third, Oğuzhan Nacaroğlu, Oktay Bektaş and Mustafa Tüysüz in the article "Examination of Science Self-Regulation Skills of Gifted and Non-Gifted Students" examine and compare the science self-regulation skills of gifted and non-gifted. Forth, Salih Yıldırım in the article "Use of Screencast in Distance Education GIS Lessons: Students' Views" analyzes the views of geography teacher students on the GIS course conducted with screencasts during the distance education process. Fifth, Aneta Mazouchová, Tereza Jedličková and Lucie Hlaváčová in the article "Statistics Teaching Practice at Czech Universities with Emphasis on Statistical Software" observed the beliefs of students and fresh university graduates about teaching statistics using statistical software. Finally, in the sixth article "A Joint Assessment of Reasoning about General Statements in Mathematics and Biology" from Libuše Samková, Lukáš Rokos and Lukáš Vízek focused on future primary school teachers and their modes of argumentation in mathematics and science, namely in arithmetic, geometry and biology.

We hope that all our readers will find this last issue of the year 2021 interesting. We also hope that ERIES Journal will contribute to the field of efficiency and responsibility in education and science as it has contributed so far. With the end of the year 2021, we would like to thank all the authors who have submitted their manuscripts to ERIES Journal, to all reviewers who carefully reviewed all these manuscripts, as well as to all members of the Editorial board who contributed to increase the ERIES Journal quality.

We wish you Merry Christmas and Happy New Year 2022.

Sincerely

Martin Flégl
Executive Editor
ERIES Journal
www.eriesjournal.com
www.linkedin.com/company/eriesjournal/
www.erie.pef.czu.cz

CONTENT

Full research paper

INVESTIGATING PROFESSIONAL TEACHER IDENTITY THROUGH ESP COURSES: VOICES FROM THREE EFL INSTRUCTORS VIA CASE STUDIES

Zülal Ayar

Ankara University, Turkey

✉ ayarz@ankara.edu.tr

ABSTRACT

The current zeitgeist in language teacher education dwells on teacher identity regarding it as one of the big buzzwords to explore and critically reflect teacher qualities from a socio-cognitive perspective. Drawing from this current trend, the research intended to disambiguate three English language instructors' sense of self-efficacy, perceptions, professional identity and professional competence in one of the most established and prestigious state universities in the country through the qualitative case study method. After selecting ESP practitioners following convenience sampling, autobiographies, informal dialogues, classroom observations, opening interviews, post-observation interviews, and field notes were utilized to gather data. Having scrutinized the professional identities of instructors through the lens of self-efficacy beliefs, perceptions, and professional competence, the study exposed that professional competence came to the fore being the best mediator to gain awareness of professional teacher identity. However, self-efficacy did not subserve as a predictor in exploring the complexity of teacher identity due to discordance with perceptions and realities of identity issues. Finally, some suggestions for further considerations were stated to be operationalized within the EFL context of in-service language teacher education.

KEYWORDS

Case study, perception, professional competence, self-efficacy, teacher identity

HOW TO CITE

Ayar Z. (2021) 'Investigating Professional Teacher Identity through ESP Courses: Voices from Three EFL Instructors via Case Studies', *Journal on Efficiency and Responsibility in Education and Science*, vol. 14, no. 4, pp. 204-216. http://dx.doi.org/10.7160/eriesj.2021.140401

Article history
Received
May 14, 2021
Received in revised form
August 17, 2021
Accepted
November 1, 2021
Available on-line
December 20, 2021

Highlights

- *The research has disambiguated English language instructors' sense of self-efficacy, perceptions, professional identity and professional competence.*
- *The study has exposed that professional competence came to the fore being the best mediator to gain awareness of professional teacher identity.*
- *Self-efficacy has not subserved as a predictor in exploring the complexity of teacher identity.*

INTRODUCTION

In language teacher education, Varghese et al. (2005) emphasize the necessity of understanding teachers' reasoning about their professional identities to interpret language teaching and learning thoroughly. Likewise, Sachs (2005:15) defines:

'Teacher identity provides a framework for teachers to construct their ideas of 'how to be', 'how to act', and 'how to understand' their work and their place in society. Importantly, it is not something that is fixed nor is it imposed; rather it is negotiated through experience and the sense that is made of that experience.'

This explanation lays the weight on cognitive aspects of identity in addition to stressing how crucial it is to decipher this concept within a social framework. The post-modernist views of Duff and Uchida (1997), Olsen (2008), Vygotsky (1978), Wenger (1998) also suggest a multi-pronged approach to be espoused while treating teacher identity and decrying its non-linear process (Yuan and Lee, 2016). Then, it can be postulated that the teacher's self is not restricted to his/her beliefs, ideas and manners, but rather that it interacts with their expectations and experiences as well. This dynamic and selfhood human agency and intrapersonal identity, which turns into 'situated identities' (Alsup, 2006), and finally 'selves' evolving in a social milieu, shape his/her understanding, school of thought, principles for the profession and helps to build the feeling of belonging (Skaalvik and Skaalvik, 2011). Thereby its connection with teacher perception, self-efficacy, and professional competence need to be further explored to ascertain the teacher professional identity issue in-depth.

Teacher perceptions

In a general sense, this concept refers to teachers' apprehension of their roles within a profession, and with class practices. Atkinson et al. (1987) clarify that teachers' perceptions towards teaching can be thought of as representative of their awareness of professional activities. In other words, their mindset, expectations, and attitude towards teaching performance have to do with their personal and professional development. Beijaard, Verloop and Vermunt (2000) also touch upon the fact that teachers' perceptions on professional identities have an impact on efficacy beliefs, eagerness to approach educational innovations, professional development and in turn being able to put these changes into practice. This means; perception is an intricate concept, demanding versatile thinking in examining its salient role in teacher identity.

Professional competence

The terms of professional learning, teacher capability, respecting the utility of professional development activities and managing professional socialization, which is a *rite of passage* in identifying their social status, usually spring to mind concerning competency. Steiner (2010) affirms that competencies ought to be viewed as a means of fostering professional development among teachers. In a similar vein, Pantic, Wubbels and Mainhard (2011) characterize its multi-faceted aspect in that professional competence comprises emotions, attitudes, qualities, decision-making ability, communication skills, and knowledge as the linchpin of professional identity.

Górska-Poręcka (2013) points out the dichotomy between instructors giving general English lessons and the ones lecturing English for Specific Purposes (ESP) who need more specific subject matter knowledge in their discipline. Deriving from the significance of pedagogical knowledge in professional competence as featured by Blömeke and Delaney (2012), professional beliefs about teaching and learning, motivation and self-regulation need to be emphasized. Furthermore, distinct types of knowledge, such as Pedagogic Content Knowledge (PCK) or content knowledge put forward by Shulman (1987) must be regarded as the flip side of the coin. Along those lines, Ibarra (1999), and Schein (1978) pronounce that teacher beliefs, motivation, and values originated from professional self-concepts, besides the years of teaching experience, will, in turn, shape professional identity. Taking all these facts into consideration and respecting the inextricable nature of professional thoughts, Eraut (1994: 162) outlines that 'competence should be viewed as an appropriate cut-off point on a learning continuum, not as a state of mastery.'

Self-efficacy

Bandura (1997) explicates self-efficacy as one's performance appraisal to develop and display specific behaviours to reach the objectives. It has been the subject of profound debate and concern and vetted primarily from different angles in quantitative research designs through efficacy scale instruments; nevertheless, these scales have been mostly criticized. These criticisms were owing to agency problems (Ross, 1995), the treatment of similar topics (Smylie, 1996), excluding contextual factors (Henson, 2002), and breaking confidentiality (Tschannen-Moran and Hoy, 2001). Another expostulation was about the dearth of qualitative analyses of salient teaching modus operandi, such as observations or interviews to decipher this issue in detail (Tschannen-Moran, Hoy and Hoy, 1998). Considering these evaluations, teacher efficacy has been determined to be associated with mainstream agents, such as competency, and preferences to discover whether they interlock with professional teacher identity.

The interrelationship among identity, perception, self-efficacy, and competence

Having identified these four concepts, they seem to be tightly correlated with one another in terms of enhancing teachers' academic commitment and productivity. To date, some ground-breaking academic research has been carried out to examine teacher identity and its connection with autonomy and/or agency (Derakhshan et al., 2020; Huang, 2009; Lennert da Silva and Mølstad, 2020; Teng, 2019). Similarly, some has explored its potential link with beliefs (Basturkmen, 2012), level of awareness towards profession (Cooney and Shealy, 1997), self-efficacy (Day, 2002), professional competence (Charters, 1976; Hurt-Avila and Castillo, 2017; Lauermann and König, 2016; Pajares, 1992), and teacher learning (Richards, 2021). Dikilitaş and Mumford (2019) also handled autonomy from agency, motivation, and identity aspects in the teacher development process and concluded that teachers could fulfil Continuing Professional Development (CPD) needs by nurturing their autonomy. Likewise, Buchanan (2015), Lasky (2005), McNicholl (2012), and Vähäsantanen (2015) underlined the interrelatedness of agency and professional identity from a sociocultural perspective. A wealth of studies has been primarily operationalized in ESL (Fotovatian, 2015; Harun, 2019) or administered on prospective teachers (Babanoğlu and Ağçam, 2019; Salinas and Ayala, 2018; Yazan, 2016). However, there are few if any examinations in in-service education aiming to have a thorough understanding of teacher identity and exploring its potential correlations with efficacy, competence, and perceptions, particularly in the Turkish EFL context. It follows that the current research has aimed to fill a niche in the literature by addressing the following research questions:

1. What are the perceptions of EFL instructors about ESP practices and their professional identities?
2. Are instructors' sense of self-efficacies interrelated with their teaching practices and professional identities? If so, how?
3. Are instructors' stated professional competencies overlap with their professional identities? If so, how?

MATERIALS AND METHODS

Case studies have constantly been in the eye of the storm in the research paradigm due to its non-transferable feature of research findings to other stakeholders and inability to be generalized to universes (Firestone, 1993; Robinson and Norris, 2001; Ruddin, 2006; Tellis, 1997; Woodside, 2010; Yin, 2003). Lincoln and Guba (2002), and Merriam (2009) asserted that case studies in qualitative methodology must perform analytical conclusions after an in-depth examination of the specific context in which they occurred (Lincoln and Guba, 1985). In addition, Nunan and Bailey (2010) underlined that a case can be either *bounded* or *integrated* depending on the physical and temporal boundaries. As this qualitative design was based on in-depth and holistic analyses of ESP instructors and their teaching experiences in a particular context bounded by time and activity (i.e. institution), it can be described as a case study research (Merriam and Tisdell, 2016).

Research context and participants

The study was performed at a comprehensive university in the central Anatolia region. The researcher adopted the convenience sampling method to identify the samples from within the instructors working at the school of foreign languages of this university. Initially, the study was planned to be conducted on eight EFL instructors as Table 1 indicates; however, considering the caveat of Duff (2008) in terms of incorporating a maximum of four or five participants in this kind of research, this number was reduced to four. Hence, the first four volunteer participants were purposefully incorporated into the design, whereas the other four were excluded from the prospective attendees. Notwithstanding this fact, one of the participants (labelled I4 in the table below) could not allocate enough time to take part in the post-observation interviews due to a large number of projects and a tight schedule. Thus, three instructors were determined to be the research participants in the end.

Instructors	Age	Gender	Years of teaching experience	Major	Educational Background
I1	55	Male	31	English Language Teaching	Bachelor's Degree
I2	41	Female	15	English Translation and Interpretation	Master's Degree
I3	28	Female	9	English Language Teaching	Doctor of Philosophy
I4	48	Male	24	English Language and Literature	Doctor of Philosophy
I5	51	Female	28	English Language Teaching	Bachelor's Degree
I6	54	Female	30	English Language and Literature	Master's Degree
I7	57	Male	34	English Language Teaching	Doctor of Philosophy
I8	52	Female	26	English Linguistics	Bachelor's Degree

Table 1: Demographics of instructors

Instructor 1 ('I1' henceforth) graduated from the English Language Teaching (ELT) department in 1989, and abandoned his Master's at the faculty of fine arts before writing his thesis. Initially, he delivered English lessons at two private schools in distinct regions of the country for three years and thereafter he held a temporary position as a part-time teacher at a university's school of foreign languages. Still, owing to an educational reform which did not stipulate students to take a preparation class in their first year at university, he could not see a promising future as a teacher in this setting. Then, a prestigious publishing house of foreign origin recruited him as an in-house trainer, editor, and material developer for about eight years. Yet, due to some serious problems in marketing and training, he reconsidered teaching English again at the same university he had previously worked for as a full-time instructor, though he would hold his position in the publishing house simultaneously throughout his academic career.

He mostly lectured ESP courses at faculties and graduate schools as well as taught basic English at the foreign language school. Since 2004, he had been teaching medical English, English for media, diplomatic English and vocational English for students at the faculty of political science in addition to English for TOEFL IBT, and IELTS exams. Besides having a lot of published work in English and Turkish, he had certificates of attendance from a variety of conferences, conventions, and seminars. As a seasoned instructor, he described that the bases of his teaching philosophy were against the rote-learning and Presentation-Practice-Production (PPP) model, but since he principally focused on learner characteristics, rather relied on interactive communication, the Socratic method, critical thinking, flipped learning, contextualization, using technology and digital materials, such as Google class and WhatsApp. Accordingly, he would appeal to online role-plays, video recordings, or written assignments which did not demand learners to take high risks by thrusting themselves into the forefront. The implication is that he had a high opinion of his learners and their uniqueness besides his attempts to discover new routes of what and how to teach. I1 also added how meticulous he was in selecting lead-in, warm-up, and motivating activities beyond giving instructions and providing elicitations in his class. Additionally, he attached importance to increasing the awareness of learners towards learning a foreign language and using it in their daily lives to consolidate language knowledge.

Instructor 2 ('I2' henceforth) completed her Bachelor of Arts in English Translation and Interpretation in Iran in 2004. She stated that the basic reason behind choosing a career in language teaching was the four years she spent in London in primary school. However, despite her graduation from an English language department, she preferred to hold her master of education in journalism from 2004 to 2007 while working both as a journalist and a translator in international business and public policy in Iran for 12 years. She did not have any teaching experience through practicum as in the department of teacher education nor did she have a chance of getting related vicarious experience. Thus, she gained her first field-

based experience in a primary school only for a few years as a part-time substitute. Subsequently, her teaching experience as a contractual instructor started with the compulsory journalism ESP courses (e.g. Professional English for media and English for professional life) at this university in 2011. The only pedagogical support she was able to receive was the professional online courses about designing instructional programmes and Non-Governmental Organization (NGO) courses in addition to face-to-face Mathematics, English, and Turkish lessons in the Persian language in İstanbul and Ankara three times a week. However, she has been planning to apply for new teaching qualification programmes in the near future.
She confessed that at first, she had difficulty in delivering business English classes due to her insufficient subject matter knowledge and pedagogical content knowledge in foreign language teaching. Thereafter, she determined to be more fluent, self-confident, and efficient in language teaching, which led to changes in her perception towards learner needs. This included planning the curriculum, building rapport with learners via short dialogues during the breaks rather than creating teacher dominance, assigning them tasks to increase collaboration and limiting expectations in line with their progress. As to her conception of teaching, on one hand, she has principally followed creative teaching, and computer-assisted language learning. On the other hand, she has employed translation-focused language education, the repetitive language instructions instead of leading students to discovery (or inquiry)-based learning, and she is completely opposed to teaching grammar and including the PPP approach in language lessons.
Instructor 3 ('I3' henceforth) held her undergraduate, graduate, and postgraduate degree in English language teaching. After her graduation in 2012, she was recruited by a state university in the Marmara region as a full-time instructor. Three years later, she changed her workplace and maintained her teaching experience at this university while writing her dissertation. She lectured basic English lessons, ESP courses at different faculties of two distinct universities and also attended some international conferences both as a speaker and a listener. Having completed her Doctor of Philosophy and becoming an expert in language teacher education, she has published academic articles in national and international journals. Furthermore, she has regarded the importance of professional development practices throughout her teaching career. She explicated that she highly regarded the significance of her self-efficacy and commitment to teaching while giving the lessons of English for academic purposes and diplomatic English. She stressed that an indispensable part of foreign language education must be making students question their knowledge and find a correlation between what they aimed to learn and what they already knew. In other words, she claimed that the best learning technique was hidden in their minds waiting to be discovered in order to wield it for long-term language attainments.
Instructor 4 received his academic degrees from English Language and Literature, ELT, and English Translation and Interpretation departments, respectively. Before being an instructor at a state university located in a different city, he acquired teaching certificates through an instructional programme and online distance education. After three years, he started teaching at this university by giving ESP courses besides the lessons for general English and national and international language exams. In addition to some of his published books and articles, he still works as a referee in an international academic journal.

Participants	Name of the course	Course	Type	Students' year of study
I1	Medical English	English for Medical Purposes	EOP	Freshman (first year)
I2	English for Professional Life	English for Business Purposes	EOP	Sophomore (second year)
I3	Academic English	English for Academic Purposes	EAP	Junior (third year)

Table 2: Instructors' courses

ESP courses in Table 2 were called relying on the branches created by Dudley-Evans and Johns (1997) who categorized ESP as English for Academic Purposes (EAP) and English for Occupational Purposes (EOP) with their sub-classes. EAP includes English for Science and Technology, English for Medical Purposes, English for Legal Purposes, and English for Management Finance and Economics. As for EOP, English for Professional Purposes and English for Vocational Purposes would also be grouped under this title. Furthermore, English for Professional Purposes covers English for Medical Purposes (EMP) and English for Business Purposes (EBP), whereas English for Vocational Purposes encompasses Pre-Vocational English and Vocational English. Despite this variety in ESP types, only EMP, EBP, and EAP were involved in this investigation to analyse each of their practices thoroughly.

Data Collection

This study was administered within a postmodern framework by incorporating different data collection instruments and creating relationships or kinships among them (Yin, 2003). Furthermore, in the manner that Atkinson (2000) defines postmodernism as the promise of uncertainty considering the fact that it allows for teachers to question their identity and language besides stressing the weight of criticism in ambivalent nature of knowledge, the study also favoured this approach to finding out experiences and professional behaviours of instructors. To that end, the syllabi of the course, autobiographies, resumes from the official websites, and specially-developed extracurricular materials by the instructors-if there are any-were determined as data collection tools. Additionally, non-participant classroom observations (see Patton, 2002), informal dialogues, opening interviews (see Appendix A) conducted once at the onset of the research were all incorporated in the study after attaining the approvals of two ELT professors in content validity. The prompts for the autobiographies were to cross-check them with the resumes, and not to miss any details regarding the attendees' demographic information.

The observations were conducted using Sendan's (1995) checklist which consists of two parts: the repertory grid elicitation sheet, and repertory grid observation checklist. Moreover, post-observation interview questions (see Appendix B) were also addressed in Turkish, and they took approximately 65 minutes in total (17:49 minutes for I1, 25:24 minutes for I2, and 22:25 minutes for I3). They were adapted from Long (2014) and Teng (2019) after a comprehensive literature review. They were also audio-recorded after obtaining the participants' consents in order to reveal reflections about their teaching performance and to find out the justifications behind their particular behaviours in classes. Thus, the answers of interviewees would lead the researcher to pose follow-up questions to delve deeper into the point. Although Turkish was chosen to make them feel relaxed during their self-expressions, all field notes were written in English.

Participant	Date	Data Collection Tool(s)	Purpose of the Research
I3	11th February 2020	1. Autobiography 2. Opening interview 3. Classroom observation 4. Post-observation interview 5. Teaching documents	Assessment about their teaching methods, reflections to ESP course in the faculty, and teacher development opportunities besides reviewing their stance, manners, and pedagogical content knowledge
I1	17th February 2020	1. Autobiography 2. Opening interview 3. Classroom observation 4. Post-observation interview 5. His critical reflections 6. Casual conversations	
I3	18th February 2020	1. Classroom observation 2. Post-observation interview 3. Her critical reflections	
I3	19th February 2020	1. Informal dialogues through text messaging	
I1	24th February 2020	1. Classroom observation 2. Post-observation interview 3. Informal face to face conversation 4. The main course-book	
I3	25th February 2020	1. Classroom observation 2. Post-observation interview 3. Teaching documents 4. Reflections of the researcher 5. Her critical reflections	
I1	26th February 2020	1. Informal conversation through e-mail	
I2	27th February 2020	1. Autobiography 2. Opening interview 3. Classroom observation 4. Post-observation interview	
I2	28th February 2020	1. Reflections of the researcher 2. Casual conversations	
I1	2nd March 2020	1. Classroom observation 2. Post-observation interview 3. Reflections of the researcher 4. His critical reflections 5. The course-book	
I3	3rd March 2020	1. Classroom observation 2. Post-observation interview	
I2	5th March 2020	1. Classroom observation	
I2	6th March 2020	1. Post-observation interview 2. Reflections of the researcher 3. Her critical reflections 4. Teaching documents	
I1	9th March 2020	1. Classroom observation 2. Post-observation interview 3. Informal face to face conversation	
I3	10th March 2020	1. Classroom observation 2. Post-observation interview 3. Informal face to face conversation	

Table 3: Timeline for data collection

Table 3 portrays the data collection process which took more than one month from February to March 2020. During this period, the researcher was always in touch with the attendees to inquire about their instructions, provide missing information, or to remind them of the schedule.

Data Analysis

For the analysis, though the current study was not based upon a longitudinal design, raw data was reported after every one of the field visits due to their sheer amount (Campion, Palmer and Campion, 1997). When the researcher completed the fieldwork, the interviews, observations, and reflections were all vetted painstakingly to shed light on all facets of the concept. Informed by the former studies (e.g. Teng, 2019), she employed inductive thematic analyses to interpret all the data in the instructors' cases by looking for the patterns within them (Yin, 2003). Initially, she analysed each participant's responses as the first coder, then it was individually reviewed by a colleague with a PhD degree in ELT in a well-regulated way. Those two coders created major and sub-themes by comparison of the data and repeated reading as is highlighted by Merriam (2009). Moreover, cross-case analysis was utilized to integrate or modify the themes within and across the data (Creswell, 2007; Merriam and Tisdell, 2016; Miles and Huberman, 1994; Mills, Durepos and Wiebe, 2010). Subsequently, it was treated *en masse* and juxtaposed through a holistic lens.

These multiple data sets were analysed to ensure thick description (Geertz, 1973; Lincoln and Guba, 1985; Ryle, 1949), trustworthiness (Creswell, 2007), and transferability of the study. Moreover, most of the data collection instruments were implemented at least three times on the same instructor to triangulate the data as cited by Denzin (1978) and Patton (2002) and to enhance credibility and dependability (Lincoln and Guba, 1985). The researcher also used peer debriefing to ensure data validity (Lincoln and Guba, 2002). As suggested by Campion, Palmer and Campion (1997), the researcher filed each instructor's document per week and described the cases in a research journal format for higher validity. She also respected Conway, Jako and Goodman (1995) and handled one to one interviews with standard questions to have the highest reliability.

RESULTS

Perceptions of EFL instructors about ESP courses and echoes to professional identities

In light of the responses to an array of data sources, three major themes and nine sequentially identified sub-themes were tabulated in the matrix to represent perceptions of each instructor to the ESP courses. They were labelled with 'S' or 'D' letters according to similar and different answers of the attendees.

Central themes	Sub-themes	Cases		
		I1	I2	I3
1. Meaning of ESP	Improve language skills	S	S	S
	Academic attribution	S	D	S
	Cultivate a sense of community	S	D	S
	Use of strategies in tasks	S	D	S
2. Management of the course	Adopt top-down strategy	S	S	D
	Unsettled behaviour(s)	S	S	D
3. Convenience of the programme to learners	The correlation between the programme and learner needs	S	S	D
	Mismatching expectations	S	S	D
	Personal and professional growth	S	S	D

Table 4: A cross-case analysis of perceptions

Table 4 indicates all instructors' common views about ESP definitions regarding its feature in improving language skills. However, I1 denoted that his efforts were chiefly to achieve teaching basic language skills and sub-skills rather than enhancing EFL learners' academic literacy in ESP.

I cannot say that I have covered academic English in today's lesson. Only grammar, vocabulary, and pronunciation did constitute the basis of my lecture as in split items. Though the lesson was in EAP or ESP context, they would be touched merely on micro levels (Interview).

Similarly, I3 reported that though her students were second years, they had some difficulties in making a simple sentence in English. Additionally, she mentioned another disjuncture, creating a sense of belonging in the class owing to the lack of interpersonal competence which originated from their diffidence to express themselves effectively. Yet, she attempted to discover whether they were able to utilize the strategies they learned in the lessons while designing an essay through writing assignments. As to I2, she seemed to be content with the senior students since she sensed their academic attributions, team spirit, and the practices of foregoing techniques based on a theoretical framework. She added that she assured herself about their competency in using the target language with the help of presentations and group work, therefore, they could easily notice their weaknesses and meet deficiencies supposing that an error was conducted.

Classroom observations, interviews, field notes revealed that I1 and I2 opted for top-down instructions by plunging into the language as a whole, and immersing them in the big picture, whereas I3 stated that she embraced bottom-up processes by virtue of learners' inefficacy in language use, thereby she felt obliged to take a macro view in teaching. Furthermore, she acknowledged her approach-avoidance tendency during instruction, and hence acknowledged her apprehension and unsettled behaviours now and then.

In general terms, I feel qualified for my position and consider myself fortunate in respect to receiving my doctorate in ELT. I always regard that the more I learn about how to teach, *deepen my content knowledge, and keep myself informed about the current trends in foreign language education, the better I am in conveying information to my students with enthusiasm. Nevertheless, this was my first experience in this faculty and I had not taught EAP before. Moreover, there was not any curriculum or course schedule followed by the former instructor who had retired and I took over that responsibility, thus I was puzzled indeed.*

The responses shaping the third major theme, convenience of the programme to learners, unearthed to what extent they would overlap with one another. Likewise, the sub-themes, correlation between the programme and learner needs, and personal, professional growth of learners strengthened this parallelism. Accordingly, I3 highlighted:

I tried to keep the lessons conforming to hearsay reports of other lecturers and former English instructors. What's more, only a few students got into class due to lack of compulsory attendance as a matter of the decision taken by the authorities in the faculty. As a result, I could not conduct a needs analysis at the outset of my lesson. I somehow managed to survive the class by exploring potentially the best techniques related to the topic that I determined weekly on my own (field notes).

Regarding EFL instructors' perceptions of ESP practices and professional identities, their values of teaching, experiences, beliefs, professional roles, and knowledge came forth as the cardinal factors. To put it more clearly, their awareness allowed them to inquire about their mindsets, philosophy, and classroom practices judiciously. In what follows, they mostly seemed to respect their performance, attitudes towards the lesson, and subject knowledge critically from a supervisor or mentor's point of view. As a consequence, I1 and I2 appeared to have strong perceptions of their own professional identities by internalizing ESP practices thanks to overlapping expectations, priorities, and achieving classroom management skills, while I3 did not seem to adjust herself to the course and its entailments. Namely, her perceived professional identity was not revealed to be in parallel with her perceptions of the course.

Instructors' self-efficacy beliefs, classroom practices, and professional identities

Instructors' sense of self-efficacy that shape their teaching perspective, approach, and belief was scrutinized in conjunction with professional identity and classroom routines to unearth whether they are correlated and if so how.

Themes	Issues	Case(s)
1. Critical thinking	Socratic questioning, elicitations, thought-provoking tasks	I1, I2
2. Technology Literacy	Google classroom, games, social media, web applications	I1, I2, I3
3. Creativity	Sparking curiosity, extracurricular materials, differentiated instructions	I1, I2
4. Flexibility	Alternative strategies or techniques	I1, 12, 13
5. Collaboration	Drama	I1, I2

Table 5: Instructors' sense of self-efficacy and practices

Socratic questioning, elicitations, concept-check questions were listed by I1 as his core principles concerning the first theme. Similar to I1, I2 reported that thought-provoking tasks comprised her indispensable techniques to enhance the analytic approach and independence. I2 particularly emphasized this in her opening interview:

Journalism calls for a discipline of mind, and analytic thinking, which will then be reflected in language learning skills.

Nonetheless, I3 did not give consequence to this point in her interviews or observation notes nor did memos unpack any information about critical thinking as a motive embedded in skills. I1's use of Google classroom, I2's Kahoot games in addition to her attempts to meet the following requirements of the 21st century, such as the adoption of social media rather than print media, teaching how to write an e-mail instead of a letter, the design of a text depending on Associated Press (AP) style, and finally, the web applications of I3 proved that they were all into technology-assisted language teaching. As for creativity, I1 emphasized the importance of sparking inquiry while avoiding rote-learning strategies by providing background information at first about medical science, and terminologies, such as gastroscopy. In fieldwork, he once drew a stomach to explain the examination process with an endoscope and wrote the following questions on the board: "what preparation do I need to do?, what can I expect after a gastroscopy?, is gastroscopy reliable?, are there any side-effects or complications from having a gastroscopy?" He also pointed this out during the interview that followed:

Because they are first years, and hence they do not know anything about medicine, I cannot expect them to remark on specific terms or treatments stated in the course-book. Though it is not my field of study, I feel obliged to keep them informed as much as I can, seeing that I am not the slave of any books, and there is no best teaching material on the markets to improve your teaching quality in class. Teachers are the only ones who can make the lesson perfect through student-centred teaching.

I2 took the issue further with her claims that utilizing extracurricular materials, differentiated instructions, peripheral learning, and reducing teacher talk time to motivate learners besides considering their capability of gauging the appropriateness of the activity were ways in which to improve creative teaching. Likewise, she stated that thanks to her profound knowledge in NGO, she had a high level of academic self-efficacy.

Another pioneering aspect seemed to be nurturing flexibility and adaptability to facilitate learner autonomy via minimizing the salience of dependence, but increasing self-control strategies, and self-driven problem-solving skills. To that end, I1 assigned them to read academic articles and write summaries in essay formats, and in turn, he skipped the reading and writing tasks in the course book. At that point, he affirmed:

Not every teacher can accomplish fostering learner autonomy by scaffolding academic literacy as much as possible (interview).

I2 claimed to have pedagogy of practicality in that she would instantly propose alternative ways to receive learner output, such as guiding them to write when they failed in speaking to be able to iterate their attainment. Similarly, I3 slightly mentioned her beliefs in designing an academic writing course prompting learners to take charge of their learning and decision-making.

The last theme, collaboration, appeared due to the drama method handled by I1, and I2 also used imparting knowledge to empower team spirit in a group setting as well as operating poster presentations through which she endeavoured to self-actualize herself. She echoed that this self-efficacy relied on her work experience in the United Nations. On the other hand, she declared that owing to the dominancy of teacher-centred education in her class, she could not manage to stimulate learners enough to take the floor, which was observed during fieldwork, and which created a discrepancy with the results.

Concerning the effects of self-efficacy on professional identity, teacher agency, and beliefs were explored as basic themes from the responses of the participants as is seen in Table 6.

Themes	Issues	Case(s)
1. Teacher agency	Teacher awareness, readiness to learn	I1
2. Teacher beliefs	Identifying macro-level plans, using integrated skills, learner-centred education	I1

Table 6: Sense of self-efficacy and professional identity

Even though the perceived self-efficacy of I1 decreased at times as was emphasized in the analysis of the previous research question. He must be given due recognition for showing high self-efficiency during the observations when he cleverly managed to disentangle himself from having to explain a word for which he could not remember the meaning. This is indicative of his constructive manner and propensity to persevere. Furthermore, he recognized the importance of exploring learners' ways of thinking, conceptualizations, reasoning, and readiness to learn; hence he managed to generate accurate predictions about the points where they would get lost thanks to his agency.

In his second lesson, having introduced common prescription abbreviations of drugs, such as measurements (mg, ml), frequencies (on, bd, tds, qds, 4-6h), and the routes (IV, IM, SC, PR, INH, NEB), he integrated power-point presentations along with audio records to elucidate the abbreviations with other examples. Only then did he ask learners to read the chart containing several abbreviations with which to make sentences. Similarly, his sufficient subject matter knowledge in this profession was detected while teaching noun clauses, adverbial clauses, word formation process, and modal verbs, though they were not covered in the syllabus, and instantly appeared with learners' immediate needs. Thus, his beliefs in delivery techniques were built upon establishing macro-level plans rather than designing the lesson from the ground up.

As to another instructor, 12, her perceived self-efficacies could not be observed during the lectures. The data gathered from the instruments revealed that she focalized on knowledge transmission instead of knowledge construction. Moreover, she tried to learn through reflection-on-doing rather than being in an attempt to identify a permanent solution, build a career development plan to remedy her deficiencies, and thus have a professional identity. The other classroom management problem she experienced and that was noted, was keeping the learners silent in class when others held the floor now and then. Consequently, she resorted to threatening:

Please be quiet! Otherwise, I may get annoyed. What happens when I get annoyed? I cannot concentrate on my profession. Therefore, while I am reading your papers, I might grade you down!… (Notes from the fieldwork)

Professional competencies and identities of the instructors

Individualized teaching, diagnosed as the only central theme, embodies some pivotal constituents, such as; attitudes, leadership, reflectivity, knowledge, expectations, and CPD activities. Therefore, it necessitates teachers to be self-directed in order to attain both skills and PCK as a life-long learning requirement. I3 mentioned that to maintain professional discipline, academic commitment, and keep abreast of recent trends in the field, which she has had to do with reading research papers, attending seminars, and writing articles even though none of which could be observed by the researcher in any judicious applications of her academic skills, such as academic writing or with reference to any corresponding domains of competency of the teacher. She was also unable to justify her specific practices in the class by alleging that these behaviours were of little consequence. In like manner, I2 confirmed that she did not have any professional competence to shape her identity, since she majored in journalism. Therefore, she could not get value out of teaching EBP, and thus she did not regard it as a direct investment in her teacher identity. Moreover, she sometimes felt lost while choosing the correct teaching method and managing the class when learners could not answer any questions even in the follow-up session, which also proved her problem in teacher agency.

Contrary to the others, I1 underlined the strong correlation between professional competence, professional development, self-awareness, and knowledge construction seeing that he hitherto attended some conferences on Socratic questioning. Consequently, he availed himself of the training and was successful in implementing this strategy in his ESP classes besides giving seminars referring to this issue. Additionally, he was reported to be mindful of his learners and their differentiated needs and enhancing their engagement in the lesson by setting dynamic goals as well as facilitating their inquiry skills.

DISCUSSION

Throughout the research, I1 remained in the forefront due to the fact that he perpetually accentuated his strong sense of self-efficacy by providing structured reasoning; creativity, technology literacy, and cooperative learning opportunity to learners in class (see Table 4). In concert with the deductions of Basturkmen (2012), the correspondence between I1's beliefs and autonomy besides its reflection to classroom practices indicated his flexibility in teaching and his engagement in rational thinking as was displayed in Table 5. This is also in line with the assertion of Charters (1976), Derakhshan et al. (2020) and Pajares (1992) in that teacher beliefs, efficacy, and competencies would attribute to their professional independence. His stated professional competence attempts to increase CPD activities via attending conferences, maintaining the position in a publishing house, and reflecting his attainments to teachers or prospective teachers through seminars, and most importantly his awareness towards the requirements of his profession, displayed the power between two concepts. Additionally, his justification of positive perceptions and self-confidence in lecturing ESP was substantiated by Atkinson et al. (1987), Day (2002) and Qi, Sorokina and Liu (2021) in that teachers' sense of professional identity combined in with their commitment and self-efficacy since they symbolize teachers' professional prospects concerning their career. Furthermore, congruent with Dikilitaş and Mumford (2019) and Lennert da Silva and Mølstad (2020), I1 exposed the consistency between professional development activities and autonomy. As such, it would be compatible to postulate that his perceived ability deeply coincided with his classroom practices and then formed the backbone of a range of intertwined characteristics of his professional teacher identity.

The critical stance of I1 in exploiting differentiated instruction revealed his capability (see Table 6), self-efficacy, problem-solving skill (Deci and Ryan, 2000; Richards, 2021), and teacher competence in respect of his professional identity. By the same token, he synthesized topics along with fine-tuning of their functionality in the target language, and regarded the gravity of motivation and autonomization process for learners by avoiding spoon-feeding. Furthermore, similar to Buchanan (2015), Lasky (2005), McNicholl (2012), and Vähäsantanen (2015), I1 displayed how interdependent the agency, professionalism, and teacher identity were indeed in addition to verifying that teacher competencies were in tandem with their professional identity in the Turkish context. In more concrete terms, the professional competence of I1 turned out to be a predictor of his 'professional wellbeing' (Lauermann and König, 2016: 18). Finally, in a similar manner to Hurt-Avila and Castillo (2017), I1 bolstered the scholars to allege interplay between professional competence and teachers' identity in the current study.

As to I2, a majority of her recounting was detected to be contradictory. She mentioned that she occasionally executed student-led learning activities, such as arranging their NGO within a group development plan. She asked them to define the details of foundation (vision, mission, objectives), resource gathering capacity (existing and required resources, strategy development, accounting controls, and budget prioritization), adaptability (strength, weakness, opportunities, and threats analysis), and communication, networking and advocacy after conducting a pilot activity and creating a template. Correspondingly, she was able to embed extracurricular activities in the course, hold pedagogy of practicality, organize a collaborative learning environment to build a sense of community, and above all encouraged learners to openly share their views as is seen in Table 5. Yet, even though her sense of self-efficacy looked similar to I1, the reports in informal dialogues and memos disclosed the adoption of direct instructions in teaching, and the difficulty of reducing teacher talking time. The researcher also extrapolated from the fieldwork that she embarked on substitution drills to teach the language functions. It reveals the fact that she did not have a comprehensive or solid grasp of PCK. In other words, she was not able to specialize her attainments in an educational setting to cultivate creativity and practice efficient activities to increase student outcomes. However, she declared that she appreciated her content knowledge in NGO which enabled her to have such a high sense of efficacy in teaching while referring to her perceptions of giving an ESP course.

I2 could not display specific subject matter knowledge of EOP, thus she sided with Górska-Poręcka (2013) due to the requirements of PCK on ESP practitioners in terms of having professional knowledge. It follows that she disaccorded with Blömeke and Delaney (2012). Likewise, her responses cannot attest to the existence of her strong beliefs, nor was its reflection to professional teacher identity observed as Table 6 indicates. Therefore, her case was in contrast with Ibarra (1999), Schein (1978) and Qi, Sorokina and Liu (2021) as well.

Different from the other two instructors, I2 had put forth the academic attributions of learners in the class and also mentioned the convenience of the programme with their needs (see Table 4). This was despite her counterclaims that she could not expect them to accomplish the objectives of the lesson thoroughly due to their low level of proficiency in the English language. Moreover, in the way that Newton and Newton (2001) emphasized, I2 could address neither casual nor subject-relevant questions. Yet, they are of importance in guiding learners to think analytically and activate their subject knowledge, which must have ensued from her paucity of pedagogical background. Similarly, as Derakhshan et al. (2020) put it, the impact of professional teacher identity is deeply reflected upon the decisions they make about their teaching techniques and the pedagogical choices. Thereby, she could not administer the lesson properly considering the lack of professional help, the praise of rote-learning strategy, her confession about obtaining no benefit from the course, unsatisfactory concept-check questions, reiterations of quite similar questions, and failure in enabling a smooth transition from NGO to describing how to write a business plan. Scientifically speaking, her case clashed with Atkinson et al. (1987), Day (2002), Buchanan (2015), Lasky (2005), McNicholl (2012), and Vähäsantanen (2015) in that her sense of high self-efficacy, which may have appeared thanks to the absence of vicarious experience, did not have a part in teacher agency, and professional identity. Nonetheless, in concert with Hurt-Avila and Castillo (2017), her tenuous professional competence attributed to the shortage of quality in teacher identity besides the interconnection between her perceptions and self-efficacy which reinforced the findings of Austin et al. (2015). Finally, there were two main areas where her professional identity did not correlate with her performance.

These were her perceived ability and her emotional self-efficacy to boost learner outcomes as an integral part of this concept.

The last participant, I3 was identified as being the least mentioned instructor in the tables owing to her unassertive manner and self-restraint to the investigation. It must derive from conspicuous problems with course management, job involvement at a new faculty, unsettled behaviours in adopting a set of procedures, serious difficulties in selecting supplementary materials and seeking a way to change the current, elusive syllabus. Hence, her endeavour was mostly to make the lesson look attractive by using integrated skills and student-led activities as well as effective and practical materials, such as visuals, audio-visuals and technology. Her perceived high level of professional identity was directly related to her being a master of writing skills and other key issues in ELT. Still, she neither managed to display her independence or agency, nor did she adjust expectations to the current students on account of her lack of experience in this department. This was dissimilar to the other two instructors in the research. Taking up a new course, ESP should have allowed room to probe strategies to manage this lesson much better.

It can be straightforwardly asserted that I3's beliefs enacted as the saying goes a *self-fulfilling prophecy*. To put it more clearly, her low perception, self-efficacy, and professional competence all correlated and thus disclosed that her beliefs directly influenced her teaching performance, behaviours, and the trajectory of the course despite her so-called high level of professional identification. Likewise, I3 experienced a similar case with I2 in the fact that she could not utilise subject knowledge due to being incompetent at general content knowledge and low professional competence; therefore, they both failed to self-actualize themselves.

I3 conflicted with Atkinson et al. (1987), Day (2002) and Richards (2021) in that her perceived professional identity did not go hand in hand with her self-efficacy. She was in discord with Dikilitaş and Mumford (2019) and Engeness (2021) because of her inconsistency in autonomy, professional development activities and professional knowledge (see Derakhshan et al., 2020 and Eraut, 1994) which she reported to be the requisite of CPD in teacher education. Notwithstanding this fact, she supported Buchanan (2015), Lasky (2005), McNicholl (2012), and Vähäsantanen (2015) since her low agency, as a part of self-efficacy (see Table 6), and professional competence were compatible to one another. In a similar vein, she was in good agreement with Austin et al. (2015) thanks to the parallelism between her low-level perceptions, and self-efficacy beliefs. She also upheld Hurt-Avila and Castillo (2017) due to a linear, strong link between professional incompetence and professional identity with a low level of respect. She compromised with Charters (1976) and Pajares (1992) on account of her inability to display independence.

In short, considering the teachers' judgments, experience, values, and 'standards' (Varghese et al., 2005), all ESP practitioners indicated the correlation between self-efficacy beliefs and teacher perception as alleged by Austin et al. (2015). By the same token, similar to Pantic, Wubbels and Mainhard (2011), versatile dimensions of professional competence were disclosed by appraising teacher knowledge, professional knowledge, attitude, problem-solving skill, and professional identity. Taken together, professional competence turned out to be the leading factor in describing professional identity (Lauermann and König, 2016).

CONCLUSION

In this case study, three EFL instructors' discrepancies in perceptions of ESP courses, self-efficacy, and their impacts on teaching practices as well as professional identity were evident from a variety of facets, such as teacher agency, beliefs, critical thinking, technology literacy, creativity, flexibility, collaboration, and management of the class. The analyses were carried out through a set of data collection instruments and melting the gathered data in the same pot. At the end of the research, all EFL instructors' perceived professional identities were confirmed to be high. I1 managed to demonstrate his teacher identity much better through linear associations with professional competence, self-efficacy, and perceptions about the course and merit of his identity. Consequently, the intricacy of ESP teacher identity was unfolded as in the study of Tao and Gao (2018).

Motives behind the disjuncture of the other two instructors in identifying their teacher identities can be categorized distinctly. According to data analysis, both were professionally less competent in the teaching profession. The basis of I2's problem can be appointed to her self-assurance and self-esteem, whereas I3's problem was due to her low degree of perceptions towards practicing the course and her own professional identity. That is, I2's vigorous effort to maintain the lessons was predicated on her alleged self-efficacy and capacity which fell behind the pacing of the lesson, while I3 affirmed to have provided consistent results thanks to her values towards competence, efficacy beliefs, and perceptions. In conclusion, professional competence was reported to be the best representative of professional teacher identity.

The importance of PCK and specific subject matter knowledge to attain professional competence (Blömeke and Delaney, 2012; Górska –Poręcka, 2013; Shulman, 1987) was illuminated throughout the analysis. This is because they would both pave the way for CPD activities (Steiner, 2010) and the repercussions of teacher perception on raising awareness of professional identity amidst instructors (Atkinson et al., 1987). In this way, the multi-faceted (Duff and Uchida, 1997; Olsen, 2008), and non-linear nature (Yuan and Lee, 2016) of identity issue was clarified.

After exploring professional identity, and teachers' jurisdictions about the profession regarding the analyses of Buchanan (2015), Varghese et al. (2005) through the qualitative case study method as is required in the research paradigm (Tschannen-Moran, Hoy and Hoy, 1998), some suggestions for further considerations ought to be uttered as well. Initially, future studies can systematically examine professional competence, CPD practices, and their potential effects on professional teacher identity. Moreover, professional identity might be investigated with some other characteristics, such as teachers' years of experience, and motivation levels (Ibarra, 1999; Schein, 1978). Additionally, 'situated identities' (Alsup, 2006) can be closely associated with the sense of belonging in detail by putting more emphasis on assessing learners' co-working opportunities in class in light of social constructivism (Vygotsky, 1978; Wenger, 1998).

Vicarious experience as an essential source of self-efficacy beliefs also cannot be left unattended while delving into professional identity in teacher education. Thus, care must be taken to this issue on a deeper level, which seems to be underrepresented in recent research. Last but certainly not least, instructors' identity development and changes may be investigated by stretching the data collection process over a long period to determine the trajectory of ESP courses, shed light on necessary amendments in the curriculum, and finally offer supplementary materials to be adopted by practitioners.

As for the limitations of the study, some quantitative measurements might be involved to triangulate the data, or data collection may be regulated periodically at different times over a year to unveil their perceptions, senses, and anticipations of teachers over the long haul. In addition, the number of participants could be increased to four to interpret the cases with data-enrichment tools.

ACKNOWLEDGMENT

I gratefully thank anonymous instructors for participating in the study, and the colleague who has shared invaluable comments with me during the data collection and analysis process. Finally, I would like to express my gratitude and appreciation to Melanie Gradden who has painstakingly proofread the article.

REFERENCES

Alsup, J. (2006) *Teacher identity discourses: Negotiating personal and professional spaces*, London: Routledge.

Atkinson, E. (2000) 'The Promise of Uncertainty: Education, Postmodernism and the Politics of Possibility', *International Studies in Sociology of Education*, Vol. 10, No. 1, pp. 81–99. https://doi.org/10.1080/09620210000200050

Atkinson, R. L., Atkinson, R. C., Smith, E. E. and Hilgard, E. R. (1987) *Introduction to psychology*, 9th edition, San Diego, New York: Harcourt Brace Jovanovich Publishers.

Austin, B., Adesope, O. O., French, B. F., Gotch, C., Bélanger, J. and Kubacka, K. (2015) 'Examining School Context and Its Influence on Teachers: Linking Talis 2013 with PISA 2012 Student Data', *OECD Education Working Papers*, No. 115. http://dx.doi.org/10.1787/5js3f5fgkns4-en

Babanoğlu, M. P. and Ağçam, R. (2019) 'Turkish EFL Teacher Candidates' Early Teacher Identity' *Asian Journal of Education and Training*, Vol. 5, No. 2, pp. 386–391. https://doi.org/10.20448/journal.522.2019.52.386.391

Bandura, A. (1997) *Self-efficacy: The exercise of control*, New York, NY: W.H. Freeman.

Basturkmen, H. (2012) 'Review of Research into the Correspondence between Language Teachers' Stated Beliefs and Practices', *System*, Vol. 40, No. 2, pp. 282–295. https://doi.org/10.1016/j.system.2012.05.001

Beijaard, D., Verloop, N. and Vermunt, J. D. (2000) 'Teachers' Perceptions of Professional Identity: An Exploratory Study from A Personal Knowledge Perspective', *Teaching and Teacher Education*, Vol. 16, No. 7, pp. 749–764. https://doi.org/10.1016/S0742-051X(00)00023-8

Blömeke, S. and Delaney, S. (2012) 'Assessment of Teacher Knowledge across Countries: A Review of the State of Research', *ZDM Mathematics Education*, Vol. 44, pp. 223–247. https://doi.org/10.1007/s11858-012-0429-7

Buchanan, R. (2015) 'Teacher Identity and Agency in an Era of Accountability', *Teachers and Teaching*, Vol. 21, No. 6, pp. 700–719. https://doi.org/10.1080/13540602.2015.1044329

Campion, M. A., Palmer, D. K. and Campion, J. E. (1997) 'A Review of Structure in the Selection Interview', *Personnel Psychology*, Vol. 50, No. 3, pp. 655–702. https://doi.org/10.1111/j.1744-6570.1997.tb00709.x

Charters, W. W. (1976) *Sense of teacher work autonomy: Measurement & findings*, Eugene, OR: Center for Educational Policy and Management, University of Oregon.

Conway, J. M., Jako, R. A. and Goodman, D. F. (1995) 'A Meta-Analysis of Interrater and Internal Consistency Reliability of Selection Interviews', *Journal of Applied Psychology*, Vol. 80, No. 5, pp. 565–579. https://doi.org/10.1037/0021-9010.80.5.565

Cooney, T. J. and Shealy, B. E. (1997) 'On understanding the structure of teachers' beliefs and their relationship to change', in Fennema, E. and Nelson, B. S. (ed.) *Mathematics teachers in transition* (pp. 87–109), Mahwah, NJ: Lawrence Erlbaum Associates.

Creswell, J. (2007) *Qualitative inquiry & research design*, London: Sage Publications.

Day, C. (2002) 'School Reform and Transitions in Teacher Professionalism and Identity', *International Journal of Educational Research*, Vol. 37, No. 8, pp. 677–692. https://doi.org/10.1016/S0883-0355(03)00065-X

Deci, E. L. and Ryan, R. M. (2000) 'Self-determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-being, *American Psychologist*, Vol. 55, No. 1, pp. 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Derakhshan, A., Coombe, C., Arabmofrad, A. and Taghizadeh, M. (2020) 'Investigating the Effects of English Language Teachers' Professional Identity and Autonomy in Their Success', *Issues in Language Teaching*, Vol. 9, No. 1, pp. 1–28. https://doi.org/10.22054/ilt.2020.52263.496

Denzin, N. K. (1978) *Sociological methods: A sourcebook*, New York, NY: McGraw-Hill.

Dikilitaş, K. and Mumford, S. E. (2019) 'Teacher Autonomy Development through Reading Teacher Research: Agency, Motivation and Identity', *Innovation in Language Learning and Teaching,* Vol. 13, No. 3, pp. 253–266. https://doi.org/10.1080/17501229.2018.1442471

Dudley-Evans, T. and John, M. (1997) *Developments in English for specific purposes: A multi-disciplinary approach*, Cambridge: Cambridge University Press.

Duff, P. A. (2008) *Case study research in applied linguistics*, New York, NY: Lawrence Erlbaum Associates.

Duff, P. A. and Uchida, Y. (1997) 'The Negotiation of Teachers' Sociocultural Identities and Practices in Postsecondary EFL Classrooms', *TESOL Quarterly*, Vol. 31, No. 3, pp. 451–486. https://doi.org/10.2307/3587834

Engeness, I. (2021) 'Developing Teachers' Digital Identity: Towards the Pedagogic Design Principles of Digital Environments to Enhance Students' Learning in the 21st Century', *European Journal of Teacher Education*, Vol. 44, No. 1, pp. 96–114. https://doi.org/10.1080/02619768.2020.1849129

Eraut, M. (1994) *Developing professional knowledge and competence*, London: Falmer Press.

Firestone, W. A. (1993) 'Alternative Arguments for Generalizing from Data as Applied to Qualitative Research', *Educational Researcher*, Vol. 22, No. 4, pp. 16–2. https://doi.org/10.2307/1177100

Fotovatian, S. (2015) 'Language, Institutional Identity and Integration: Lived Experiences of ESL Teachers in Australia', *Globalisation, Societies and Education*, Vol. 13, No. 2, pp. 230–245. https://doi.org/10.1080/14767724.2014.934072

Geertz, C. (1973) *The interpretation of cultures*, New York: Basic Books.

Górska-Poręcka, B. (2013) 'The Role of Teacher Knowledge in ESP Course Design', *Studies in Logic, Grammar and Rhetoric*, Vol. 34, No. 1, pp. 27–42. https://doi.org/10.2478/slgr-2013-0021

Harun, R. N. S. R. (2019) 'The Construction and Reconstruction of ESL Student Teachers' Professional Identity', *Studies in English Language and Education*, Vol. 6, No. 1, pp. 1–12. https://doi.org/10.24815/siele.v6i1.13073

Henson, R. K. (2002) 'From Adolescent Angst to Adulthood: Substantive Implications and Measurement Dilemmas in the Development of Teacher Efficacy Research', *Educational Psychologist*, Vol. 37 No. 3, pp. 137–150. https://doi.org/10.1207/S15326985EP3703_1

Huang, J. (2009) *Autonomy, agency and identity in foreign language learning and teaching* [Unpublished PhD Thesis], Hong Kong: The University of Hong Kong.

Hurt-Avila, K. M. and Castillo, J. (2017) 'Accreditation, Professional Identity Development, and Professional Competence: A Discriminant Analysis', *Journal of Counsellor Leadership and Advocacy*, Vol. 4, No. 1, pp. 39–51. https://doi.org/10.1080/2326716X.2017.1282331

Ibarra, H. (1999) 'Provisional Selves: Experimenting with Image and Identity in Professional Adaptation', *Administrative Science Quarterly*, Vol. 44, No. 4, pp. 764–791. https://doi.org/10.2307/2667055

Lasky, S. (2005) 'A Sociocultural Approach to Understanding Teacher Identity, Agency and Professional Vulnerability in A Context of Secondary School Reform', *Teaching and Teacher Education*, Vol. 21, No. 8, pp. 899–916. https://doi.org/10.1016/j.tate.2005.06.003

Lauermann, F. and Konig, J. (2016) 'Teachers' Professional Competence and Wellbeing: Understanding the Links between General Pedagogical Knowledge, Self-efficacy and Burnout', *Learning and Instruction*, Vol. 45, pp. 9–19. https://doi.org/10.1016/j.learninstruc.2016.06.006

Lennert da Silva, A. L. and Mølstad, C. E. (2020) 'Teacher Autonomy and Teacher Agency: A Comparative Study in Brazilian and Norwegian Lower Secondary Education', *The Curriculum Journal*, Vol. 31, No. 1, pp. 115–131. https://doi.org/10.1002/curj.3

Lincoln, Y. S. and Guba, E. G. (1985) *Naturalistic inquiry*, Newbury Park, CA: Sage Publications.

Lincoln, Y. and Guba, E. (2002) 'The only generalization is: There is no generalization', in Gomm, R., Hammersley, M. and Foster, P. (ed.) *Case study method* (pp. 27–44), London: Sage.

Long, N. (2014) *Teacher autonomy in a context of Chinese tertiary education: case studies of EFL teachers*, [Online], Available: https://repository.hkbu.edu.hk/etd_oa/103 [12 Apr 2020]

McNicholl, J. (2012) 'Relational Agency and Teacher Development: A CHAT Analysis of A Collaborative Professional Inquiry Project with Biology Teachers', *European Journal of Teacher Education*, Vol. 36, No. 2, pp. 218–232. https://doi.org/10.1080/02619768.2012.686992

Merriam, S. B. (2009) *Qualitative research and case study applications in education: Revised and expanded from case study in education*, San Francisco, CA: Jossey-Bass.

Merriam, S. B. and Tisdell, E. J. (2016) *Qualitative research: A guide to design and implementation*, 4th edition, San Francisco, CA: Jossey-Bass.

Miles, M, B. and Huberman, A. M. (1994) *Qualitative data analysis: An expanded Sourcebook*, 2nd edition, Thousand Oaks, CA: Sage.

Mills, A. J., Durepos, G. and Wiebe, E. (ed.) (2010) *Encyclopaedia of case study research*, Thousand Oaks, CA: Sage Publications. http://dx.doi.org/10.4135/9781412957397

Newton, D. P. and Newton, L. D. (2001) 'Subject Content Knowledge and Teacher Talk in the Primary Science Classroom', *European Journal of Teacher Education*, Vol. 24, No. 3, pp. 369–379. https://doi.org/10.1080/02619760220128914

Nunan, D. and Bailey, K. M. (2010) *Exploring second language classroom research: A comprehensive guide*, Beijing: Foreign Language Teaching and Research Press.

Olsen, B. (2008) *Teaching what they learn, learning what they live*, Boulder, CO: Paradigm Publishers.

Pajares, M. F. (1992) 'Teachers' Beliefs and Educational Research: Cleaning up A Messy Construct', *Review of Educational Research*, Vol. 62, No. 3, pp. 307–332. https://doi.org/10.3102/00346543062003307

Pantic, N., Wubbels, T. and Mainhard, T. (2011) 'Teacher Competence as A Basis for Teacher Education: Comparing Views of Teachers and Teacher Educators in Five Western Balkan Countries', *Comparative Education Review*, Vol. 55, No. 2, pp. 165–188. https://doi.org/10.1086/657154

Patton, M. Q. (2002) *Qualitative research and evaluation methods*, 3rd edition, London: Sage Publications.

Qi, W., Sorokina, N. and Liu, Y. (2021) 'The Construction of Teacher Identity in Education for Sustainable Development: The Case of Chinese ESP Teachers', *International Journal of Higher Education*, Vol. 10, No. 2, pp. 284–298. https://doi.org/10.5430/ijhe.v10n2p284

Richards, J. C. (2021) 'Teacher, Learner and Student-Teacher Identity in TESOL', *RELC Journal*, pp.1–15. https://doi.org/10.1177/0033688221991308

Robinson, J. E. and Norris, N. F. J. (2001) 'Generalisation: The Linchpin of Evidence Based Practice?', *Educational Action Research*, Vol. 9, No. 2, pp. 303–310. https://doi.org/10.1080/09650790100200155

Ross, J. A. (1995) 'Strategies for Enhancing Teachers' Beliefs in Their Effectiveness: Research on A School Improvement Hypothesis', *Teachers College Record*, Vol. 97, No. 2, pp. 227–251.

Ruddin, L. P. (2006) 'You Can Generalize Stupid! Social Scientists, Bent Flyvbjerg, and Case Study Methodology', *Qualitative Inquiry*, Vol. 12, No. 4, pp. 797–812. https://doi.org/10.1177/1077800406288622

Ryle, G. (1949) *The concept of mind*, London: Hutchinson.

Sachs, J. (2005) 'Teacher education and the development of professional identity: Learning to be a teacher', in Denicolo, P. and Kompf, M. (ed.) *Connecting policy and practice: Challenges for teaching and learning in schools and universities* (pp. 5–21), Oxford: Routledge.

Salinas, D. and Ayala, M. (2017) 'EFL Student-Teachers' Identity Construction: A Case Study in Chile', *HOW*, Vol. 25, No. 1, pp. 33–49. https://doi.org/10.19183/how.25.1.380

Schein E. H. (1978) *Career dynamics. Matching individual and organizational needs*, Reading, MA: Addison-Wesley.

Sendan, F. (1995) *Patterns of development in EFL student teachers' personal theories: A constructivist approach*, [Unpublished PhD Thesis] Reading: University of Reading.

Shulman, L. (1987) 'Knowledge and Teaching: Foundations of the New Reform, *Harvard Educational Review*, Vol. 57, No. 1, pp. 1–23. https://doi.org/10.17763/haer.57.1.j463w79r56455411

Skaalvik, E. M. and Skaalvik, S. (2011) 'Teachers' Feeling of Belonging, Exhaustion, and Job Satisfaction: The Role of School Goal Structure and Value Consonance', *Anxiety, Stress & Coping*, Vol. 24, No. 4, pp. 369–385. https://doi.org/10.1080/10615806.2010.544300

Smylie, M. A. (1996) 'From Bureaucratic Control to Building Human Capital: The Importance of Teacher Learning in Education Reform', *Educational Researcher*, Vol. 25, No. 9, pp. 9–11. https://doi.org/10.3102/0013189X025009009

Steiner, L. (2010) 'Using competency-based evaluation to drive teacher excellence: Lessons from Singapore', *Public Impact*, [Online], Available: https://opportunityculture.org/images/stories/singapore_lessons_2010.pdf [26 Mar 2020].

Tao, J. and Gao, X. (2018) 'Identity Constructions of ESP Teachers in A Chinese University', *English for Specific Purposes*, Vol. 49, pp. 1–13. https://doi.org/10.1016/j.esp.2017.09.003

Tellis, W. M. (1997) 'Application of A Case Study Methodology', *The Qualitative Report*, Vol. 3, No. 3, pp. 1–19. https://doi.org/10.46743/2160-3715/1997.2015

Teng, M. F. (2019) *Autonomy, agency, and identity in teaching and learning English as a foreign language*, Singapore: Springer.

Tschannen-Moran, M. and Hoy, A. W. (2001) 'Teacher Efficacy: Capturing an Elusive Construct', *Teaching and Teacher Education*, Vol. 17, No. 7, pp. 783–805. https://doi.org/10.1016/S0742-051X(01)00036-1

Tschannen-Moran, M., Hoy, A. W. and Hoy, W. K. (1998) 'Teacher Efficacy: Its Meaning and Measure', *Review of Educational Research*, Vol. 68, No. 2, pp. 202–248. https://doi.org/10.3102/00346543068002202

Vähäsantanen, K. (2015) 'Professional Agency in the Stream of Change: Understanding Educational Change and Teachers' Professional Identities', *Teaching and Teacher Education*, Vol. 47, pp. 1–12. https://doi.org/10.1016/j.tate.2014.11.006

Varghese, M., Morgan, B., Johnston, B. and Johnson, K. A. (2005) 'Theorizing Language Teacher Identity: Three Perspectives and Beyond', *Journal of Language, Identity, and Education*, Vol. 4, No. 1, pp. 21–44. https://doi.org/10.1207/s15327701jlie0401_2

Vygotsky, L. S. (1978) *Mind in society: The development of higher psychological processes*, Cambridge, MA: Harvard University Press.

Wenger, E. (1998) *Communities of practice: Learning, meaning, and identity*, Cambridge: Cambridge University Press.

Woodside, A. G. (2010) *Case study research: Theory, methods, practice*, Bingley, UK: Emerald.

Yin, R. K. (2003) *Case study research: Design and methods*, 3rd edition, Thousand Oaks, CA: Sage.

Yazan, B. (2016) 'Early Career EFL Teachers' Instructional Challenges', *Journal of Theory and Practice in Education*, Vol. 12, No. 1, pp. 194–220.

Yuan, R. and Lee, I. (2016) 'I Need to Be Strong and Competent: A Narrative Inquiry of a Student-Teacher's Emotions and Identities in Teaching Practicum', *Teachers and Teaching*, Vol. 22, No. 7, pp. 819-841. https://doi.org/10.1080/13540602.2016.1185819

APPENDIX

APPENDIX A. OPENING INTERVIEW

1. What are the qualities of a good teacher?
2. How can you portray yourself as an EFL instructor?
3. What are the perceptions of other instructors about your profession?
4. Can you talk about your (continuing) professional development activities? What are they principally oriented to? Why would you like to be informed about these issues?
5. Do you think that you can implement these ideas in class? How?
6. Do you think that these activities influence your identity in real terms?
7. How do you describe your professional teacher identity?
8. How do you invest in your professional teacher identity?
9. To what extent do you believe that your professional development "adventure" affects the learners?
10. How do you plan your teaching career as an instructor in the near future?

APPENDIX B. POST-OBSERVATION INTERVIEW

1. What do you think about your teaching performance today?
2. What were the things you adopted from your professional learning practices and implemented in the lesson?
3. Can you refer to any justifications of your specific behaviours in class?
4. What was the overlap between your sense of self-efficacy and classroom practice?

Full research paper

THE PRODUCTIVE EFFICIENCY OF SCIENCE AND TECHNOLOGY WORLDWIDE: A FRONTIER ANALYSIS

Gustavo Ferro[1]✉
Carlos A. Romero[2]

[1]Universidad del CEMA (UCEMA) and CONICET, Argentina

[2]CONICET-Universidad de Buenos Aires. Instituto Interdisciplinario de Economía Política (IIEP-BAIRES), Argentina

✉ gferro05@yahoo.com.ar

ABSTRACT

We are interested in how codified knowledge is produced around the globe (which inputs are used to produce scientific articles and patented inventions) and the efficiency of the process (how do the best performers produce more with the same inputs or produce the same with less inputs). Using a Data Envelopment Analysis (DEA) efficiency frontier approach, we aim to determine which countries are more efficient at producing codified knowledge. We proxy knowledge production by publications and patents, obtained through human (researchers) and non-human (R&D expenditure) resources. We built a 15-year database with more than 800 observations of these and other variables. Our findings enable us to distinguish efficiency by country, geographical region, and income area. We run four different specifications and correlate the results with partial productivity indexes seeking consistency. Under constant returns to scale, the most traditional producers of knowledge are not fully efficient. Instead, small countries with limited resources appear to be efficient. When we add environmental conditions, both sets of countries are efficient producers of knowledge outputs. High-income regions, on the one hand, and East Asia, North America, and Europe and Central Asia, on the other, are the most efficient regions at producing knowledge.

KEYWORDS

Knowledge, science, productive efficiency; Data Envelopment Analysis, research

HOW TO CITE

Ferro G., Romero C. A. (2021) 'The Productive Efficiency of Science and Technology Worldwide: A Frontier Analysis', *Journal on Efficiency and Responsibility in Education and Science*, vol. 14, no. 4, pp. 217-230. http://dx.doi.org/10.7160/eriesj.2021.140402

Article history
Received
February 8, 2021
Received in revised form
May 17, 2021
Accepted
September 9, 2021
Available on-line
December 20, 2021

Highlights

- *We build and examine a proxy of a production frontier for national innovation systems using DEA methodology.*
- *We can detect with this method which countries achieve more output of its resources.*
- *Our concern is to concentrate on the relationship between outputs and inputs. The method permits to distinguish between countries which are the best achievers.*
- *Policy design could use this to help respond "where we are and where we are intended to be", and as a catalyst for change (deepening on why? and how?).*

INTRODUCTION

Public, private, and third sectors in every country devote resources through institutions (mainly universities, laboratories and research centers) to research activity. Produced knowledge can ultimately be applied to technology and yield developments. The knowledge production function is a multi-input and multi-output activity in which quality and environmental issues matter. The inputs consist of both human, non-human (scientific instruments, materials or financial resources), and intangible resources (accumulated knowledge, formal or informal networks of scientists and practitioners). The outputs can also be tangible (embedded in publications, patents, conference presentations, databases, etc.) and intangible (tacit knowledge, common practices, etc.). Important activity areas for economic growth include the creation of private and public knowledge, human capital building, and knowledge infrastructure production (Abramo, D'Angelo and Di Costa, 2015).

The concept of National Innovation Systems holds that innovation results from complex interactions between actors who generate, diffuse, and apply knowledge. This concept was applied in several contexts: national, technological, and sectoral. Innovation Systems can be understood as a set of relationships between private firms, public authorities, research organizations, and other bodies, ideally structured and co-ordinated in some way so that linkages between actors stimulate collective learning, continuous innovation, and entrepreneurial activity (Njøs and

Jakobsen, 2018). Thus, a country's innovative achievements (research and development, R&D) depend on how those actors link up with each other as components of a collective creation system. Their contribution can be divided into knowledge production and application (Choi and Zo, 2019).

For R&D policy makers this kinds of study are key elements to allocate resources, to establish priorities, to set goals, to evaluate past initiatives, to compare with similar countries with best achievements, to extract lessons, to change course, to avoid misleading objectives or instruments, and to trigger delving deeper in details about the "why" and "how" of the observed performance. They are useful to project possible future paths, to identify commonalities and differences. The type of assessment we offer is an instrument to evaluate and monitoring the impact of national policies. The impacts can be measure according its absolute performance (production, inputs, evolution across time, etc.) and with respect to its relative performance (productivity and efficiency, across time, places, and productive units). The first approach is relatively simple and useful; the second one is superior since relates results to resources, compare best practices with standard (or even substandard) ones, identify costly ways of achieving the results, and challenge the researcher to define, conceptualize, and measure the phenomenon under study. The cost of more complete and deep assessments is some loss of simplicity, since simple ratios are easier to understand than more complex studies of efficiency and productivity. However, certain frontier techniques, as those here presented offer a reasonable trade-off between deepness of the analysis and loss of simplicity. An efficiency assessment helps identify typologies of knowledge generation in different countries and provides policy and managerial implications for each case, as well as detecting best-practices to identify benchmarks and discover weaknesses (Ferro and D'Elia, 2020). Thus, it is possible to evaluate whether some policy or line of incentives to research had some impact, such as budgetary funds allocated to certain goals or rewards and disincentives to certain practices. Most previous studies that examined the efficiency of National Innovation Systems comprise two stages: knowledge production and knowledge application (Lundvall, 2007; Marxt and Brunner 2013). The former is defined as generating knowledge outcomes by using research-related inputs. The latter stage consists of transforming the outcomes of the previous process into inputs for economic results. In developing countries, an additional component is knowledge absorption coming from developed countries (Choi and Zo, 2019). We focus our study on knowledge production.

The differences in efficiency (achievements given resources) between countries lying on the frontier and countries lying below the frontier, also show the possibilities of improvement. To catch-up the best performers at macro level, the following step is to explain those differences delving into the details of each national innovation system (organizations, institutions, incentives, policies to identify and retain talent, tools to induce certain research lines, etcetera). This deserves a deeper analysis at a micro level which is beyond our goals, nevertheless measuring a phenomenon can be a good first step for diagnosis, which in turn can be useful to develop policy guidelines as a second step. To apply policies aimed to generate knowledge, planning is essential, and planning mean a diagnosis which begins in data collection, continues with transformation of data in information, follows with the analysis of information, and once some conclusions are drawn it is time for policy implementation, conducing to goals. Policy design could use this kind of analysis as a compass to help respond "where we are and where we are intended to be", and as a catalyst for change (if we are here, and we can be there, "why?" and "how?").

To guarantee the effectiveness of this instrument, it is necessary to analyze in detail certain features of the information and modelling. With respect to data, we build a homogeneous international database, paying attention to its coherence, quality, extension, comparability, and time span. And concerning models, pros and cons of different approaches are balanced. Simplicity is important for applying and for interpreting results within a interdisciplinary atmosphere such as national innovation systems. Data Envelopment Analysis (DEA), for instance, is particularly friendly to apply, and it is relatively straightforward to understand its results.

Based on data from R&D statistics on human and non-human resources as input variables, and publications and patents as output variables, we use an efficiency frontier method to determine what the countries obtain from them. It is not possible to draw a coherent conclusion from the patterns arising from the partial productivity indicators, such as papers per researcher or patents per unit of financial resources, since they yield, on occasion, contradictory rankings (Battese and Coelli, 1988). That is essentially why we have analyzed global indexes, such as frontier efficiency scores. Efficiency is understood as the relationship between inputs and outputs, paying due attention to environmental factors (contextual and mainly uncontrollable) and quality conditions (distinctive, and depending on volition and deliberate resource allocation decisions).

This study contributes to the literature by measuring the knowledge production of publications and patents, obtained from human and non-human resources at the country (national) level. This paper, builds and examines a proxy of a production frontier for national innovation systems using DEA methodology, contributing to a more complete evaluation of knowledge production than focusing either in comparing outputs, inputs, or simple ratios output/input (average productivity indexes). The method we apply allows us to determine with which combination of human and non-human resources (inputs), a certain set of outputs are produced (codified knowledge in the form of papers and patents over certain quality threshold). Moreover, we can detect with this method which countries achieve more output from its resources. Our concern is to concentrate on the observed relationship between outputs and inputs. The method permits to distinguish between units (countries in this case) which are the best achievers in comparison with other units which counting with the same resources achieve smaller results. Countries with the best results are in the efficiency frontier and countries with achievements below the frontier, by definition, produce only a fraction of the outputs the countries which lies in the frontier do. Quality and environmental issues should be addressed to differentiate the productivity of inputs and to compare the final outcomes. Quality results from deliberate actions to improve inputs and outputs, while environmental conditions are non-discretionary inputs.

The above discussion allows us to formulate our research questions: What is a meaningful proxy for the production function of a national R&D system? What are the inputs, the outputs, and environmental conditions when considering a frontier where the unit of analysis is a country? What are the levels of productivity measured through conventional indicators, such as average productivity or average costs of such a system? What levels of productive efficiency can be estimated for country systems? What are the drivers that explain the differences in efficiency between knowledge producing countries? How can this study be enriched in the future?

After motivating the discussion, defining the objective, setting the perimeter, establishing the possibilities and limitations of the study and formulating research questions in this introduction, The second section, presents materials and method: it starts with the database and its analysis, describes the method to explain the efficiency differences, their drivers, presents the models to be run and the empirical results. The third section, of discussion, evaluates studies relevant for the specification of outputs, inputs, quality and contextual variables, as well as further explains the utility of the empirical results for policies at the stage of diagnosis and performance monitoring, and the fourth section concludes.

MATERIALS AND METHOD

Material: the Database

Our database covers the years 2003-2017. The original database is an unbalanced panel for 206 countries and territories across 16 years, which was shortened by balancing to 60 countries across 15 years. The outputs considered are published papers included in the Scimago (2020) database and patents (applications and grants) from the World Intellectual Property Organization (WIPO) database.

A publication, according to Scimago (2020), is a document published and indexed in a specific year, which satisfies certain scientific protocols (blinded refereeing and indexing for a database which admits publications over a certain quality threshold). According to our source (WIPO, 2020), a patent is 'An exclusive right granted for an invention, which is a product or a process that provides, in general, a new way of doing something, or offers a new technical solution to a problem. To get a patent, technical information about the invention must be disclosed to the public in a patent application […] patent protection means that the invention cannot be commercially made, used, distributed, imported or sold by others without the patent owner's consent. Patents are territorial rights. In general, the exclusive rights are only applicable in the country or region in which a patent has been filed and granted, in accordance with the law of that country or region. The protection is granted for a limited period, generally 20 years from the filing date of the application.'.

Knowledge production uses human resources, physical productive capital, research funds, knowledge embedded in human resources, machinery and equipment, public involvement in R&D, and agglomeration effects. They are devoted to produce codified outputs (such as publications and patents), yielding non-codified outputs embedded in researchers, students and the community (extension or service activities). The inputs considered are non-human resources, proxied by R&D Expenditures (in constant 2005 Purchase Price Parity or PPP USD) and human resources proxied by Full-Time Equivalent (FTE) Researchers are both from UNESCO's compilation of statistics. In this study, each country is the decision-making unit of DEA, which uses human and non-human resources to produce scientific publications and patents. Quality is addressed through citations (for publications) and patent grants (for patents). The former to proxy the publication impact and the latter to proxy the patent's commercial application. According to WIPO (2020), 'Licensing a patent simply means that the patent owner grants permission to another individual/organization to make, use, sell etc. his/her patented invention. This takes place according to agreed terms and conditions, for a defined purpose, in a defined territory, and for an agreed period. Unlike selling or transferring a patent to another party, the licensor continues to have property rights over the patented invention.' Citation lag is not necessarily the same as publication lag, and patent grant lags are not necessarily the same as patent claim lags. Concerning publications, in an aggregated database, on the one hand, it is not possible to attribute each citation to its product. On the other hand, lags imply losing observations in a not very extensive series. Thus, we first correlate contemporaneous data of inputs to outputs and do the same for two-year lags. The correlations remain quite similar. Consequently, we opted to run the estimates in contemporaneous data for inputs and outputs.

Another decision concerns impact itself. Since we cannot trace citations to publications, and patent grants to patent claims, we run a Cobb-Douglas[1] cost function in logarithms with and without correction for quality. In the first version, we estimated costs with respect to outputs, the relative price of outputs and a time trend. Marginal costs of publications and patents proved to be highly significant and positive, as was expected. When we added two variables for quality consideration (citations for publications and patent grants for patents), the variable for citations was not significant and the inclusion of patent grants was highly significant, but invalidated the significance for patents. These results suggest that the marginal cost is positive for publications and for patents with commercial applications, but not necessarily so for citations (which depend on several factors) and for patent claims (not all patents have commercial potential). The Appendix presents the cost function estimates. Our outputs, corrected by quality (in the sense of a costly attribute added to the product to improve it) are, thus, publications and patent grants.

For environmental variables, we use data on GDP[2] as a proxy of the material resources each economy is able to produce, the population as a proxy of the potential dimension of the country's human resources, per capita GDP as the quotient of the former two, and the percentage of GDP[3] devoted to R&D

1 We also ran a Translog version to address quadratic and crossed effects, but such terms were nonsignificant.
2 We use GDP at constant prices without correcting for PPP values, which is a better proxy of the size of each economy than the PPP value.
3 In this case, we use PPP values since it is a cost.

expenditure as a proxy of the relevance of the activity in the country under study.
We built two groupings for the countries based on World Bank criteria to classify countries into high-, low-, lower-middle, and upper-middle income, on the one hand, and employed geography, on the other. We also distinguish those countries where English is the official language de jure or de facto. In each case, we studied correlations between these variables and outputs. We also present partial productivity ratios, which later compare with efficiency scores. Table 1 summarizes the variables: definition, source, units, and their role in the estimates.

Variables	Unit	Type	Concept
ydocum	Number of Documents published during a specific year	Output	Publications
qcitati	Total number of Citations contained in the documents published during the source year.	Quality	Quotes
ypateap	Number of Total Patent Applications (direct and Patent Cooperation Treaty PCT national phase entries). Yearly flow of new applications	Output	Patent applications
qpategr	Number of Total Patent Grants (direct and Patent Cooperation Treaty PCT national phase entries). Yearly flow of grants.	Quality	Patent grants
xfteres	Function-total Researchers Full Time Equivalent (FTE).	Input	Labor input
xgerdpp	Gross Domestic Expenditures in Research and Development (GERD) in Purchasing Power Parities (PPPs) at constant 2005 values.	Input	Non-labor input
GDPpc	GDP in 2010 constant million USD/Inhabitants	Environmental	GDP / Inhabitants
Partial productivity ratios			
doc_res	ydocum/xfteres (human resources partial productivity)		
doc_gerd	Ydocum/xgerdpp (non-human resources partial productivity)		
qpategr_res	Qpategr/xfteres (human resources partial productivity)		
qpategr_gerd	Qpategr/xgerdpp (non-human resources partial productivity)		

Table 1: Variable Definitions (sources: Scimago (2020) for publications, WIPO (2020) for patents, UNESCO (2020) for human and non-human resources and its elaborations, The World Bank (2020) for GDPpc and Population, World Bank Atlas for country classifications by income, 2019 GNI per capita, and geographical according WB. Wikipedia for "English as official language")

Table 2 presents the descriptive statistics of the database. Our final database is comprised of 803 annual observations of 60 countries over 15 years (2003-2017) with an average of 13.4 observations per country (minimum 10, maximum 15).

Variable	Observations	Mean	Total	Std Dev	Minimum	Maximum
ydocum	803	41,732	33,511,125	94,378	92	669,588
qcitati	803	725,181	582,320,089	1,970,961	780	18,095,159
ypateap	803	38,871	31,213,618	124,776	1	1,306,080
qpategr	803	16,914	13,581,586	51,013	1	352,567
xfteres	803	113,256	90,944,836	262,111	142	1,740,442
xgerdpp	803	20,623,366	16,560,562,919	60,325,656	3,419	426,030,391
GDP pc	803	25,310	20,323,615	22,466	456	111,968
doc_res	803	0.533	428.335	0.414	0.028	3.437
doc_gerd	803	0.005	4.228	0.006	0.000	0.075
pategr_res	803	0.081	65.424	0.105	0.000	0.777
pategr_gerd	803	0.001	0.490	0.001	0.000	0.008

Table 2: Descriptive statistics of the variables (source: own elaboration on sources of Table 1)

The correlation between our quality variables and their output is high and positive: 0.8510 (qcitati and ydocum) and 0.9306 (qpategr and ypateap). Also, the correlations between inputs and outputs are high and positive: 0.9015 (ydocum and xfteres), 0.9630 (ydocum and xgerdpp), 0.8637 (ypateap and xfteres) and 0.8782 (ypateap and xgerdpp). GDPpc has a 0.1797 correlation with ydocum and 0.1151 with ypateap.
English as an official de jure or de facto language in our sample exhibits a positive but low correlation with outputs: 0.3333 (d_eng and ydocum) and 0.1048 (d_eng and ypateap). As for impact, the English language is found to be more important: 0.3951 (d_eng and qcitati) but not so for patents: 0.1128 (d_eng and qpategr).
Thus, with respect to the correlation analysis and the correlation between geographical regions and outputs, two regions exhibit positive values: 0.5891 (d_nac and ydocum), 0.2900 (d_nac

and ypateap), 0.2250 (d_eas and ydocum), and 0.4951 (d_eas and ypateap). The last values also denote a relative pattern of specialization in publications in North America and patents in East Asia.

METHODOLOGY

The standard methods to estimate efficiency are the non-parametric Data Envelopment Analysis (DEA) and the parametric Stochastic Frontier Analysis (SFA). The respective advantages and disadvantages are well discussed in the literature. Parametric methods assume a specific functional form for the frontier, departing from some behavioral objectives (such as profit maximization or cost minimization); non-parametric methods do not given their greater flexibility to consider different decision-making unit behaviors. These methods can be deterministic or stochastic. Deterministic methods attribute the distance of a given decision-making unit from its frontier to inefficiency; stochastic methods assume that some of them can be attributed to randomness ("noise") and try to separate both components from the error term.

We use DEA to determine which decision-making units (in this case, countries) form an envelope surface of the sample to which they belong. The efficient decision-making units are those yielding on the frontier, while those below it are deemed inefficient since they produce less than their "peers" in the frontier with the same inputs (or produce the same with fewer inputs). A score is attributed to each decision-making unit based on how much it differs from the most efficient "peers". For each country, DEA solves an optimization problem seeking the optimal weights for the inputs and for the outputs, which maximize the ratio between the weighted sum of output divided by the weighted sum of inputs.

The efficiency measure (score) for any decision-making unit is obtained as the maximum ratio of weighted outputs to weighted inputs, subject to similar ratios for every decision-making unit being less or equal to unity. Following the Charnes, Cooper and Rhodes (1978) notation for n decision-making units ($j = 1,\ldots, n$), s outputs and m inputs the problem is:[4]

$$Max\,\theta = \frac{\sum_{r=1}^{s} u_r\, y_{r0}}{\sum_{i=1}^{m} v_i\, x_{i0}} \qquad (1)$$

Subject to:

$$\frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1;\; j = 1,\ldots,n$$

$$u_r, v_i \geq 0;$$

$$r = 1,\ldots,s;$$

$$i = 1,\ldots,m$$

Where θ is the maximum ratio for decision-making unit 0, y_r are the outputs (for $r = 1,\ldots, s$), x_i are the inputs (for i = 1,…, m), outputs and inputs being positive. The u_r, $v_i \geq 0$ are the weights yielded by the solution of the problem, which relies on all decision-making units used as a reference set.

The efficiency of one decision-making unit of the sample is to be rated relative to the others, distinguishing it by "*0*" in the functional (but preserving its original subscript in the constraints). This decision-making unit has the most favorable weighting allowed by the constraints (Charnes, Cooper and Rhodes, 1978). An optimal $\theta^* = \max\ \theta$ will always satisfy $0 \leq \theta^* \leq 1$ with optimal solution values u_r^*, $v_i^* > 0$ (Banker, Charnes and Cooper, 1984). Efficiency is defined as the score $E_r = y_r/Y_r$, where y_r is the actual output r produced by the decision-making unit under analysis, and Y_r is the maximum feasible output obtained by the same input set, where $0 \leq E_r \leq 1$. The weights are objectively determined to obtain a dimensionless E_r scalar measure of efficiency from observational data, subject only to the constraints established in (1). Therefore, no other set of common weights will give a more favorable rating relative to the reference set (Charnes, Cooper and Rhodes, 1978).

In the so-called CCR Model (Charnes, Cooper and Rhodes, 1978), the set of efficient decision-making units forms an envelope relative to observational data from all $j = 1,\ldots, n$ decision-making units. The envelopment can differ because of the scale assumption with respect to the phenomena under study. They are customarily constant returns to scale (CRS) or variable returns to scale (VRS), encompassing both increasing and decreasing returns to scale. CRS implies that output will change by the same proportion as inputs do (at the same scale), while VRS assumption reflects that production will change in different proportion as input do (differents scales, increasing, decreasing and as a special case, constant, that is, at the same scale). A priori can be reasons to assume certain returns to scale in each investigation, while the practice indicates that comparing results from different assumptions can be useful in another circumstances (Cooper, Seiford and Tone, 2007). Productivity and technical efficiency are equivalent only when the technology exhibits CRS and the model produces an "overall efficiency" rating. The BCC Model (Banker, Charnes and Cooper, 1984) applies to technologies with VRS, which helps compare the maximum average productivity attained at the most productive scale size with the average productivity at the actual scale of production to measure scale efficiency. Under VRS, it is possible to separate pure technical inefficiency from scale inefficiency. In this case, we only compare decision-making units of a similar scale. Units deemed inefficient under the CRS assumption can be efficient once we allow for VRS. As DEA is mainly a deterministic method, no accommodations have been made for bias resulting from environmental heterogeneity, external shocks, measurement errors, and omitted variables (Rhaiem, 2017).

4 This version of the problem is known as output-oriented. It can be also formulated as input-oriented, or not oriented at all. For the sake of brevity, we omit the last specifications. The way it is written is one of the three possibilities of presenting the problem. It intuitively presents the problem as maximizing the ratio between a weighted sum of outputs divided a weighted sum of inputs, where the key elements are the weights, different among efficient and non efficient decision-making units.

Models

We run two versions (CRS and VRS) for the two models, the first considering non environmental variables (CORE) and the second including the latter (ENV). We use GDPpc as a synthetic variable for development: GDP gives an indication of the economic size of the country, and the population proxies the country's potential in terms of human resources. GDPpc normalizes the first variable on the second. Moreover, GDP and population are partly exogenous (accounting for the endowments of resources) and partly endogenous (accounting for public policies, institutions and human capital accumulation). From the point of view of knowledge generation, we consider it a non-discretionary input and have treated it as such in the estimates. It is reasonable to assume it thus over a short period, such as that of the sample. It is also plausible that the accumulation of physical and human capital plus technology change transforms the variable into an endogenous factor in knowledge production in the long run. The models we present in Table 3 are the result of several alternatives that we have tested and the reasoning underlying each one is more easily understood when examining the results below.

	Outputs + Inputs (CORE CRS Model)	Outputs + Inputs (CORE VRS Model)	Outputs + Inputs + Environmentals (ENV CRS Model)	Outputs + Inputs + Environmentals (ENV VRS Model)
Outputs				
Ydocum	X	X	X	X
Qpategr	X	X	X	X
Inputs				
Xfteres	X	X	X	X
Xgerdpp	X	X	X	X
Environmental				
GDPpc			X	X

Table 3: DEA Model Specifications (source: own elaboration)

RESULTS

Table 4 presents the results of the sample as a whole by year, while Table 5 shows the figures aggregating countries by region (according to the geographical and income classifications of the World Bank). Table 4 shows the results for an unbalanced panel, which includes a minimum of 46 (2017) and a maximum of 59 countries (2011), and for four Models: CORE CRS, CORE VRS, ENV CRS, and ENV VRS.

Year	#	Core Model							Env Model						
		CRS			VRS			Scale Effic.	CRS			VRS			Scale Effic.
		# Eff	AVG	Std Dev	# Eff	AVG	Std Dev		# Eff	AVG	Std Dev	# Eff	AVG	Std Dev	
2003	49	5	0.564	0.269	15	0.753	0.249	1.337	16	0.750	0.238	19	0.804	0.211	1.072
2004	48	5	0.540	0.266	15	0.771	0.232	1.427	14	0.736	0.241	18	0.814	0.204	1.106
2005	50	4	0.472	0.251	13	0.738	0.223	1.562	14	0.713	0.252	18	0.794	0.203	1.115
2006	52	4	0.505	0.228	14	0.713	0.233	1.411	12	0.690	0.235	17	0.758	0.214	1.098
2007	56	5	0.499	0.233	17	0.716	0.243	1.434	15	0.685	0.242	20	0.755	0.225	1.102
2008	57	7	0.521	0.237	19	0.743	0.234	1.426	14	0.714	0.233	20	0.777	0.208	1.088
2009	58	6	0.477	0.242	16	0.693	0.254	1.455	13	0.662	0.246	18	0.728	0.229	1.101
2010	54	4	0.496	0.241	13	0.692	0.250	1.394	12	0.676	0.238	15	0.727	0.225	1.076
2011	59	4	0.487	0.235	17	0.707	0.259	1.450	12	0.689	0.246	18	0.742	0.239	1.078
2012	57	3	0.440	0.236	14	0.699	0.260	1.589	11	0.691	0.250	20	0.767	0.249	1.111
2013	56	4	0.402	0.233	11	0.660	0.275	1.642	10	0.653	0.258	17	0.745	0.260	1.141
2014	56	2	0.413	0.213	12	0.688	0.278	1.667	12	0.685	0.264	20	0.766	0.260	1.119
2015	53	2	0.389	0.183	13	0.715	0.263	1.838	14	0.716	0.257	22	0.788	0.246	1.100
2016	52	4	0.404	0.221	14	0.726	0.268	1.794	17	0.727	0.267	20	0.775	0.255	1.065
2017	46	3	0.400	0.222	14	0.726	0.270	1.814	11	0.725	0.247	17	0.778	0.245	1.073
AVG	**54**	**4**	**0.467**	**0.234**	**14**	**0.716**	**0.253**	**1.549**	**13**	**0.701**	**0.248**	**19**	**0.768**	**0.232**	**1.096**

Table 4: Core and ENV Models, by year (CRS and VRS versions)(source: own elaboration)

All models show a slight decrease in the average efficiency scores, considering the extreme points of the series. The CORE CRS model, as was expected, registers few efficient countries: 4 on average[5] (or 7.73 percent). The mean efficiency scores exhibit a decreasing pattern, averaging 0.467. The standard deviation, the dispersion and the minimum value show a quite stable behavior in the period. This model is influenced by the performance of a reduced set of small efficient countries. When we run the model omitting these countries to check for robustness, the average efficiency improves. Nonetheless, the objective is not to eliminate observations but to find a model that solves said aspect. Thus, we extend the models following two strategies: first, considering the potential influence of size (scale effect) in research and development activity (CORE VRS), and second considering the level of economic development as an environmental condition (ENV CRS and ENV VRS models).

The Core VRS Model exhibits 14 efficient countries on average[6] (or 27.10 percent). The mean efficiency scores average 0.716, starting above, and the standard deviation is almost the same as in the CRS version. By using a VRS model, we assume that scale is important. The estimate seems to indicate that. In effect, the scale efficiency resulting from the quotient of the average efficiency scores under VRS on the same average under CRS is surprisingly high (1.550 on average and has an increasing pattern, reaching more than 1.80 in some years). These results should be taken with caution: technology enables us to share a significant degree of knowledge and experiences internationally, and a great deal of intellectual production comes from transboundary collaboration. Unfortunately, our database does not allow us to address that issue. Up to this stage, the results suggest strong economies of scale. Nonetheless, in that regard, some environmental conditions, correlated with the size of the country, can explain those results.

When accounting for environmental conditions, the scale efficiency presents moderate values. What hypothesis would explain the absence of economies of scale in knowledge generation? In a globalized world, researchers and resources for R&D can collaborate and even move at a relatively low cost. Mobility is a great incentive to concentrate human and non-human resources in developed countries. Yet, there are some reasons to limit the international mobility of researchers, which implies a sort of "Home Bias"[7] in residence, research interests, and regional specialization. Differences in the cost of living, social status and growing communication possibilities are forces favoring low-scale knowledge production.

Thus, considering CRS ENV, the inclusion of GDPpc as an environmental variable yields results that are quantitatively similar to the Core VRS model. On average, 13 countries are efficient (24.72 percent of the sample), the mean efficiency score is 0.701, with similar dispersion and minimum to Core VRS model.

When analyzing VRS ENV, 19 out of 54 countries are efficient on average (or 34.88 percent), and standard deviation holds more or less the same average value as in the other three models. The minimum efficiency score is higher than in the rest of the models. However, the scale efficiency now shows a more stable and moderate pattern, averaging 1.096. We cannot confirm that variable returns to scale exist, moreover when considering aggregated information. The inclusion of a plausible environmental variable reduces the efficiency scores gap between CRS and VRS versions, in addition to the number of efficient countries. The difference in the average efficiency scores between both ENV model versions drops.

Table 5 shows that the production frontier's average efficiency scores yield higher efficiency levels in regions and countries with a tradition in academic production. Thus, East Asia (EAS), Europe and Central Asia (ECS) and North America (NAC) exhibit comparatively high levels of productive efficiency, while Latin America (LCN), the Middle East and North Africa (MEA) and Sub-Saharan African Countries (SSA) show comparatively low levels of productive efficiency, the first case being more consistent across the models. In contrast, the figures improve in the second case if GDPpc is included as an environmental variable. However, the performance of the models differs.

The CORE CRS model is the most pressing one because it does not consider the size of the scientific community or the development level of the countries. In this model, the performance of the countries with the most noteworthy academic tradition (such as the United States, the United Kingdom and Germany) does not behave efficiently and the scores of these countries are not very far from the average performance of the sample. China, the rising star in publishing and patents claims and licensing – because of its impressive growth in the years of the sample – behaves poorly in efficiency terms. The reason for these results can be attributed to the efficient behavior of small countries with comparatively scarce resources but comparatively abundant production. In turn, this admits several possible explanations. For instance, the international financing of research and development has a great impact in countries with a low R&D budget, a critical weight of small high-productivity research groups, data errors, among others.

Thus, one interpretation is that the CORE CRS Model yields a higher inclusion error (i.e., considering a country efficient when it is not) and exclusion (considering a country inefficient when it is not). The adoption of the hypothesis of VRS or the consideration of GDPpc as an environmental variable reduces the standard deviation of the most traditional R&D producers, which implies a lower exclusion error. Nevertheless, it does not correct the inclusion error. The CORE VRS, ENV CRS and ENV VRS models (whose average level of efficiency and the number of efficient countries is quite similar, but not coincident in each model's set of efficient countries, or in the same order)

5 The countries vary each year. Those which appear frequently are the Netherlands, Luxemburg, Japan, Republic of Korea.

6 The countries vary each year. Those which appear more frequently, besides those which appear as being efficiency under CRS are: United States, United Kingdom, Italy, Germany, China in the more recent years, as well of Chile, Russia, Poland, and Spain in some years.

7 It is a established hypothesis in finance. Portfolios in practice exhibit a lower internationalization in their composition than the levels suggested by theoretical portfolio models (Lewis, 1999).

show more foreseeable results for regions and countries with a greater tradition in R&D, particularly for these selected countries (see Table 5). The same is valid for China. Efficiency scores also increase monotonically with income level areas.

Region*/Selected country	Mean 2003-2017 CORE CRS	Mean 2003-2017 CORE VRS	Mean 2003-2017 ENV CRS	Mean 2003-2017 ENV VRS
EAS	0.554	0.896	0.959	0.970
ECS	0.494	0.852	0.845	0.868
LCN	0.235	0.435	0.501	0.533
MEA	0.534	0.628	0.775	0.861
NAC	0.441	0.980	0.910	0.915
SSF	0.418	0.659	0.751	0.803
High Income	0.536	0.934	0.919	0.935
Low Income	0.031	0.051	0.040	0.067
Lower Middle Income	0.276	0.437	0.627	0.753
Upper Middle Income	0.260	0.675	0.801	0.813
Sample Average	0.467	0.716	0.701	0.768
China	0.235	0.809	1.000	1.000
Germany	0.598	0.945	0.921	0.945
USA	0.448	1.000	1.000	1.000
UK	0.437	1.000	1.000	1.000

** Each country score within the region is weighted by its GDP for summing up.*

Table 5: Core and ENV models by region (CRS and VRS versions) (source: own Elaboration)

DISCUSSION

Conceptual discussion

This study contributes to the literature by measuring the knowledge production of publications and patents, obtained from human and non-human resources at the country (national) level. As stated, a national innovation system can produce knowledge, apply knowledge, and in certain context absorb knowledge coming from developed countries. Choi and Zo (2019) concentrates in the three stages and with cross sectional information estimates efficiency in each of the stages. Their contribution is important for its reach; however, they are considering only one year of information, and a subset of countries. Cross sectional databases have its merits however they do not allow to examine evolution over time. More mundane, outliers or error in data are hardly detected. We develop from Choi and Zo (2019). Our purpose is first to expand its time frame, second to extend geographically the reach (both things imply taking several instead of one picture, to detect evolution patterns), and third to focus the reach of the analysis to knowledge generation for deepening on this issue. In Choi and Zo (2019), papers and patents are the knowledge outputs. In our contribution we add quality dimensions (papers over certain internal quality assessment, and patents not only registered but licensed, that is, with current application potential) and environmental or contextual issues to address for differences in standard of living among countries. Also, we delve into the role of scale in knowledge production, that is, whether size of national innovation system is conductive (or not) to high efficiency.

As research activity is a production process, we analyze it from the perspective of the microeconomic theory of production. Performance should be evaluated with respect to aims and stated in measurable terms that represent the desired outcome (Abramo and D'Angelo, 2014). Aksnes et al. (2017) investigate methodological problems in measuring research productivity at the national level. Problems arise with the comparability of input and output statistics, as well as with the different National Innovation Systems themselves. Reports on resources and outcomes are often presented separately by different reporters, instead of combining them in measurements of productivity or efficiency.

Note that the productive units have different aims and institutional contexts. Fair comparisons of output-input efficiency must be managed by means of environmental variables, considering the uncontrollable inputs of knowledge production. The aims and rewards of private sector producers are supposedly different to public sector institutions' objectives and incentives. Nevertheless, public research organizations and universities face increasing demands to extend their teaching and research activities through the licensing of inventions, spin-off creation, research collaborations, and partnerships with private companies, etc. (Abramo, D'Angelo and Di Costa, 2015).

A time lag exists between the start of research and when the results ultimately appear as published articles. A standard in scientific production is that It can take about two years, on average, to achieve publication (Leydesdorff and Wagner, 2009). Lags also exist between the claim and award of patents, between the time patents are awarded and licensed with commercial aims, and from the time a publication is edited and cited. Nevertheless, the precise attribution of lags in empirical work is feasible if the data make it possible to attribute the citations to the publications, or the award to the patent, which is difficult when the level of aggregation of the data increases.

We base our selection of the outcomes that synthesize the countries' knowledge codified production on the character of public and private new knowledge. It is unlikely that the output of basic research is patentable since it is difficult or almost impossible to exclude third parties from using it (free access). Thus, private incentives to produce that knowledge are low, while public or third sector bodies interested in the production of public goods and positive externalities will be biased to produce them. This kind of knowledge can also surge and disseminate through accumulation and casual events (serendipity). New knowledge, a rival of established knowledge, and excludable, is deemed private or exclusive knowledge. Indeed, a patent confers exclusive rights on its owner. A publication is free knowledge (you pay for its supporting platform) (Elías, Ferro and García, 2019).

The results of research in the business sector are rarely published in scientific journals. If they are, it is often because the research was conducted together with actors from other sectors. Thus, we can expect publications to proxy well (mainly) publicly aimed and financed research, and that patents proxy well (mainly) privately aimed and financed research. Specific empirical evidence supports this assertion (Schmid and Fajebe, 2019). However, the boundary between knowledge outcomes is blurring. The recent worldwide moves towards incentivizing "impact" within the research funding system pose a growing challenge to academic research practices that produce both scientific and social impact, which differ in terms of their epistemic qualities and value as academic work. A growing number of countries are adopting incentives for research "relevance" to fund research. In this "culture of accountability", universities are becoming increasingly more "entrepreneurial" at raising funds (Bandola-Gill, 2019).

Rhaiem (2017) distinguishes the following outputs for knowledge production: refereed articles, books, book chapters, refereed conferences, professional publications, dissertations, other deliverables, and quality indicators to build a hierarchy among them. Refereed articles ensure sufficient homogeneity, which is a necessary condition to estimate a production function. To further homogenize outcomes, they can be corrected by quality, proxied by its impact, and measured by its citations. In addition to publications, several authors have included industry grants, third-party contracts, technology transfer to businesses, the number of patents, the number of expert reports, and teaching activities in the case of universities. It seems, thus, that publications and patents are good candidates for proxy knowledge production.

Research production is usually understood as the number of publications produced by a given unit and analyzed in such a way as to consider impact, efficiency, or quality components in research production to determine its drivers (Thelwall and Fairclough, 2017). Publications are external, objective means to measure outcomes, unlike internal and more subjective metrics. In effect, in assessing research group output, two approaches are generally employed: peer reviews and bibliometric methods. Each method has its pros and cons. However, publications appear to summarize the consensus of committees. The peer-review method is based on perceptions of well-informed experts about different quality dimensions of research production. It is subjective and depends on the committee's composition. Groot and García-Valderrama (2006) conclude that publications well proxy the scientific output as scientists understand it. A scientific paper has become the reference unit of bibliometric research because of the rules of the game of such outputs (blind refereeing, the transparency of the process, the need to present evidence and more recently, anti-plagiarism software), and the possibility of standardizing and auditing those rules through indexing in different databases (Abramo and D'Angelo, 2014). While the core bibliometric indicator is the number of published papers by an institution or country within a period, nevertheless, it should take the resources that are correlated with intellectual production into consideration (Bornmann et al., 2020). Partial productivity indexes allocates the papers to full-time equivalent researchers or non-human resources used in the productive process.

Bibliometric methods also have their limitations since they are restricted to written outputs. The evaluation results are influenced by the measurement methods applied, and the assessment of which publications are acceptable depends on publication and citation practices in different disciplines (Groot and García-Valderrama, 2006). Uncodified new knowledge is hard to measure, and when codified, some conventions should be included to identify and measure its various forms. A publication is an established measure of codified knowledge in most sciences, while it is not the standard output for the arts, humanities, and part of the social sciences. Compilations of science indicators measuring national research performance in the international context describe the development of a field of science or a production unit with the help of bibliometric means. Nevertheless, not all researchers produce bibliometric outputs, and these are heavily concentrated in a minority of researchers (Glänzel, 2003).

In specific environments, some scholars challenge publication in international journals (see Sahoo et al. (2015) for a discussion). They consider them biased against developing countries' problems, issues, data, and researchers. The prejudice extends to language since the international scientific community works almost exclusively in English, with little room for other native languages. The discussion concerns the legitimacy of using indexed international publishing as the main output of the scientific community. Nevertheless, the growing tendency to collaborate through international co-authorship is an important counterargument against the isolation argument ("a native writes for natives") (Sahoo et al., 2015).

The use of publications as an output measure of research has the disadvantage of being retrospective. Likewise, the use of research grants in efficiency studies has its proponents and opponents. Research grants are defined as additional resources in an institution's budget to promote research and the work of young scientists. Research grants proxy the market value of financed research and can also be considered a proxy of its quality. Nevertheless, the funds are spent not only on research assistance but also on other facilities, which are inputs for production. Thus, some authors consider research grants to be a measure of research output (because they are assigned to "virtuous" researchers or valuable proposals), but they mostly consider them an input because they support research projects (Gralka, Wohlrabe and Bornmann, 2019).

Given the diversity of research activities, several experts have proposed including indicators, such as patenting and spin-offs, in performance measurements (Gralka, Wohlrabe and Bornmann, 2019). Patents are often published to be presented to scientists (Abramo, D'Angelo and Di Costa, 2015). However, knowledge used for practical applications can differ from academic knowledge in terms of the type of problems being dealt with, incentives, timelines, accountability standards, procedures, and institutions (Bandola-Hill, 2019).

With respect to inputs of knowledge production, Rhaiem (2017) distinguishes human resources (academic staff and non-academic staff), physical productive capital (building spaces, laboratories, equipment, libraries, computers, etc.), research funds (capital and operating), knowledge embedded in human resources, machinery and equipment, public involvement in R&D, and agglomeration effects, which refer to a regional concentration of research effects and the links that could appear in an entrepreneurial environment. The identification of production factors other than labor, and the assessment of their value and share by field is often hard to quantify. Sometimes, unobserved effort could yield no results, and conversely, given that 'serendipity and luck may yield huge returns at little cost' (De Fraja and Valbonesi, 2012: 322). At the same time, some factors can be independent of the capacities of the staff for the units under examination, because of returns to scale, returns to scope or available capital resources (Abramo, D'Angelo and Di Costa, 2015).

Quality can and ideally should be assessed either in outputs or inputs for fair and meaningful comparisons through different coefficients or dummy variables (See Abramo, D'Angelo and Murgia, 2016, for seniority and qualification of researchers and its impact, for example). In bibliometric studies, research productivity distinguishes a researcher's publications from their impact (Abramo and D'Angelo, 2014). Although the databases differ in scope, the volume of data and coverage policies, the countries' outputs (papers) and impacts (citations) are extremely correlated (Archambault et al., 2009). Co-authorship, references, and citations are qualitative elements that denote the impact of the contribution. Processes are important when considering whether an output is deemed a scientific product: sources, procedures, and techniques should be reliable and documented and the reproducibility of results should be guaranteed. The screening procedures for publication ensures that scientific character in particular. Scientific knowledge production has historically developed within an international community of scholars for whom values such as objectivity and de-contextualization are epistemic virtues, and prerequisites for communication (Bandola-Gill, 2019).

There is a growing demand to expand the societal impacts of academia although the logic behind the achievement of scientific impact and societal relevance is different: aims (reputation or "scientific quality" against the practical application or "societal relevance") and processes diverge, as well as reward mechanisms. The reshaping of incentives (funding in particular) eventually reconciles the divergence of skills and interests. It also facilitates specialization among those that continue to produce knowledge (seeking peer recognition) and those who redirect their efforts towards communication and fundraising (D'Este et al., 2018).

The value of research is measured by its impact on scientific advancements. The impact is proxied by citations, which reveal knowledge dissemination. Not all given citations indicate quality. Heavy criticism can reflect their true impact. Irrelevance is a major reason not to be cited. Citation impact is mainly influenced by the subject matter, the paper's age, its "social status", the document type, and the observation period (Glänzel, 2003). Citation behavior differs across fields. At times, citation scores can be inflated, favoring popular authors, topics, fields, and established journals (Groot and García-Valderrama, 2006). An otherwise important paper that is casually dismissed as common knowledge may not be cited at all. Authors working on niche areas are cited less. Citation-based analyses can also be biased due to selective citations or self and mutual citations (Sahoo et al., 2015).

When measuring research productivity, as Abramo, Costa and D'Angelo (2015) recommends the specifications for the exercise must also include the publication period and the ''citation window'' (to address the already mentioned time lag of publications and the so called shelf life of papers, see Ferro and D'Elia (2020)). The "publication window", again according Abramo, Costa and D'Angelo (2015) refers both to the date of a paper's original submission to a journal to its date of acceptance, and then from acceptance to its actual publication. These vary greatly within the same discipline. Publication delays differ across fields. Publication intensity is linked to the type of research, as well as to the entire research life cycle: a scientist, say, could appear to be completely unproductive if evaluated during the launch of a new research program. The reliability of citations to approximate the publication impact is higher when the length of the "citation window" is longer (Abramo, Costa and D'Angelo, 2015).

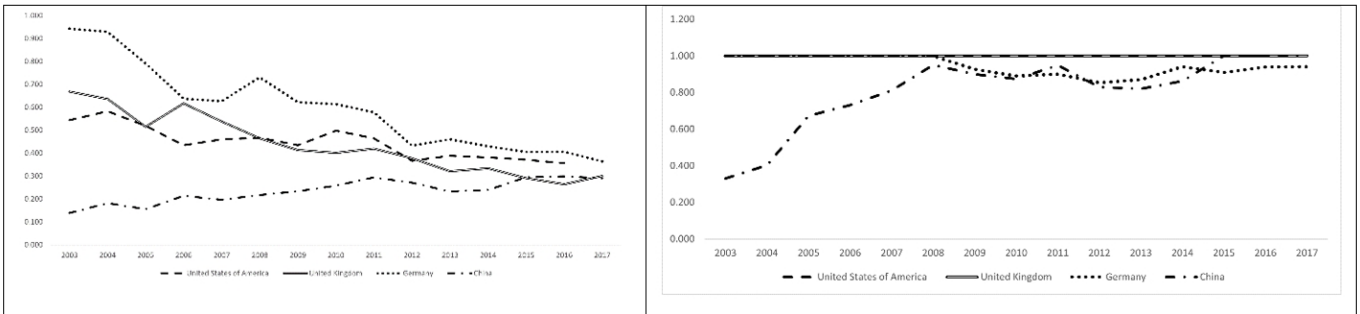
The most important indicators of co-authorship are the number and share of co-authored papers of a unit, joint publications of different units, the strength of co-authorship links, and the profile and citation impact of co-publications. Co-authorship weight can be assigned following fractional counting, first address count, or full or integer counting for each contributor. The first and the second are problematic when the practice of the discipline is simply alphabetical order. In some fields, instead, the tendency is to put the most relevant contributor first (Glänzel, 2003).

Many publications are internationally co-authored and result from collaborative efforts involving more than one country. While different principles and counting methods can be applied in bibliometric studies, the most common method is whole counting. Thus, every country receives full credit. An alternative is fractionalized counting, where the credit is divided proportionally between the participating countries. Whole counting reflects the number of papers in which the country

has "participated". The choice of counting method influences the output variable because the proportion of internationally co-authored publications varies across countries. According to Aksnes et al. (2017), small countries tend to obtain higher productivity results from whole counting than fractional counting.
The importance of patents refers to the subsequent technological change they induce, and it can be measured by citations received from new patents. Forward citations are positively correlated with the market value of a patent. The generality of patents concerns the technological scope of a technology's impact on subsequent innovation and is frequently measured by a concentration measure of a patent's forward citations. Patents can be issued for trivial or incremental inventions, while other patented innovations can yield subsequent technological progress for decades. During the application process, patent applicants and the patent examiner are required to cite antecedents for the proposed innovation. Experts consider highly cited patents important. Patents cited by patents from vast technology fields are deemed more general than those cited by patents with applications from few subfields. Patent quality is also expected to be positively correlated with the number of jurisdictions in which it has been filed (Schmid and Fajebe, 2019). An alternative measure of patents is less academic and more commercial, attributing impact on patents that reveal, through grants, economic impact.
Once inputs and outputs have been specified, it is important to identify possible environmental variables that may explain the differences in the efficiency scores of decision-making units. Environmental variables make it possible to address observable heterogeneity owing to uncontrollable factors. The main difference between environmental and production drivers is that the latter influence technology structure, while the former influence the efficiency with which the drivers are converted into outputs.

Empirical results discussion

One value added of the paper is to study the evolution of efficiency across the time: take a subset of best achievers in terms of gross production (USA, the UK, Germany and China), the evolution of efficiency scores across the time provides an impressive tendency to convergence in efficiency (clear in CRS CORE Model, and less clear but also present in VRS CORE Model) (See Figure 1). It is explained by the growth of China: in the period, this country increased its ydocum production by 553 percent, its qpategr in 2,842 percent, its full-time equivalent researchers in 102 percent and its R&D expenditures (at PPP values in 568 percent). In the same period, its GDP increased 226 percent. The ratio ydocum / researcher increased 223 percent in the period; and the ratio qpategr / researcher increased in 1357 percent. In 2003 the comparative productivity in terms of papers produced by researcher of China was slightly less than 20 percent the average researcher in the USA. In the last observation the value is almost 60 percent. The average researcher produced in China 1/13 patents with respect to the USA. In the last observation, the proportion converged. The comparison with USA, UK and Germany is presented in Table 6.



CRS Core Model Efficiency Scores	VRS Core Model Efficiency Scores

Figure 1: The evolution of efficiency in selected countries

GROWTH RATE		ydocum	qpategr	xfteres	xgerdpp	GDP pc
China	2017 / 2003	553%	2,842%	102%	568%	226%
Germany	2017 / 2003	76%	46%	54%	51%	23%
UK	2017 / 2003	81%	37%	34%	30%	12%
USA	2016 / 2003	55%	82%	21%	37%	14%
MEAN PRODUCTIVITY GROWTH		**Ydocum / xfteres**	**Ydocum / xgerdpp**	**Qpategr / xfteres**	**Qpategr / xgerdpp**	
China	2017 / 2003	223%	-2%	1,357%	340%	
Germany	2017 / 2003	15%	16%	-5%	-3%	
UK	2017 / 2003	35%	39%	3%	6%	
USA	2016 / 2003	28%	13%	50%	33%	

Table 6: The evolution of outputs, inputs, environmental variable and productivities (selected countries) (source: own elaboration)

The method we apply does not compare performance among incomparable units, instead it does compare "peers". Among "peers" (in our case, countries with similar resources) you can find best performers, middle performers, and low performers. If we say that (for example in CRS ENV Model) Chile is fully efficient and Japan is also fully efficient, we are not saying that Chile and Japan produce the same quantity or quality of research products, what we are saying is that Chile performs better than countries with similar resources (its peers) and the same is true for Japan. Chile and Japan are in different points of the frontier. On average for the whole period, Chile ydocum figure is 8,770 papers and qpategr is 258 patents, while Japan ydocum figure is 127,816 papers and qpategr is 262,009 patents. To focus on policy, one national authority can watch macrotendencies. The first question after seen these evidence is which micro determinants explain the results?

The scale consideration, in turn, tries to address whether is there an advantage of being big, which can have some rationale at laboratory level, and to certain extent, but it is not for sure at country levels. Finally, the environmental condition we test is another way to be fair in the comparison. Argentina and Brazil, for instance, are in a similar 0.19 low efficiency level in the CRS CORE model; if corrected by per capita GDP (the environmental condition), a bit lower in Brazil than in Argentina, Brazil goes to 0.55 and Argentina to 0.34. The interpretation is that Argentina wastes part of its slight per capita GDP advantage against Brazil, at least producing knowledge, all the other things equal. On the other hand, if compared the CRS CORE with the VRS CORE, Argentina's score is 0.33 and Brazil 0.40, the difference being attributable to the scale advantage due to the greater size of Brazil.

The correlation between partial productivity ratios and efficiency scores is comparatively higher in CRS models, and the highest corresponds to the CORE CRS model. On the other hand, the efficiency scores correlate 0.66 (both CRS) and 0.88 (both VRS), while CORE CRS and CORE VRS correlate 0.69, and ENV CRS and ENV VRS correlate 0.85. The low positive correlations between researchers' partial productivity ratios indicate some degree of complementarity between both outputs, while the low but negative productivity of financial resources devoted to both outputs suggests that they compete for the R&D funds.

CONCLUSION

Our study builds and examines a proxy of a production frontier for national science systems using DEA methodology. To that end, we concentrate on two outputs: publications and patents, and on two inputs – researchers (human resources) and R&D expenditure (non-human resources). We aggregate our measures to country levels as the unit of analysis.

Conventional indicators of our sample show that the average researcher in the sample produces 0.53 publications per year (3.43 maximum) and 0.17 patents (1.28 maximum). We go beyond partial productivity analysis and estimate production frontiers. We do not make behavioral assumptions about the mechanisms of national innovation systems. As the frontier is a non-parametric estimate, the orientation (to inputs or to outputs) is only a criterion to determine which variable is discretional. A non-oriented method is appropriate when both inputs and outputs can be modified discretionally. We estimate a "CORE" model following the economic theory (production requires human and non-human resources to produce outputs) and explore a comprehensive environmental variable adjusting that core model in an "ENVIRONMENTAL" (ENV) one. We do not conjecture a priori in favor of or against the presence of economies of scale. Thus, we estimate CRS and VRS versions of CORE and ENV models.

The results appear to be broadly consistent (when considering average values) with partial productivity indexes: efficiency and productivity have relatively high and positive correlations. The CORE CRS model reveals that the expected regions are the most efficient and that efficient areas are coincident with their affluence. Nevertheless, there are some surprises when analyzing individual countries. Small countries are positioned as efficient in the CORE CRS model while, contrary to what we expected, the most traditional producers of knowledge do not have outstanding efficiency results.

When VRS is incorporated, for instance, the United States, the United Kingdom become efficient, and Germany is quite close to being fully efficient. We follow another road, incorporating an encompassing environmental variable: GDPpc. When the ENV model is run under CRS, Germany, the USA, and the United Kingdom improve, the latter two becoming fully efficient. On the other hand, when GDPpc is incorporated as an environmental variable, they become efficient. The same is true for China, which is by no means efficient in the CORE model.

The role of variable returns to scale is unclear, especially when the level of aggregation is as high as it is in this analysis. The CRS assumption is only fitting when all of the decision-making units are operating at an optimal scale. Banker et al. (1984) suggest extending the CRS DEA model to account for variable returns to scale (VRS) situations. The use of the CRS specification when not all decision-making units are operating at the optimal scale will result in measures of total efficiency. One shortcoming of this measure of efficiency is that the value does not indicate whether the producer is operating in an area of increasing or decreasing returns to scale. This can be determined by running an additional DEA problem with a non-increasing return to scale (NIRS) assumption.

Is it meaningful to speak of constant or variable returns to scale in this context? It is probably useful, since a CORE CRS model yields a higher inclusion error (i.e., considering a country efficient when it is not) and exclusion (considering a country inefficient when it is not). The adoption of the hypothesis of VRS, or the consideration of GDPpc as an environmental variable, reduces the standard deviation of the most traditional R&D producers, which implies a lower exclusion error. However, it does not correct the inclusion error. More levels of disaggregation of the information would improve the results.

For R&D policy makers this kinds of study are key elements to allocate resources, to establish priorities, to set goals, to evaluate past initiatives, to compare with similar countries with best achievements, to extract lessons, to change course, to avoid misleading objectives or instruments, and to trigger delving deeper in details about the "why" of the observed performance. They are useful to project possible future paths, to identify commonalities and differences.

REFERENCES

Abramo, G. and D'Angelo, C. A. (2014) 'How do you define and measure research productivity?', *Scientometrics*, Vol. 101, No. 2, pp. 1129–1144. https://doi.org/10.1007/s11192-014-1269-8

Abramo, G., D'Angelo, C. A. and Di Costa, F. (2015) 'A New Approach to Measure the Scientific Strengths of Territories', *Journal of the Association for Information Science and Technology*, Vol. 66, No. 6, pp. 1167–1177. https://doi.org/10.1002/asi.23257

Abramo, G., D'Angelo, C. A. and Murgia, G. (2016) 'The combined effects of age and seniority on research performance of full professors', *Science and Public Policy*, Vol. 43, No. 3, pp. 301–319. https://doi.org/10.1093/scipol/scv037

Abramo, G., Costa, C. and D'Angelo, C. A. (2015) 'A multivariate stochastic model to assess research performance', *Scientometrics*, Vol. 102, pp. 1755–1772. https://doi.org/10.1007/s11192-014-1474-5

Aksnes, D. W., Sivertsen, G., van Leeuwen, T. N. and Wendt, K. K. (2017) 'Measuring the productivity of national R&D systems: Challenges in cross-national comparisons of R&D input and publication output indicators', *Science and Public Policy*, Vol. 44, No. 2, pp. 246–258. https://doi.org/10.1093/scipol/scw058

Archambault, É., Campbell, D., Gingras, Y., and Larivière, V. (2009) 'Comparing bibliometric statistics obtained from the Web of Science and Scopus', *Journal of the American Society for Information Science and Technology*, Vol. 60, No. 7, pp. 1320–1326. https://doi.org/10.1002/asi.21062

Bandola-Gill, J. (2019) 'Between relevance and excellence? Research impact agenda and the production of policy knowledge', *Science and Public Policy*, Vol. 46, No. 6, pp. 895–905. https://doi.org/10.1093/scipol/scz037

Banker, R., Charnes, A. and Cooper, W. W. (1984) 'Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis', *Management Science*, Vol. 30, No. 9, pp. 1078–1192. https://doi.org/10.1287/mnsc.30.9.1078

Battese, G. and Coelli, T. (1988) 'Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data', *Journal of Econometrics*, Vol. 38, No. 3, pp. 387–399. https://doi.org/10.1016/0304-4076(88)90053-X

Bornmann, L., Gralka, S., de Moya Anegón, F., and Wohlrabe, K. (2020) 'Efficiency of Universities and Research-Focused Institutions Worldwide: An Empirical DEA Investigation Based on Institutional Publication Numbers and Estimated Academic Staff Numbers', *CESifo Working Paper*, No. 8157.

Charnes, A., Cooper, W. W. and Rhodes, E. (1978) 'Measuring the Efficiency of Decision- Making Units', *European Journal of Operational Research*, Vol. 2, No. 6, pp. 429–444. https://doi.org/10.1016/0377-2217(78)90138-8

Choi, H. and Zo, H. (2019) 'Assessing the efficiency of national innovation systems in developing countries', *Science and Public Policy*, Vol. 46, No. 4, pp. 530–540. https://doi.org/10.1093/scipol/scz005

Cooper, W. W., Seiford, L. M. and Tone, K. (2007) *Data Envelopment Analysis. A Comprehensive Text with Models, Applications, References and DEA-Solver Software*, 2nd Edition, Boston, MA: Springer. https://doi.org/10.1007/978-0-387-45283-8

De Fraja, G. and Valbonesi P. (2012) 'The Design of the University System', *Journal of Public Economics*, Vol. 96, No. 3–4, pp. 317–330. https://doi.org/10.1016/j.jpubeco.2011.10.005

D'Este, P., Ramos-Vielba, I., Woolley, R. and Amara, N. (2018) 'How do researchers generate scientific and societal impacts? Toward an analytical and operational framework', *Science and Public Policy*, Vol. 45, No. 6, pp. 752–763. https://doi.org/10.1093/scipol/scy023

Elías, J., Ferro, G. and García, A. (2019) 'A Quest for Quality: Creativity and Innovation in the Wine Industry of Argentina', *AAWE Working Paper* 242, 09/2019.

Ferro, G. and D'Elia, V. (2020) 'Higher Education Efficiency Frontier Analysis: A Review of Variables to Consider', *Journal of Efficiency and Responsibility in Education and Science*, Vol. 13, No. 3, pp. 140–153. https://doi.org/10.7160/eriesj.2020.130304

Glänzel, W. (2003) *Bibliometrics As A Research Field. A Course on Theory and Application of Bibliometric Indicators*, Course Handouts.

Gralka, S., Wohlrabe, K. and Bornmann, L. (2019) 'How to measure research efficiency in higher education? Research grants vs. publication output', *Journal of Higher Education Policy and Management*, Vol. 41, No. 3, pp. 322–341. https://doi.org/10.1080/1360080x.2019.1588492

Groot, T. and García-Valderrama, T. (2006) 'Research Quality and Efficiency. An Analysis of Assessments and Management Issues in Dutch Economics and Business Research Programs', *Research Policy*, Vol. 35, No. 9, pp. 1362–1376. https://doi.org/10.1016/j.respol.2006.07.002

Leydesdorff, L. and Wagner, C. (2009) 'Macro-level indicators of the relations between research funding and research output', *Journal of Informetrics*, Vol. 3, No. 4, pp. 353–362. https://doi.org/10.1016/j.joi.2009.05.005

Lewis, K. K. (1999) 'Trying to Explain Home Bias in Equities and Consumption', *Journal of Economic Literature*, Vol. 37, No. 2, pp. 571–608. https://doi.org/10.1257/jel.37.2.571

Lundvall, B.-Å. (2007) 'National Innovation Systems—Analytical Concept and Development Tool', *Industry and Innovation*, Vol. 14, No. 1, pp. 95–119. https://doi.org/10.1080/13662710601130863

Marxt, C. and Brunner, C. (2013) 'Analyzing and Improving the National Innovation System of Highly Developed Countries—The Case of Switzerland', *Technological Forecasting and Social Change*, Vol. 80, No. 6, pp. 1035–1049. https://doi.org/10.1016/j.techfore.2012.07.008

Njøs, R. and Jakobsen, S.-E. (2018) 'Policy for Evolution of Regional Innovation Systems: The Role of Social Capital and Regional Particularities', *Science and Public Policy*, Vol. 45, No. 2, pp. 257–268. https://doi.org/10.1093/scipol/scx064

Rhaiem, M. (2017) 'Measurement and Determinants of Academic Research Efficiency: A Systematic Review of the Evidence', *Scientometrics*, Vol. 110, No. 2, pp. 581–615. https://doi.org/10.1007/s11192-016-2173-1

Sahoo, B., Singh, R., Mishra, B. and Sankaran, K. (2015) 'Research Productivity in Management Schools of India: A Directional Benefit-of-Doubt Model Analysis', *MPRA Paper*, No. 67046. https://mpra.ub.uni-muenchen.de/67046/

Schmid, J. and Fajebe, A. (2019) 'Variation in patent impact by organization type: An investigation of government, university, and corporate patents', *Science and Public Policy*, Vol. 46, No. 4, pp. 589–598. https://doi.org/10.1093/scipol/scz010

Scimago (2020) *Scimago Journal & Country Rank*, [Online], Available: https://www.scimagojr.com/countryrank.php [20 Mar 2020].

Thelwall, M. and Fairclough, R. (2017) 'The research production of nations and departments: A statistical model for the share of publications', *Journal of Informetrics*, Vol. 11, No. 4, pp. 1142–1157. https://doi.org/10.1016/j.joi.2017.10.001

UNESCO (2020) *UNESCO Institute for Statistics (UIS)*, [Online], Available: http://data.uis.unesco.org/ [20 Mar 2020].

WIPO (2020) *Information Resources on Patents*, [Online], Available: https://www.wipo.int/patents/en/ [20 Mar 2020].

World Bank (2020) *World Bank Open Data*, [Online], Available: https://data.worldbank.org/ [20 Mar 2020].

APPENDIX

A LOG-LOG COBB-DOUGLAS COST FUNCTION

The model estimates the financial resources devoted to R&D (in logarithms) as a function of the logs of both outputs (ydocum and ypateap), the relative price of publications in terms of patents (estimated as the quotient of zgerdpp on xfteres) and a time trend (model 1). After reaching a satisfactory core model, we try to address quality issues, approximating the logs of qcitati for publications and qpategr for patents. The first variable proves to be non-significant and the second is significant, but the log of ypateap loses its statistical significance (Model 2). We hold that the marginal cost of publishing is positive (in human and non-human resources), while it is also positive for licensing patents. Thus, we run a third model, using the logs of publications and licensing of patents and do not include the logs of qcitati and ypateap (Model 3). We use this as a guide to design the CORE and ENV DEA models in Section 3.

Dependent Variable: lxgerdpp / Independent variables (*t*):	Model 1	Model 2	Model 3
lYdocum	0.8080*** (25.29)	0.7916*** (16.04)	0.7945*** (26.61)
lYpateap	0.0749*** (4.03)	0.0230 (0.98)	
Zgerd_fteres	0.0026*** (4.03)	0.0026*** (10.23)	0.0026*** (10.27)
Trend	-0.2069*** (-6.62)	-0.0212*** (-4.24)	-0.0210*** (-6.89)
lQpategr		0.0682*** (3.62)	0.0807*** (5.28)
lQcitati		-0.0231 (-0.07)	
Constant	6.4566*** (31.84)	6.5837*** (29.20)	6.6190*** (31.86)
Observations	930	930	930
Groups	102	102	102
R-sq: within	0.3857	0.3935	0.3956
R-sq: between	0.9367	0.9366	0.9356
R-sq: overall	0.9387	0.9391	0.9381
Wald chi2(4)	1,832.9500	1,864.4800	1,783.7700
Prob > chi2	0.0000	0.0000	0.0000
Sigma_u	0.5596	0.5579	0.5800
Sigma_e	0.2633	0.2610	0.2621
Rho (fraction of variance due to u_i)	0.8187	0.8124	0.8304

**** 0.99 significant, ** 0.95 significance, * 0.90 significance*

Table A1: Cobb-Douglas in logs cost function. Random-effects GLS regression, unbalanced panel. (102 groups, minimum observations per group 1, maximum observations per group 15) (ource: own elaboration)

Full research paper

EXAMINATION OF SCIENCE SELF-REGULATION SKILLS OF GIFTED AND NON-GIFTED STUDENTS

Oğuzhan Nacaroğlu[1]✉
Oktay Bektaş[2]
Mustafa Tüysüz[3]

[1]Malatya Science and Art Center, Turkey

[2]Faculty of Education, Erciyes University, Kayseri, Turkey

[3]Van Yüzüncü Yıl University, Science Education Department, Van, Turkey

✉ onacaroglu44@gmail.com

ABSTRACT

The study aimed to examine and compare the science self-regulation skills of gifted and non-gifted students in this study. Survey design, one of the quantitative methods, was utilized in the research. The sample of the study consisted of 263 gifted students enrolled in science and art center and 482 non-gifted students located in the Eastern Anatolia Region of Turkey. Science Self-regulation Scale was used as a data collection tool in the research. Independent samples t-test and one-way analysis of variance were used in the analysis of the data. The findings showed that gifted and non-gifted students had high self-regulation skills towards science. In addition, it was found that although there was no statistically significant difference between the average scores of gifted female and male students on the overall scale, there was a significant difference in the other group. Moreover, while the difference between the mean scores obtained in the dimensions of Refinement, Time Organizing, Organizing, Help Seeking, Metacognitive Self-regulation, and Repetition was in favor of gifted students, it was in favor of non-gifted students regarding the mean scores of critical thinking and effort regulation dimensions. The conclusion and implication were discussed in line with these findings.

KEYWORDS

Science, self-regulation, gifted students, non-gifted students, survey

HOW TO CITE

Nacaroğlu O., Bektaş O., Tüysüz M. (2021) 'Examination of Science Self-Regulation Skills of Gifted and Non-Gifted Students', *Journal on Efficiency and Responsibility in Education and Science*, vol. 14, no. 4, pp. 230-246. http://dx.doi.org/10.7160/eriesj.2021.140403

Article history
Received
September 30, 2020
Received in revised form
May 3, 2021
Accepted
September 14, 2021
Available on-line
December 20, 2021

Highlights

- *Gifted and non-gifted students have high self-regulation skills towards science.*
- *There was no statistically significant difference between the average scores of gifted female and male students on the overall scale.*
- *A statistically significant difference was found between the mean scores of non-gifted female and male participants in the overall scale.*
- *It was determined that gifted students' self-regulation skills increase as the age gets older while the self-regulation skills of non-gifted students decrease.*
- *A statistically significant difference in the overall scale was found between gifted and non-gifted students' mean scores in favor of gifted students.*

INTRODUCTION

Today, rapid changes have been taking place in the field of industry and science. As the pioneer of this change, Industry 4.0 industrial revolution affecting individual and social areas have been shown (Vaidya, Ambad and Bhosle, 2018). With the last industrial revolution where machine power has replaced human resources, individuals have become crucial to make radical changes in education to have 21st-century skills and to produce and use knowledge (Puncreobutr, 2016). The Industrial 4.0 and production, innovation-oriented Education 4.0 education revolution have made it compulsory for students to develop appropriate learning strategies, learn to maintain their academic goals, and organize their actions towards this purpose (Ciolacu et al. 2017), thus enabling students to have self-regulation skills.

Self-regulation

The fact that self-regulation, which we encounter particularly in education and neuroscience (McClelland et al. 2018), is cognitively, affectively, and behaviorally versatile, revealing different views regarding its dimensions and definition. For example, Kopp (1982) emphasized the behavioral dimension of self-regulation as initiating and maintaining movement according to the situation while Posner and Rothbart (1998) stated the cognitive dimension of self-regulation by defining it as being able to control the desires and orientations of the individual and focusing on the target by maintaining attention. In general, self-regulation, which allows individuals to relate to their environment and control their behaviors by observing themselves, is defined as the processes that students use to activate their emotions and behaviors in line with their goals (Can Aran, 2015; Zimmerman and Kitsantas, 2014). In other words, self-regulation is the capacity of an individual to manage and change his/her ego's responses by controlling innate desires and tendencies (Bauer and Baumeister, 2011). The behavioral theory defines self-regulation as "learned self-control"; in the socio-cultural context, self-regulation is considered within the relationship between the individual and the environment. Since self-regulation includes different dimensions and definitions, it has led to the emergence of other self-regulated learning models.

Self-Regulated Learning Models

In self-regulatory learning within the scope of the process-oriented metacognitive model, an appropriate learning strategy is developed and implemented while considering the performance of the student in accordance with their goals (Borkowski, 1992). Another model is the four-stage self-regulated learning model. This model includes the stages of determining tasks and goals, implementing planning and strategies, and regulating metacognition (Winne and Hadwin, 1998). Another model is adaptive self-regulated learning. The most important stage of this model is the preparation stage in which the student evaluates (Boekaert, 1997). Pintrich's self-regulated learning model, based on social learning theory, examines self-regulation in terms of prediction, monitoring, control, and reflection (cited in Zat, 2018). According to Pintrich et al. (1991), the student should continuously improve himself/herself regarding content, behavior, cognition, and motivation. Zimmerman's approach is based on the Social Cognitive Learning Theory, which was the first to introduce a self-regulated learning model in the literature. This model includes prediction, performance, and self-reflection stages (Zimmerman, 1989).

Regardless of the model, the common point of self-regulation models is to enable students to take responsibility for their own learning, to gain the ability to control and evaluate their behavior in line with their goals (Dembo and Eaton, 2000). Hence, it is crucial for all individuals to have self-regulation skills to adapt to the rapidly changing new world order and control and evaluate their academic lives (Siegle, 2013). It is thought that individuals need to gain self-regulation skills to contribute to the countries gaining an advantageous position with other countries owing to increasing in knowledge-intensity with technological developments (Hilbert, 2014), the emerging of new business areas, and getting of the importance of lifelong learning skills (Chan and Rao 2010). In this context, it is necessary to examine the level of self-regulation skills of individuals in the present century and investigate the underlying reasons for this situation in depth if individuals with low self-regulation skills. For this reason, many countries have been working to identify and develop students' self-regulation skills at all levels of education. Self-regulated learning, which includes a dynamic and complex process by its nature (Muis, Chevrier and Singh, 2018), could be considered to be especially crucial for gifted students with higher level skills than their peers to recognize and improve their interests and abilities (Obergriesser, Steinbach and Stoeger, 2013; Tortop and Eker, 2014).

Gifted Students and Science Self-Regulation Skills

Gifted students are defined as individuals with high level abilities and competencies in one or more fields (Philips, 2019). Renzulli (1986) discussed the characteristics of gifted students under three headings creativity, being above average in general abilities, and task commitment. These individuals can exhibit many different behaviors in their social and educational lives. Gifted students focus on areas of interest, produce innovative solutions to problems, demonstrate a high level of self-awareness, transfer their knowledge to changing situations, and prefer challenging tasks (Cavilla, 2019; Kutlu Abu, 2018). Additionally, gifted students are shown as the most important human resource for economic and social welfare (Rinn and Bishop, 2015). Due to these characteristics of gifted students, it is necessary for different and challenging educational environments/teaching strategies such as acceleration, enrichment, and deepening (Renzulli and Reis, 2000). Because in the education of these students, suitable school environments are of great importance besides the education in the family (Lipovská and Fischer, 2016).

Heilbronner (2009) suggested that gifted students should be educated in accordance with their learning styles, interests, and needs. These students' curriculum should be organized in a flexible way that allows them to manage their own learning (Hockett, 2009). Otherwise, the teaching activities carried out within a particular program may not contribute to the developing of these students' skills, such as self-regulation, problem-solving, and creativity (Stoltz et al. 2015). In this context, the education of gifted students is carried out in Science and Art Centers (SAC). In SACs, different curriculums are applied to reveal the interests and abilities of gifted students and enable them to use them. In these educational centers, different curriculums have been implemented to demonstrate the interests and skills of gifted students, and allow them to use these skills. These programs are Adaptation Training (AT), Support Training (ST), Individual Talent Recognition (ITR), Special Skill Development (SSD), and Project Production and Management (PPM). It is important that gifted students studying in these centers have the outcomes of the determined courses. One of these courses is the science that attaches importance to individuals' science literacy and 21st-century skills.

The science course has general objectives, such as providing students with basic science-related skills, creating social awareness for the sustainability of resources, and understanding the stages of the formation of scientific knowledge and the nature of science (Ministry of National Education [MoNE], 2018). In line with this general purpose, it is important that gifted students in science courses should emphasize their scientific literacy, use their skills, and have self-regulatory learning skills by controlling their learning in the SAC. Because it is stated that an individual, who have self-regulation in the science course, is able to hold her/his cognitive processes by being aware of his/her own characteristics, has self-confidence in receiving new information, is also highly motivated in academic studies, and arranges the environments for working efficiently (Montalvo and Torres, 2004; Schunk, 2009). When the literature was examined, it was seen that the number of studies on the self-regulation skills of gifted students was limited. Kutlu Abu (2018) stated that differentiated science activities for the inclusion of gifted students positively improved their motivations and strategies for self-regulation. In addition, Kirişçi (2013) investigated the self-regulated learning and motivational beliefs of non-gifted and gifted students in mathematics. The findings obtained from this research indicated a difference in favor of female students in the use of extrinsic focus, exam anxiety, self-regulation, and cognitive strategy in non-gifted students and only in self-regulation for gifted students.

Purpose of the Study

Gifted students studying in the same learning environments with their peers face many problems such as managing time, not being able to evaluate their learning, not being able to set goals, and not being aware of their abilities (Stoeger and Ziegler, 2010). Appropriate content studies have been prepared in the SAC to solve these problems experienced by gifted students in having self-regulation skills that include a complex process. In this context, it is essential to examine and compare the gifted and non-gifted students' self-regulation skills in terms of creating appropriate self-regulated learning content for these students. Hence, in this study, it was aimed to investigate the self-regulation skills of gifted and non-gifted students regarding different variables and to compare the self-regulation skills of these students for science. Determining the skill levels of gifted and non-gifted students such as setting goal, developing and applying appropriate learning strategy, time management, effort regulation is thought to help researchers, teachers, experts, and families in terms of preparing suitable curriculum for these students. Because, with the results obtained, it is crucial to examine the students' level of the science self-regulation skills in terms of correct understanding and guidance of students. The research was assumed that gifted and non-gifted students respond sincerely to the items on the scale and that the sample represents the accessible population. One of study's limitations was that the study included non-gifted students and gifted students who continue their education in a SAC in a city in Anatolia. In this context, the following problems were sought:

1. What is the level of self-regulation skills of gifted students towards science?
2. Is there a statistically significant difference between gifted students' science self-regulation skill scores in terms of gender variables?
3. $H_0$1: There is no statistically significant difference between gifted students' science self-regulation skill scores in terms of gender variables.
4. Is there a statistically significant difference between gifted students' science self-regulation skill scores in terms of the program variable?
5. $H_0$2: There is no statistically significant difference between gifted students' science self-regulation skill scores in terms of the program variable.
6. Is there a statistically significant difference between gifted students' science self-regulation skill scores in terms of different age groups?
7. $H_0$3: There is no statistically significant difference between gifted students' science self-regulation skill scores in terms of different age groups.
8. What is the level of self-regulation skills of non-gifted students towards science?
9. Is there a statistically significant difference between non-gifted students' science self-regulation skill scores in terms of gender variables?
10. $H_0$4: There is no statistically significant difference between non-gifted students' science self-regulation skill scores in terms of gender variables.
11. Is there a statistically significant difference between non-gifted students' science self-regulation skill scores in terms of different age groups?
12. $H_0$5: There is no statistically significant difference between non-gifted students' science self-regulation skill scores in terms of different age groups.
13. Is there a statistically significant difference between gifted and non-gifted students' science self-regulation skill scores in terms of total and sub-dimensions of the scale?
14. $H_0$6: There is no statistically significant difference between gifted and non-gifted students' science self-regulation skill scores in terms of the scores of total and sub-dimensions of the scale.

MATERIALS AND METHODS

Research Model

In this study, it was aimed to seek and compare the self-regulation skills of gifted and non-gifted students towards science. For this purpose, survey research was utilized as a type of quantitative research method design. This design, which purposes to describe the past or present situation, event, opinion, and phenomenon (Büyüköztürk et al. 2010), is the research design carried out on relatively larger samples (Fraenkel and Wallen, 2006).

Population and Sample

The research sample consisted of 745 students, including 263 gifted students studying at a Science and Art Center and 482 non-gifted students continuing their education in public schools in the Eastern Anatolia Region of Turkey. Moreover, the sample was chosen from the accessible population by using the cluster random sampling. In cluster sampling, the accessible universe is divided into clusters and a sample is created by random selection from the created groups (Creswell and

Clark, 2016). In this context, the schools where the study was conducted were numbered and five schools were determined by lot to apply the scale to non-gifted students. Demographic information of the participants in the sample obtained by reaching at least 10% of the accessible population is given in Table 1. Demographic information of the participants in the sample obtained by reaching at least 10% of the accessible population is provided in Table 1.

Group	Variables	Demographic information	*f*	%
Gifted Students	Gender	Female	132	50.2
		Male	131	49.8
	Age	6-10	110	41.8
		11-15	120	45.6
		16 and over	33	12.6
	SAC Program	Support training	77	29.3
		Individual Talent Recognition	74	28.1
		Special Skill Development	67	25.5
		Project Production and Management	45	17.1
Non-gifted Students	Gender	Female	222	46.1
		Male	260	53.9
	Age	6-10	195	40.5
		11-15	190	39.4
		16 and over	97	20.1

Table 1: Demographic information for participants (source: own calculation)

As seen in Table 1, 50.2% of the gifted students were female while 49.8% were male participants. Likewise, 53.9% of the non-gifted students were male participants and 46.1% were female participants. Moreover, 77 gifted students continue their education in support education, 74 in individual talent recognition, 67 in special skill development, and 45 in project production and management programs. The ages of all participants ranged from 6 to 18 years.

Data Collection Tools

The self-regulation scale developed by Pintrich et al. (1991) for university students and called the "Motivated Strategies for Learning Questionnaire" (MSLQ) was utilized as the data collection tool in the study. The scale includes three parts and 81 items, and within the scope of the research, the study was conducted with 50 items of the "self-regulation strategies" section of the scale. While forming the scale, the views of two science educators who are experts in their fields were taken. In this context, the scale prepared in 7-point Likert-type was converted to 5-point Likert-type and arrangements were made in articles 4, 16, 21, 27, 29, 48, and 50 items. For example, Item 29, which was initially written as "If a subject is difficult, I will either stop working or study the easy parts. ", was changed to "If a science topic is difficult, I will either give up studying or only study the easy parts." in line with the opinions of experts. In this way, the students were enabled to express their views about the science course. Factor analysis was conducted to reveal the theoretical structure of the scale (Pallant, 2016). As a result of the analysis, the Kaiser-Meyer-Olkin [KMO] value was found as.89. According to the exploratory factor analysis outcome, it was decided to collect the scale under eight factors and the determined factors explain 49.25% of the total variance. Conformity index values were also calculated by conducting confirmatory factor analysis. (Root Mean Square Error of Approximation [RMSEA] = .045, Normed Fit Index [NFI] = .93, Comparative Fit Index [CFI] = .97, Incremental Fit Index [IFI] = .97, Relative Fit Index [RFI] = .93, Goodness of Fit Index [GFI] = .95, Adjusted Goodness of Fit Index [AGFI] = .66). It could be stated that the obtained indexes were within acceptable ranges and the scale forms a good fit with the determined fit indexes (Schumacker and Lomax, 2004). Sample items under each factor were indicated in Table 2.

As seen in Table 2, the items under each factor were examined and named. When the items in the first factor are concerned, this factor is called *"elaboration"* because it includes the individual's processes to make the information understandable (Salovaara, 2005). The second factor was named *"time regulation"*. The items under this factor are named in this way because they highlight the time dimension. For example, Item 49, written as "I don't find much time to review my notes or readings before the science exam." emphasizes time regulation. When the items in the third factor are checked, it is named *"organizing"* since it emphasizes schematization and connection (Wolters, Pintrich and Karabenick, 2003). For example, item 33, written as "While I am reading for a science course, I try to link what I read at the time with my previous knowledge.", is named this way because it involves the individual's organization of mental processes. When the items in the fourth factor were examined, this factor was called *"critical thinking"*. For example, the item 20 in the fourth factor, written as "I consider the topics covered in the science course as a starting point and try to form my own ideas on the relevant topics." draws attention to the individual's ability to present ideas by analyzing their situation. When the items in the fifth factor were checked, this factor was named *"effort regulation"*. Item 10 in the fifth factor, written as "If I am confused about something while reading about the science course, I go back to the beginning and try to understand.", indicates that the individual regulates his/her effort. When the items in the sixth factor were examined, this factor was named *"help seeking"*. For instance, the item 37 in this factor, written as "If I cannot understand a topic in a science course, I seek help from another student in the class.", emphasizes seeking help in learning the subject. When the items in the seventh factor were checked, this factor was named *"metacognitive self-regulation"*. The item 47 in this factor, written as "While I study science, I set goals for myself to direct my studies.", evaluates the process of planning and organizing students' learning. When the items in the eighth factor were analyzed, this factor was named *"repetition"*. The item 8 in this factor, written as "While I study science, I repeat important information inside of me over and over again.", includes situations such as underlining the subject and reading it repeatedly.

Factors	Factor names	Items in the factor.	Reliability coefficient	Sample item
Factor 1	Elaboration	3, 5, 11, 28, 32, 36, 38, 40, 41, 48	.887	While studying in the science class, I review what I have read and the notes I have taken and try to identify the most important points.
Factor 2	Time regulation	2, 6, 9, 21, 26, 29, 42, 46, 49	.892	I often do not have enough time for science because I am involved in other activities.
Factor 3	Organizing	18, 27, 30, 31, 33, 34	.889	When I am reading something for science, I try to link what I am reading with my previous knowledge.
Factor 4	Critical thinking	1, 4, 7, 12, 16, 20	.889	I see the topics covered in the science course as a starting point and try to come up with my own ideas on related topics.
Factor 5	Effort regulation	10, 22, 23, 35	.888	I constantly try to revise my ideas about what I have learned in the science class.
Factor 6	Help seeking	14, 19, 37, 44,	.886	If I do not understand a subject in science class, I ask for help from another student in the class.
Factor 7	Meta-cognitive self-regulation	13, 39, 43, 45, 47, 50	.893	When I study science, I set goals for myself to direct my studies.
Factor 8	Repetition	8, 15, 17, 24, 25	.885	When I study for science, I repeat important information many times.

Table 2: Determined factors for the scale (source: own calculation)

Data Analysis

For the statistical analysis of the research, the scale has eight dependent variables due to the sub-dimensions mentioned earlier. Moreover, there are independent variables such as gender, age, and types of curriculum. Therefore, it was considered to use the MANOVA to evaluate the collected data at the beginning of the analysis. However, since the assumptions of MANOVA could not be met, it was decided to utilize the independent sample *t-test for the comparison of the gender, and the one-way ANOVA for the comparison of the program and age as independent variables* although the type-1 error rate increased. These analysis methods, which reveal whether there is a significant difference between the averages compared, do not provide information about the size of the difference. Therefore, the effect size calculation was checked in the present study. The effect size findings revealed how much the independent variable explained the total variance in the dependent variable. While interpreting the effect size, if the calculated effect size value is between .2 and .5, the effect size is considered "small". Similarly, if the calculated effect size value is between .5 and .8, the effect size is considered "medium". Finally, if the calculated effect size value is between .8 and 1 the effect size is considered "large" (Cohen, 1988). Moreover, since the rate of type-1 error could increase in multiple analyses, Bonferroni correction was also made to reduce the effect of this situation. The formula (significance level (α) /number of group comparisons) was used for the Bonferroni correction (Vialatte and Cichocki, 2008). As a result, the significance level in the study was determined as $p < .006$. Within the scope of the research, descriptive statistical analysis was conducted by calculating frequency (f), percentage (%), mean, and standard deviation values to carry out inferential statistics for the scores obtained from the self-regulation scale. To determine whether the scores obtained from the scale differ according to gender, program types and age variable, it was first checked whether the scores were normally distributed for each variable. The determined values were presented in Table 3 (see next page).
When the total scores of the participants obtained from the scale were examined in terms of gender, SAC program types, and age, it was determined that the mean and median values were very close to each other, and the kurtosis and skewness values varied between -1 and +1. Thus, it could be said that the scores obtained from the scale indicated normal distribution regarding gender, SAC program types, and age (Tabachnick and Fidell, 2013). Finally, independent sample t-test for comparison the scores of the participants in terms of gender was chosen, and one-way analysis of variance (ANOVA) was used to compare the scores obtained from the scale regarding the SAC program types and age. "(The number of option-1)/number of options" formula was used to determine the level of participation of gifted and non-gifted students in items of dimensions in the self-regulation scale. Thus, the discontinuous answer choices become continuous and interpret the statistically obtained data facilitated by using this formula. In this context, the score ranges of the science self-regulation scale containing the five-point Likert type were given in Table 4.

Group	Variables		Mean	Median	Skewness	Kurtosis	Min	Max
Gifted Students	Gender	Male	188.4	187.0	.435	.025	153	243
		Female	188.9	188.5	.281	.175	135	236
	SAC Programs	ST	187.4	188	.487	-.062	160	235
		ITR	187.4	187.5	.219	-.327	153	226
		SSD	189.8	188	.002	-.100	135	235
		PPM	191.2	189	.559	.084	155	243
	Age	6-10	186.5	186.5	.405	-.123	153	235
		11-15	190.3	188	.218	.227	135	243
		16 and over	190.1	189	.420	.259	155	238
Non-gifted Students	Gender	Male	177.5	177	-.077	.112	108	237
		Female	184.7	186	.711	.359	88	231
	Age	6-10	177	175	.447	-.033	88	237
		11-15	184	186	-.462	.400	120	231
		16 and over	182.4	182	-.444	.032	127	227

Table 3: Descriptive statistics results of the scale scores (source: own calculation)

Science self-regulation scale	
Strongly Disagree	1.00–1.80
Disagree	1.81–2.60
Neutral	2.61–3.40
Agree	3.41–4.20
Strongly Agree	4.21–5.00

Table 4: The score ranges for scales (source: own calculation)

RESULTS

Science Self-Regulation Skills of Gifted Students

Within the scope of the research, the answer of the question *"What is the level of science self-regulation skills of gifted students towards science?"* was sought. In this context, descriptive statistics findings regarding the scores obtained by gifted students from the overall scale and each dimension were indicated in Table 5.

Self-regulation Scale	*N*	$\bar{X}$
Elaboration	263	3.83
Time Regulation	263	3.56
Organizing	263	3.90
Critical Thinking	263	3.76
Effort Regulation	263	3.75
Help Seeking	263	3.83
Meta-cognitive Self-regulation	263	3.91
Repetition	263	3.72
Total Score	263	3.78

Table 5: Descriptive Statistics on the Scale Scores of Gifted Students (source: own calculation)

When Table 5 was examined, it was seen that the total score average was found to be 3.78. Therefore, gifted students stated their opinions as *Agree* with 3.78 degrees of participation in the self-regulation scale. Likewise, the views of the gifted students regarding the sub-dimensions of the scale were examined, the students expressed their opinions as *Agree* in the sub-dimensions of *Elaboration* ($\bar{X}$ = 3.83), *Time regulation* ($\bar{X}$ = 3.56), *Organizing* ($\bar{X}$ = 3.37), *Critical thinking* ($\bar{X}$ = 3.76), *Effort regulation* ($\bar{X}$ = 3.75), *Help-seeking* ($\bar{X}$ = 3.83), *Meta-cognitive self-regulation* ($\bar{X}$ = 3.91), and *Repetition* ($\bar{X}$ = 3.72). Moreover, it was among the findings that gifted students agreed at least with the items in the *Time regulation* dimension and the most with the items in the *Meta-cognitive self-regulation* dimension.

Examination of Science Self-Regulation Skills of Gifted Students in Terms of Gender Variable

Within the scope of the research, the response of the question *"Is there a statistically significant difference between gifted students' science self-regulation skill scores in terms of gender variables?"* was sought. In this context, an independent sample *t*-test analysis was conducted to examine whether the total mean scores of male and female gifted students in the overall and sub-dimensions of the scale differ significantly. Analysis results were given in Table 6:

Dimension	Gender	*N*	$\bar{X}$	*sd*	*df*	*t*	*p*	Effect Size (η^2)
Elaboration	Male	131	3.83	5.24	261	-1.089	.277	.004
	Female	132	3.86	4.66				
Time Regulation	Male	131	3.60	4.03	261	1.419	.157	.007
	Female	132	3.51	4.55				
Organizing	Male	131	3.91	3.28	261	.219	.765	.001
	Female	132	3.89	3.48				
Critical Thinking	Male	131	3.80	3.32	261	1.110	.268	.004
	Female	132	3.72	3.06				
Effort Regulation	Male	131	3.73	2.81	261	-.265	.791	.002
	Female	132	3.76	2.75				
Help Seeking	Male	131	3.84	2.36	261	.408	.684	.006
	Female	132	3.81	2.27				
Meta-cognitive Self-regulation	Male	131	3.80	3.34	261	-2.779	.006	.028
	Female	132	3.96	2.87				
Repetition	Male	131	3.70	3.21	261	-.406	.685	.006
	Female	132	3.73	3.27				
Total Score	Male	131	3.76	2.33	261	-.246	.806	.002
	Female	132	3.77	2.26				

Table 6: Mean, Standard Deviation, *t* and *p* Values Regarding Total Scores of Male and Female Gifted Students Participants from Self-Regulation Scale (source: own calculation)

When Table 6 was evaluated, no statistically significant difference ($p > .006$, $t = -.246$) was found between the total mean scores of the female and male participants from the scale ($\bar{X}_{Female} = 3.77$, $\bar{X}_{Male} = 3.76$). Moreover, there was not any statistically significant difference between the scores obtained by male and female gifted students from each sub-dimension of the scale ($p > .006$). Therefore, $H_0 1$ hypothesis was not rejected in terms of all dimensions and total score.

Examination of Science Self-Regulation Skills of Gifted Students Regarding Program Variable

Within the scope of the research, the answer of the question *"Is there a statistically significant difference between gifted students' science self-regulation skill scores in terms of the program variable?"* was investigated. One-way analysis of variance (ANOVA), one of the parametric tests, was conducted to response this sub-problem. The group statistics obtained from the analysis was indicated in Table 7.

Program	*N*	$\bar{X}$
Support Training	77	3.74
ITR	74	3.75
SSD	67	3.79
Project Production and Management	45	3.82
Total	263	3.77

Table 7: Descriptive Statistics Results for the Program Variable (source: own calculation)

When Table 7 was analyzed, it was found that the mean of the gifted students enrolled in different SAC program types was close to each other. When the Levene's test results were examined, it was found that the assumption of homogeneity of variances was met for each dimension and scale in general ($p > .05$). The significance of the differences between the mean scores of the general and sub-dimensions of the scale was analyzed by using ANOVA. The results were given in Table 8. In Table 8, it was seen that No significant difference was found between the averages obtained from both the general [$F(3\text{-}259) = .614$; $p > .006$] and sub-dimensions of the scale in terms of SAC program types ($p > .006$). Therefore, $H_0 2$ hypothesis was not rejected in terms of all dimensions and total score.

Examination of Science Self-Regulation Skills of Gifted Students Regarding Age Variable

Within the aim of the research, the response of the question *"Is there a statistically significant difference between gifted students' science self-regulation skill scores in terms of different age groups?"* was investigated. One-way analysis of variance (ANOVA), one of the parametric tests, was utilized to respond to this sub-problem. The group statistics obtained from the analysis was shown in Table 9.

Dimensions	Source Variance	Sum of Squares	*df*	Mean of Squares	*F*	*p*	Effect Size (η^2)
Elaboration	Between Groups	66.041	3	22.014	.892	.446	.010
	Within Groups	6392.917	259	24.683			
	Total	6458.958	262				
Time Regulation	Between Groups	3.738	3	1.246	.066	.978	.007
	Within Groups	4868.954	259	18.799			
	Total	4872.692	262				
Organizing	Between Groups	56.500	3	18.833	1.662	.176	.010
	Within Groups	2935.150	259	11.333			
	Total	2991.650	262				
Critical Thinking	Between Groups	18.558	3	6.186	.601	.615	.006
	Within Groups	2665.092	259	10.290			
	Total	2683.650	262				
Effort Regulation	Between Groups	33.313	3	11.104	1.440	.231	.016
	Within Groups	1996.687	259	7.709			
	Total	2030.000	262				
Help Seeking	Between Groups	10.688	3	3.563	.662	.576	.007
	Within Groups	1392.841	259	5.378			
	Total	1403.529	262				
Meta-cognition Self-regulation	Between Groups	12.653	3	4.218	.420	.739	.004
	Within Groups	2603.560	259	10.052			
	Total	2616.213	262				
Repetition	Between Groups	13.836	3	4.612	.436	.728	.005
	Within Groups	2741.244	259	10.584			
	Total	2755.080	262				
Total	Between Groups	623.250	3	207.750	.614	.606	.007
	Within Groups	87585.207	259	338.167			
	Total	88208.456	262				

Table 8: Results of ANOVA Statistics (source: own calculation)

Age	*N*	$\bar{X}$
6-10	110	3.73
11-15	120	3.80
16 and over	33	3.96
Total	260	3.77

Table 9: Results of Descriptive Statistics Related to Age Variable (source: own calculation)

According to Table 9, it was found that the mean of gifted students in different age groups was close to each other. When the Levene's test results were checked, it was seen that the assumption of homogeneity of variances was met for each dimension and scale in general ($p > .05$). The significance of the differences between the scores obtained from the general and sub-dimensions of the scale was evaluated by utilizing ANOVA. The findings were depicted in Table 10.

When Table 10 was examined, there were not any a significant difference between the averages obtained from both the general [$F(2\text{-}260) = .264$; $p > .006$] and sub-dimensions of the scale in terms of age group ($p > .006$). Therefore, $H_0 3$ hypothesis was not rejected in terms of all dimensions and total score.

Science Self-Regulation Skills of Non-Gifted Students

Within the scope of the research, the answer of the question *"What is the level of self-regulation skills of non-gifted students towards science?"* was sought. In this context, the findings of the descriptive statistics regarding the scores obtained by non-gifted students from the general and sub-dimensions of the scale were given in Table 11.

When Table 11 was examined, the mean scale was also found to be 3.68. Therefore, non-gifted students stated their opinions as "Agree" with 3.68 degrees of participation in the whole scale. Likewise, the views of the non-gifted students regarding the sub-dimensions of the scale were investigated, the students express their views as "Agree" in the sub-dimensions of *Elaboration* ($\bar{X}$ = 3.8), *Organizing* ($\bar{X}$ = 3.70), *Critical thinking* ($\bar{X}$ = 3.91), *Effort regulation* ($\bar{X}$ = 3.98), *Help-seeking* ($\bar{X}$ = 3.55), *Meta-cognitive self-regulation* ($\bar{X}$ = 3.93), and *Repetition* ($\bar{X}$ = 3.89). Moreover, it was determined that non-gifted students agreed with *Effort regulation* dimension items the most and the items in the *Time regulation* dimension the least.

Dimensions	Source of Variance	Sum of Squares	*df*	Mean of Squares	*F*	*p*	Effect Size (η^2)
Elaboration	Between Groups	64.594	2	32.297	1.313	.271	.010
	Within Groups	6394.364	260	24.594			
	Total	6458.958	262				
Time Regulation	Between Groups	14.107	2	7.054	.377	.686	.002
	Within Groups	4858.585	260	18.687			
	Total	4872.692	262				
Organizing	Between Groups	8.102	2	4.051	.353	.703	.002
	Within Groups	2983.548	260	11.475			
	Total	2991.650	262				
Critical Thinking	Between Groups	16.198	2	8.099	.789	.455	.006
	Within Groups	2667.452	260	10.259			
	Total	2683.650	262				
Effort Regulation	Between Groups	23.782	2	11.891	1.541	.216	.011
	Within Groups	2006.218	260	7.716			
	Total	2030.000	262				
Help Seeking	Between Groups	3.816	2	1.908	.354	.702	.002
	Within Groups	1399.712	260	5.384			
	Total	1403.529	262				
Meta-cognition Self-regulation	Between Groups	10.542	2	5.271	.526	.592	.004
	Within Groups	2605.670	260	10.022			
	Total	2616.213	262				
Repetition	Between Groups	48.946	2	24.473	2.351	.097	.017
	Within Groups	2706.134	260	10.408			
	Total	2755.080	262				
Total Score	Between Groups	898.929	2	449.465	1.338	.264	.010
	Within Groups	87309.527	260	335.806			
	Total	88208.456	262				

Table 10: Results of ANOVA Statistics (source: own calculation)

Self-regulation Scale	*N*	$\bar{X}$
Elaboration	482	3.81
Time Regulation	482	2.72
Organizing	482	3.70
Critical Thinking	482	3.91
Effort Regulation	482	3.98
Help Seeking	482	3.55
Meta-cognitive Self-regulation	482	3.93
Repetition	482	3.89
Total Score	482	3.68

Table 11: Descriptive Statistics on Self-Regulation Scale (source: own calculation)

Examination of Science Self-Regulation Skills of Non-Gifted students in Regarding Gender Variable

Within the scope of research, the response of the question *"Is there a statistically significant difference between non-gifted students' science self-regulation skill scores in terms of gender variables?* was investigated. In this context, an independent sample t test analysis was utilized to examine whether male and female non-gifted students' total mean scores obtained from the scale and its sub-dimensions differ significantly. Analysis findings were shown in Table 12.

As seen in Table 12, there is a statistically significant difference (p < .006, t = -3.776) between the total scores of the female and male participants ($\bar{X}_{\text{Female}}$ participants = 3.69 and $\bar{X}_{\text{Male}}$ participants = 3.55). $H_0$4 hypothesis was rejected in terms of total score. However, since the calculated effect size (η^2 = .028) is smaller than the value between .2 and .5, the difference in self-regulation scores in favor of female students has not practical significance and not generalizable to the population. Moreover, there are a significant difference between the scores obtained by the male and female participants from the sub-dimensions of *Elaboration* [t(480) = -4.384; p < .006], *Time regulation* [t(480) = 2.860; p < .006], *Effort regulation* [t(480) = -4.131; p < .006], *Help seeking* [t(480) = -3.371; p < .006], *Meta-cognitive Self-regulation* [t(480) = -3.770; p < .006], and *Repetition* [t(480) = -3.802; p < .006]. Although, $H_0$4 hypothesis was rejected in terms of sub- dimensions, the difference in self-regulation scores in favor of female students in all sub-dimensions has not a practical significance and cannot be generalized to the accessible population since the calculated effect sizes for the sub-dimensions are smaller than the value between .2 and .5. Another result obtained from the analysis was that there was no significant difference between the scores obtained from the *Organization* [$t(480) = -1.961; p > .006$] and Critical Thinking [$t(480) = -2.248; p > .006$].

Dimensions	Gender	*N*	$\bar{X}$	*sd*	*df*	*t*	*p*	Effect Size (η^2)
Elaboration	Male	260	3.67	6.24	480	-4.384	< .001	.038
	Female	222	3.93	6.31				
Time Regulation	Male	260	2.77	6.18	480	2.860	.004	.016
	Female	222	2.58	6.87				
Organizing	Male	260	3.63	4.06	480	-1.961	.060	.007
	Female	222	3.76	4.31				
Critical Thinking	Male	260	3.84	3.90	480	-2.248	.065	.010
	Female	222	3.97	3.61				
Effort Regulation	Male	260	3.84	2.92	480	-4.131	< .001	.034
	Female	222	4.11	2.60				
Help Seeking	Male	260	3.43	3.34	480	-3.371	.001	.023
	Female	222	3.69	3.45				
Meta-cognitive Self-regulation	Male	260	3.81	3.76	480	-3.770	< .001	.028
	Female	222	4.03	3.84				
Repetition	Male	260	3.76	3.59	480	-3.802	< .001	.029
	Female	222	4.00	3.02				
Total Score	Male	260	3.55	2.62	480	-3.776	< .001	.028
	Female	222	3.69	2.63				

Table 12: Mean, Standard Deviation, *t* and *p* Values Regarding Total Scores of Male and Female Participants from Self-Regulation Scale (source: own calculation)

Examination of Science Self-Regulation Skills of Non-Gifted Students Regarding Age Variable

Within the scope of the research, the answer of the question *"Is there a statistically significant difference between non-gifted students' science self-regulation skill scores in terms of different age groups?"* was sought. One-way analysis of variance (ANOVA), one of the parametric tests, was used to analyze this sub-problem. The group statistics obtained from the analysis was indicated in Table 13.

Age	N	$\bar{X}$
6-10	195	3.54
11-15	190	3.68
16 and over	97	3.64
Total	482	3.61

Table 13: Results of Descriptive Statistics Related to Age Variable (source: own calculation)

When Table 13 was examined, it was seen that there was a difference between the mean scores of non-gifted students in different age groups from the overall scale. When the Levene's test results were evaluated, it was found that the assumption of homogeneity of variances was met for each dimension and scale in general ($p > .05$). The significance of the differences between the scores obtained from the general and sub-dimensions of the scale was analyzed by using ANOVA. The results were shown in Table 14.

As seen in Table 14, there was a significant difference between the total scores obtained from the scale in terms of age variable [$F(2\text{-}479) = 6.669; p < .006$]. Moreover, a statistically significant difference was found between the scores obtained for the sub-dimensions of *Elaboration* [$F(2\text{-}479) = 9.961; p < .006$], *Effort regulation* [$F(2\text{-}479) = 6.342; p < .006$], *Help seeking* [$F(2\text{-}479) = 5.725; p < .006$], and *Metacognitive Self-regulation* [$F(2\text{-}479) = 7.700; p < .006$]. Post Hoc test (LSD) was used to determine the source of the difference. In this context, it was determined that the mean scores of the 11-15 age group for the *elaboration, effort regulation,* and *Metacognitive Self-regulation* subscales were higher than the mean scores of the participants in the other age range. Another finding was that mean scores of the sub-dimension of *Help seeking* were determined to be higher in the 6-10 age range than the mean scores of the participants in the other age range.

$H_0 5$ hypothesis was rejected in terms of total score. However, since the calculated effect size ($\eta^2 = .023$) is smaller than the value between .2 and .5, the difference has not practical significance and not generalizable to the population. Although, $H_0 5$ hypothesis was rejected in terms of sub-dimensions, the difference in all sub-dimensions has not a practical significance, and cannot be generalized to the accessible population since the calculated effect sizes for the sub-dimensions are smaller than the value between .2 and .5.

Science Self-Regulation Skills of Gifted and Non-Gifted students

Within the scope of the research, the answer of the question *"Is there a statistically significant difference between gifted and non-gifted students' science self-regulation skill scores in* *terms of total and sub-dimensions of the scale?"* was sought. In this comparison, the level of self-regulation skills of the two groups was examined separately regarding each dimension. The results obtained were given in Table 15.

Dimensions	Source of Variance	Sum of Square	*df*	Mean of Squares	*F*	*P*	Effect Size (η^2)
Elaboration	Between Groups	785.396	2	392.698	9.961	< .001	.039
	Within Groups	18883.600	479	39.423			
	Total	19668.996	481				
Time Regulation	Between Groups	249.521	2	124.761	2.921	.055	.012
	Within Groups	20459.535	479	42.713			
	Total	20709.056	481				
Organizing	Between Groups	112.958	2	56.479	3.245	.040	.013
	Within Groups	8336.430	479	17.404			
	Total	8449.388	481				
Critical Thinking	Between Groups	10.341	2	5.170	0.359	.698	.001
	Within Groups	6890.323	479	14.385			
	Total	6900.664	481				
Effort Regulation	Between Groups	99.021	2	49.511	6.342	.002	.025
	Within Groups	3739.510	479	7.807			
	Total	3838.531	481				
Help Seeking	Between Groups	132.108	2	66.054	5.725	.003	.023
	Within Groups	5526.581	479	11.538			
	Total	5658.689	481				
Meta-cognitive Self-regulation	Between Groups	222.043	2	111.021	7.700	.001	.031
	Within Groups	6906.406	479	14.418			
	Total	7128.448	481				
Repetition	Between Groups	95.617	2	47.808	4.206	.015	.017
	Within Groups	5444.143	479	11.366			
	Total	5539.759	481				
Total	Between Groups	5088.261	2	2544.131	6.699	.004	.023
	Within Groups	213844.759	479	446.440			
	Total	218933.021	481				

Table 14: Results of ANOVA Statistics (source: own calculation)

Dimensions	Development Level	*N*	$\bar{X}$	*s*	*sd*	*t*	*P*	Effect Size (η^2)
Elaboration	Non-Gifted	482	3.79	6.39	657	0.972	.332	.001
	Gifted	263	3.83	4.96				
Time Regulation	Non-Gifted	482	2.68	6.56	717	19.601	< .001	.340
	Gifted	263	3.55	4.31				
Organizing	Non-Gifted	482	3.69	4.19	640	4.397	< .001	.025
	Gifted	263	3.90	3.37				
Critical Thinking	Non-Gifted	482	3.90	3.78	618	-3.134	.002	.013
	Gifted	263	3.76	3.20				
Effort Regulation	Non-Gifted	482	3.96	2.78	545	-4.081	< .001	.021
	Gifted	263	3.75	2.82				
Help Seeking	Non-Gifted	482	3.55	3.42	710	5.214	< .001	.035
	Gifted	263	3.83	2.31				
Meta-cognitive Self-regulation	Non-Gifted	482	3.89	3.84	632	-0.482	.630	.003
	Gifted	263	3.91	3.15				
Repetition	Non-Gifted	482	3.72	3.24	559	-3.090	.002	.012
	Gifted	263	3.87	3.39				
Total Score	Non-Gifted	482	3.61	2.66	610	5.248	< .001	.035
	Gifted	263	3.77	2.29				

Table 15: Independent sample t-test results of science self-regulation skills of gifted and non-gifted students (source: own calculation)

As seen in Table 15, a statistically significant difference was found in favor of gifted students (p < .006, t = 5.248) between the total scores of the gifted and non-gifted students ($\bar{X}_{gifted} = 3.77$, $\bar{X}_{non\text{-}gifted} = 3.61$). On the other hand, although there was not any a statistical difference between the mean of the *Elaboration* and *Meta-cognitive self-regulation* sub-dimensions of the gifted and non-gifted participants in the scale, a significant difference was also found between the scores obtained in the other sub-dimensions of the scale in terms of development level. This difference was in favor of gifted students in the dimensions of *Time regulation, Organizing, Help seeking, and Repetition* whilst it was in favor of non-gifted students in *Critical thinking* and *Effort regulation* dimensions. $H_0 6$ hypothesis was rejected in terms of total score. However, since the calculated effect size (η^2 = .035) is smaller than the value between .2 and .5, the difference has not practical significance and not generalizable to the population. Although, $H_0 6$ hypothesis was rejected in terms of the sub- dimensions, the difference in all sub-dimensions except that time regulation has not a practical significance, and cannot be generalized to the accessible population since the calculated effect sizes for the sub-dimensions are smaller than the value between .2 and .5. The significant difference in time regulation in favor of gifted students has practical significance and generalizable to the population since the calculated effect size for this sub-dimension (.340) are bigger than .2. Therefore, this result has a small effect on the population.

DISCUSSION

The study aimed to examine the self-regulation skills of gifted and non-gifted students towards science. In this context, 745 students, consisting of 263 gifted students continuing in different programs in SAC, and 482 non-gifted students attending public schools, participated in the study. Gifted students expressed their views as "Agree" with 3.78 participation degrees on the whole science self-regulation scale. Similarly, the non-gifted students stated their views as "Agree" with 3.68 degrees of participation. Hence, it could be interpreted that gifted and non-gifted students have high self-regulation skills towards science. Gifted students participated the most in the items in the *Meta-cognitive self-regulation* dimension and the least in the items in the *Time regulation* dimension. Non-gifted students, on the other hand, participated most in the items in the *Effort regulation* dimension and the least in the items in the *Time regulation* dimension like gifted students. This result was different from the results obtained from some investigations (Akpınar, Batdı and Dönder, 2013; Çekim and Aydın, 2018). For example, Çekim and Aydın (2018) in their research with non-gifted students stated that they mostly used the resource management and the Help seeking in the form of "determining who to get help from when they need help in the Science course". In this research, gifted students were more involved in the item *"I try to identify the concepts that I do not understand well while studying the science course"* in the *Metacognitive self-regulation* dimension. This finding is similar to the investigation by Tanti et al. (2020). Meta-cognitive self-regulation, which emphasizes awareness of strategies, resources, and skills to fulfill certain jobs and tasks (Noushad, 2008), involves the process of planning and organizing students' learning (Uzuntiryaki Kondakçı and Aydın Çapa, 2013). Therefore, it is crucial for gifted students to participate in the items in the Meta-cognitive self-regulation dimension in terms of determining their goals, planning their learning, evaluating, and organizing learning processes (Boekaerts and Niemivirta, 2005). Supporting this situation, Li et al. (2018) stated that students with metacognitive self-regulation set higher-level goals such as deep learning in science learning. In addition, more concrete data could be obtained by observing students' metacognitive self-regulation skills with a longitudinal study (Maloney, Ryan and Ryan, 2021). On the other hand, the reasons underlying the fact that both gifted and non-gifted students attend the items in the *Time regulation* dimension at least should be examined.

Gifted students, who constitute the crucial human resource of societies, should attach importance to time management in their learning process. Because self-regulated learning is a gradual and complex process with emotional, motivational and social components that involves students not only with the academic success achieved in school life, but also to produce realistic and lasting solutions to problems that can be experienced in every moment of life (Çokçalışkan, 2019; Oppong, Shore and Muis, 2019). Therefore, it is important to investigate the reasons and solutions for the low level of participation of gifted and non-gifted students in the time regulation dimension compared to the other dimensions. In this context, it is necessary to implement self-reflective learning practices in formal and informal learning environments (Patrick, 2017; Schunk and Zimmerman, 1998). Having self-reflection skills related to monitoring behaviour and feedback of learning evaluation allows students to manage their time by observing their own learning process and evaluating their learning (Grant, Franklin and Langford, 2002; Moeder-Chandler, 2020).

It was concluded that there was no statistically significant difference between the total mean scores of the science self-regulation scales of the gifted male and female students. This situation was expected in terms of gender equality. There were investigations in the literature that support this finding (Maloney, Ryan and Ryan, 2021; Gröpel, Baumeister and Beckmann, 2014). However, a statistically significant difference was found between the total mean scores of non-gifted male and female participants on the scale. A significant difference was also determined between the scores obtained from the sub-dimensions of *Elaboration, Time regulation, Effort regulation, Help seeking, Meta-cognitive self-regulation,* and *Repetition* in terms of gender variable. The results show differences from some studies (Alcı and Altun, 2007; Bouffard et al. 1995; Ilgaz, 2011; Lee, 2002; Lynch, 2010), and similarity with some studies (Betül-Cebesoy, 2013; Çalışkan and Sezgin Selçuk, 2010; Hargittai and Shafer, 2006; Tezel Şahin, 2015; Yükseltürk and Bulut, 2009). For example, Betül Cebesoy (2013) stated that the pre-service teachers' motivational and learning strategies did not change in terms of gender in the study that investigated the self-regulation skills of participants for physics lesson; Lee (2002), on the other hand, found that female students had more difficulties in self-regulated learning environments. In this respect, researchers should have more qualitative and mixed studies investigating the reasons for the level and differences of students' self-regulation skills in terms of gender.

Another finding was that there was not any significant difference between the total score averages of the gifted students obtained from the overall scale in terms of the SAC program types. Although there was no statistically significant difference, it was found that the average score of the gifted students in the project production and management program was higher than the mean of the students in the other programs. From this point of view, it might be interpreted that the gifted students who continue to the SAC have increased their level of self-regulation skills towards science as they move on to the next program. Particularly, the reason for the high self-regulation skills of the students in the project production and management program is high because of the creation of the learning environments necessary for the gifted students enrolling this program to participate in national and international project competitions and to organize their learning in this process by experienced advisors (Powers, 2008). It was also thought that doing science-based projects in these institutions could be effective (Chiang and Lee, 2016; Girgin, 2020). Another finding of the study was that although there is no significant difference between the total score averages of gifted students in different age groups obtained from the overall scale, it was found that the average score of 16 and over participants was higher than the average score of individuals in the other age group. Since the individuals in this age range are individuals in the project production and management program, this result supports the high self-regulation skills of the students in the project production and management program. Therefore, this situation shows that activities are carried out in SACs to enhance the self-regulation skills of the students, especially in the project production and management program. This result is different from non-gifted students.

It was concluded that the self-regulation skills of gifted students increased as the age progressed while non-gifted students decreased. A statistically significant difference was found between the scores of non-gifted students in general of the scale and the sub-dimensions of *Elaboration, Effort regulation, Help-seeking, and Meta-cognitive self-regulation* regarding the age variable. In addition, participants between the ages of 11–15; It was concluded that the mean scores of the *Elaboration and Effort regulation* sub-dimensions were higher than the average scores of the participants in the other age range. In the emergence of this situation, it might be said that suitable environments have been created for individuals in the age range with high self-regulation skills to organize and evaluate their own learning. Moreover, as the age gets older, exam-oriented studies might have caused to decrease their self-regulation skills. For example, in the current study, while the gifted students attending the SAC organized their learning by only doing scientific and artistic learning without exam anxiety, the non-gifted students did exam-oriented studies in their schools. This situation could be considered the reason for the decline in non-gifted students' self-regulation skills as they get older because it was stated that performance-oriented behavioral measures decrease self-regulation skills (Parlak Yılmaz, 2005). Furthermore, insufficient time and equipment of teachers to use learning strategies is shown among the factors that prevent students from developing self-regulation skills (Zumbrunn, Tadlock and Roberts, 2011). These factors prevent students from achieving their goals and cause them to fail to monitor and evaluate themselves (Baumeister, Heatherton and Tice, 1994). Considering the factors that hinder self-regulation skills, it explains the higher self-regulation skills of gifted students in science compared to non-gifted students. The findings obtained in this context support this conclusion.

This study is quantitative research, and more in-depth findings of students' self-regulation skills can be obtained by conducting different studies using qualitative research designs. Moreover, the self-regulation skills of non-gifted and gifted students might be improved with a student-centered learning method by conducting experimental studies.

CONCLUSION

It was concluded that gifted and non-gifted students had high science self-regulation skills. Moreover, gifted students mostly participated in the metacognitive self-regulation dimension and non-gifted students in the effort regulation dimension. On the other hand, students in both groups participated least in the time regulation dimension. Although there was no statistically significant difference between male and female gifted students' mean scores, there was a statistically significant difference between the mean scores of non-gifted female and male participants. Furthermore, it was seen that there was no significant difference between the mean scores of the gifted students in the different program types. On the other hand, researchers concluded that as gifted students get older, their self-regulation skills increase while non-gifted students' self-regulation skills decrease. A statistically significant difference was found in favor of gifted students in terms of the self-regulation skills. In addition, although there was no statistically significant difference between the mean scores of gifted and non-gifted students in terms of *elaboration* and *metacognitive self-regulation*, a significant difference was found between the mean scores in the other sub-dimensions of the scale. Whilst this difference was in favor of gifted students in the *Time regulation, Organizing, Help seeking, Meta-cognitive Self-regulation,* and *Repetition dimensions*, it favored non-gifted students in *Critical thinking and Effort regulation* dimensions.

REFERENCES

Akpınar, B., Batdı, V. and Dönder, A. (2013) 'Evaluating Primary School Students' Motivation Levels in Science Education in Terms of Gender and Class Variables', *Cumhuriyet International Journal of Education*, Vol. 2, No. 1, pp. 15–26.

Alcı, B. and Altun, S. (2007) 'Lise Öğrencilerinin Matematik Dersine Yönelik Özdüzenleme ve Bilişüstü Becerileri, Cinsiyete, Sınıfa ve Alanlara göre Farklılaşmakta mıdır? [Do High School Students' Self-regulation and Metacognitive Skills for Mathematics Lesson Differentiate According to Gender, Class and Fields?]', *Journal of the Cukurova University Institute of Social Sciences*, Vol. 16, No. 1, pp. 33–44.

Bauer, I. M. and Baumeister, R. F. (2011) 'Self-regulatory strength', in Vohs, K. D. and Baumeister, R. F. (ed.), *Handbook of self-regulation: research, theory, and applications*, pp. 64–82, New York: Guilford Press.

Baumeister, R. F., Heatherton, T. F. and Tice, D. M. (1994) *Losing Control: How and Why People Fail at Self-Regulation*, San Diego, California: Academic Press.

Betül Cebesoy, Ü. (2013) 'Pre-service Science Teachers' Perceptions of Self-Regulated Learning in Physics', *Turkish Journal of Education, January*, Vol. 2, No.1, pp. 4–18. https://doi.org/10.19128/turje.181050

Boekaerts, M. (1997) 'Self-regulated learning: A new concept embraced by researchers, policy makers, educator, teacher, and students', *Learning and Instruction*, Vol. 7, No. 2, pp. 161–186. https://doi.org/10.1016/S0959-4752(96)00015-1

Boekaerts, M. and Niemivirta, M. (2005) Self-Regulated Learning: Finding a Balance Between Learning Goals and Ego-Protective Goals, in Boekarters, M., Pintrich, P. R. and Zeidner, M. (ed.) *Handbook of Self Regulation*, San Diego: Academic Press. https://doi.org/10.1016/B978-012109890-2/50042-1

Borkowski, J. G. (1992) 'Metacognitive Theory: A Framework for Teaching Literacy, Writing and Math Skills', *Journal of Learning Disabilities*, Vol. 25, No. 4, pp. 253–257. https://doi.org/10.1177/002221949202500406

Bouffard, T., Boisvert, J., Vezeau, C., and Larouche, C. (1995) 'The Impact Of Goal-Orientation On Self-Regulation And Performance Among College Students', *British Journal of Educational Psychology*, Vol. 65, No. 3, pp. 317–329. https://doi.org/10.1111/j.2044-8279.1995.tb01152.x

Büyüköztürk, Ş., Çakmak, E. K., Akgün, Ö. E., Karadeniz, Ş. and Demirel, F. (2010) *Bilimsel araştırma yöntemleri [Scientific research methods]*, Ankara: Pegem Yayıncılık.

Can Aran, Ö. (2015) 'Özdüzenleme ve Çalışma Becerileri Arasındaki İlişki', *Pegem Eğitim ve Öğretim Dergisi*, Vol. 5, No. 2, pp. 207–220. https://doi.org/10.14527/pegegog.2015.011

Cavilla, D. (2019) 'Maximizing the Potential of Gifted Learners Through a Developmental Framework of Affective Curriculum', *Gifted Education International*, Vol. 35, No. 2, pp. 136–151. https://doi.org/10.1177/0261429418824875

Chan, C. K. and Rao, N. (2010) *Revisiting the Chinese learner: Changing contexts, changing education CERC studies in comparative education* (Vol. 25), Dordrecht: Springer.

Chiang, C. L. and Lee, H. (2016) 'The Effect of Project-Based Learning on Learning Motivation and Problem-Solving Ability of Vocational High School Students', *International Journal of Information and Education Technology*, Vol. 6, No. 9, pp. 709–712. https://doi.org/10.7763/IJIET.2016.V6.779

Cohen, J. (1988) *Statistical power analysis fort he behavioral sciences*, 2nd edition, Hillsdale, NJ: Erlbaum

Çalışkan, S. and Sezgin Selçuk, G. (2010) 'Self-Regulated Strategies Used by Undergraduate Students in Physics Problems: Effects of Gender and University', *Dokuz Eylül University The Journal of Buca Faculty of Education,* Vol. 27, pp. 50–62.

Ciolacu, M., Tehrani, A. F., Beer, R. and Popp, H. (2017) 'Education 4.0-Fostering Student Performance with Machine Learning Methods', IEEE 23nd International Symposium SIITME. https://doi.org/10.1109/SIITME.2017.8259941

Creswell, J. W. and Clark, V. L. P. (2016) *Designing And Conducting Mixed Methods Research*, New York: Sage.

Çekim, Z. and Aydın, S. (2018) 'Investigation of the Relations between Middle School Student's Motivation to Learn Science and Use of Learning Strategies', *Kırıkkale University Journal of Social Sciences*, Vol. 8, No. 2, pp. 451–468.

Çokçalışkan, H. (2019) *The effect of self-regulation based science teaching on primary school fourth-grade students' self-regulated learning skills, scientific processing skills and achievement*, [PhD Thesis], Marmara University, İstanbul.

Dembo, M. H. and Eaton, M. J. (2000) 'Self-regulation of academic learning in middle-level schools', *The Elementary School Journal*, Vol. 100, No. 5, pp. 473–490. https://doi.org/10.1086/499651

Fraenkel, J. R. and Wallen, N. E. (2006) *How to design and evaluate research in education,* New York: McGraw-Hill.

Girgin, D. (2020) 'Evaluation of project-based learning process of gifted children via reflective journals', *International Journal of Curriculum and Instruction*, Vol. 12, No. 2, pp. 772–796.

Grant, A. M., Franklin, J. and Langford, P. (2002) 'The self-reflection and insight scale: A new measure of private self-consciousness', *Social Behavior And Personality*, Vol. 30, No. 8, 821–836. https://doi.org/10.2224/sbp.2002.30.8.821

Gröpel, P., Baumeister, R. F. and Beckmann, J. (2014) 'Action versus state orientation and self-control performance after depletion', *Personality and Social Psychology Bulletin,* Vol. 40, No. 4, pp. 476–487. https://doi.org/10.1177/0146167213516636

Hargittai, E. and Shafer, S. (2006) 'Differences in Actual and Perceived Online Skills: The Role of Gender', *Social Science Quarterly*, Vol. 8, No. 2, pp. 432–448. https://doi.org/10.1111/j.1540-6237.2006.00389.x

Heilbronner, N. N. (2009) *Pathways in STEM: Factors Affecting the retention and attrition of talented men and women from the STEM pipeline*, [PhD Thesis], University of Connecticut, USA.

Hilbert, M. (2014) 'How Much of the Global Information and Communication Explosion is Driven by More, and How Much by Better Technology?', *Journal of the Association for Information Science and Technology*, Vol. 65, No. 4, pp. 856–861. https://doi.org/10.1002/asi.23031

Hockett, J. A. (2009) 'Curriculum For Highly Able Learners That Conforms To General Education And Education Quality İndicators', *Journal for the Education of the Gifted*, Vol. 32, No. 3, 394–440. https://doi.org/10.4219/jeg-2009-857

Ilgaz, G. (2011) *Investigation of elementary learners' perceptions on self-regulated learning strategies, self-efficacy and learner autonomy in science and technology course*, [PhD Thesis], Gazi University, Ankara.

Kirişçi, N. (2013) *Self-regulated learning and motivational beliefs of gifted and normal intelligence level students on mathematics*, [Ms Thesis], İstanbul University, İstanbul.

Kopp, C.B. (1982) 'Antecedents of Self-Regulation: A Developmental Perspective', *Developmental psychology*, Vol. 18, No. 2, pp. 199–214. https://doi.org/10.1037/0012-1649.18.2.199

Kutlu Abu, N. (2018) *An evaluation of differentiated science activities for inclusion of gifted students*, [PhD Thesis], Amasya University, Amasya.

Lee, I.S. (2002) 'Gender Differences in Self-Regulated On-Line Learning Strategies within Korea's University Context', *Educational Technology Research and Development*, Vol. 50, No. 1, pp. 101–109. https://doi.org/10.1007/BF02504967

Li, M., Zheng, C., Liang, J. C., Zhang, Y. and Tsai, C. C. (2018) 'Conceptions, self-regulation, and strategies of learning science among Chinese high school students', *International Journal of Science and Mathematics Education*, Vol. 16, No. 1, pp. 69–87. https://doi.org/10.1007/s10763-016-9766-2

Lipovská H. and Fischer J. (2016) 'Gifted Students and Human Capital Accumulation', *Journal on Efficiency and Responsibility in Education and Science*, Vol. 9, No. 3, pp. 60–69, https://doi.org/10.7160/eriesj.2016.090302

Lynch, D. J. (2010) 'Motivational Beliefs and Learning Strategies as Predictors of Academic Performance in College Physics', *College Student Journal*, Vol. 44, No. 4, pp. 920–928.

Maloney, D. M., Ryan, A. and Ryan, D. (2021) 'Developing Self-regulation skills in second level students engaged in threshold learning: results of a pilot study in Ireland', *Contemporary School Psychology*, Vol. 25, pp. 109–123. https://doi.org/10.1007/s40688-019-00254-z

McClelland, M., Geldhof, J., Morrison, F., Gestsdóttir, S., Cameron, C., Bowers, E., Duckworth, A., Little, T. and Grammer, J. (2018) 'Self-Regulation', in Halfon, N., Forrest, C. B., Lerner, R. M. and Faustman, E. M. (ed.) *Handbook of Life Course Health Development*, Springer International Publishing: Cham, Germany, pp. 275–298. https://doi.org/10.1007/978-3-319-47143-3_12

Ministry of National Education [MoNE], (2018) *Fen bilimleri dersi öğretim programı (İlkokul ve ortaokul 3, 4, 5, 6, 7 ve 8. Sınıflar)* [Science lesson curriculum (Primary and secondary school 3, 4, 5, 6, 7 and 8th Grades)]. Ankara: Millî Eğitim Bakanlığı Temel Eğitim Genel Müdürlüğü.

Moeder-Chandler, M. (2020) *Teacher reflective practice and perceptions on overall job satisfaction* [PhD Thesis], Trident University International.

Montalvo, F. T. and Torres, M. C. G. (2004) 'Self Regulated Learning, Current and Future Directions', *Electronic Journal of Research in Educational Psychology*, Vol. 2, No. 1, pp. 1–34.

Muis, K. R., Chevrier, M. and Singh, C. A. (2018) 'The Role of Epistemic Emotions in Personal Epistemology and Self-Regulated Learning', *Educational Psychologist*, Vol. 53, pp. 165–184. https://doi.org/10.1080/00461520.2017.1421465

Noushad, P.P. (2008) *Cognitions about cognitions: The theory of metacognition*, [Online Submission], Available: http://files.eric.ed.gov/fulltext/ED502151.pdf [30 Sept 2020].

Obergriesser, S., Steinbach, J. and Stoeger, H. (2013) 'Emotional experience during participation in a program of self-regulated learning', *Symposium at the 3rd International Conference on Talent Development and Excellence (ICTDE)*, Antalya, Turkey.

Oppong, E., Shore, B. M. and Muis, K. R. (2019) 'Clarifying the Connections Among Giftedness, Metacognition, Self-Regulation, and Self-Regulated Learning: Implications for Theory and Practice', *Gifted Child Quarterly*, Vol. 63, No. 2, pp. 102-119. https://doi.org/10.1177/0016986218814008

Pallant, J. (2016) *SPSS kullanma kılavuzu SPSS ile adım adım veri analizi* [SPSS survival manual a step by step guide to data analysis using SPSS], Ankara: Anı Publishing.

Parlak Yılmaz, N. (2005) 'School Goal Structure - Teacher Scale Adapting Into Turkish', *Educational Sciences and Practice*, Vol.4, No. 7, pp. 129–144.

Patrick, P. G. (2017) Informal Science Educators and the Nine Dimensions of Reflective Practice, in Patrick, P. G. (ed.) *Preparing Informal Science Educators* (pp. 41–65), Cham: Springer.

Philips, K. (2019) *Teaching the Student*, Salem Press Encyclopedia.

Pintrich, P. R., Smith, D. A. F., García, T. and McKeachie, W. J. (1991) *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*, Ann Arbor: University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning.

Posner, M. I. and Rothbart, M. K. (1998) 'Attention, Self–regulation and Consciousness', *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, Vol. 353, No. 1377, pp. 1915–1927. https://doi.org/10.1098/rstb.1998.0344

Powers, E. A. (2008) 'The use of independent study as a viable differentiation technique for gifted learners in the regular classroom', *Gifted Child Today*, Vol. 31, No. 3, pp. 57–65.

Puncreobutr, V. (2016) 'Education 4.0: New Challenge of Learning', *St. Theresa Journal of Humanities and Social Science*, Vol. 2, No. 2, pp. 92–97.

Renzulli, J. (1986) 'The three-ring conception of giftedness: A developmental model for creative productivity', in Sternberg R. J. and Davidson J. E. (ed.), *Conceptions of giftedness*, pp. 51–92. New York: Cambridge University Press.

Renzulli, J. S. and Reis, S. M. (2000) 'The Schoolwide Enrichment Model', in Heller, K. A., Mönks, F. J., Sternberg R. J. and Subotnik, R. F. (ed.), *International handbook of giftedness and talent*, pp 367–382, Oxford, UK: Elsevier Science Ltd.

Rinn, A. N. and Bishop, J. (2015) 'Gifted Adults: A Systematic Review and Analysis of the Literature.' *Gifted Child Quarterly*, Vol. 59, No. 4, pp. 213–235. http://dx.doi.org/10.1177/0016986215600795

Salovaara, H. (2005) *Achievement goals and cognitive learning strategies in dynamic contexts of learning*, [PhD Thesis], University of Oulu.

Schumacker, R. E. and Lomax R. G. (2004) *A beginner's guide to structural equation modeling*, New Jersey: Taylor & Francis; pp.1-8.

Schunk, D. H. (2009) *Öğrenme teorileri, eğitimsel bir bakışla [Learning theories an educational perspective]*, Ankara: Nobel Yayın Dağıtım.

Schunk, D. H. and Zimmerman, B. J. (ed.) (1998) *Self-regulated learning: From teaching to self-reflective practice*, New York: Guilford Press.

Siegle, D. (2013) *The underachieving gifted child: Recognizing, understanding, & reversing underachievement*, Waco, TX: Prufrock Press.

Stoeger, H. and Ziegler, A. (2010) 'Do Pupils with Differing Cognitive Abilities Benefit Similarly from a Self-regulated Learning Training Program?', *Gifted Educational International*, Vol. 26, No. 1, pp. 110–123. https://doi.org/10.1177/026142941002600113

Stoltz, T., Piske, F., de Freitas, M., D'Aroz, M. and Machado, J. (2015) 'Creativity in education: Contributions from Vygotsky and Piaget', *Creative Education*, Vol. 6, pp. 64–70. http://dx.doi.org/10.4236/ce.2015.61005

Tabachnick, B. G. and Fidell, L. S. (2013) *Using multivariate statistics*, 6th edition, Boston, MA: Pearson Education.

Tanti, M., Syefrinando, B., Daryanto, M. and Salma, H. (2020) 'Students' self-regulation and motivation in learning science', *International Journal of Evaluation and Research in Education*, Vol. 9, No. 4, pp. 865–873. https://doi.org/10.11591/ijere.v9i4.20657

Tezel Şahin, F. (2015) 'A Study into Self Regulation Sufficiencies of the Students Attending to the College Physical Education and Sport', *International Journal of Sport Culture and Science*, Vol. 3, No. 4, pp. 425–438. https://doi.org/10.14486/IJSCS406

Tortop, H. S. and Eker, C. (2014) 'Why Should Self-regulated Learning Skills Take Place in Gifted Education Programs?', *Journal of Gifted Education Research*, Vol. 2, No. 1, pp. 23–41.

Uzuntiryaki Kondakçı, E. and Çapa Aydın,Y. (2013) 'Predicting Critical Thinking Skills of University Students Through Meta Cognitive Self-Regulation Skills and Chemistry Self-Efficacy', *Educational Sciences: Theory Practice*, Vol. 13, No. 1, pp. 666–670.

Vaidya, S., Ambad, P. and Bhosle, S. (2018) 'Industry 4.0-Glimpse', Procedia Manufacturing, Vol. 20, pp. 233–238. https://doi.org/10.1016/j.promfg.2018.02.034

Vialatte, F. and Cichocki, A. (2008) 'Split-test Bonferroni correction for QEEG statistical maps', *Biological Cybernetics*, Vol. 98, No. 4, pp. 295–303. https://doi.org/10.1007/s00422-008-0210-8

Winne, P. H. and Hadwin, A. F. (1998) 'Studying As Self-Regulated Learning', in Hacker, D. J., Dunlosky J. and Graesser, A. C. (ed.) *Metacognition in educational theory and practice*, pp. 279–306, Hillsdale, NJ: Erlbaum.

Wolters, C. A., Pintrich, P. R. and Karabenick, S. A. (2003) Assessing Academic Self-Regulated Learning. *Conference on Indicators of Positive Development: Definitions, Measures, and Prospective Validity.*

Yükseltürk, E. and Bulut, S. (2009) 'Gender Differences in Self-Regulated Online Learning Environment', *Educational Technology & Society*, Vol. 12, No. 3, pp. 12–22.

Zat, Z. (2018) *Investigation the effectiveness of EMDR focused intervention program for increasing self regulated learning on self regulation and trauma symptoms of university students with traumatic experiences related to academic life*, [Ms Thesis], Eskişehir Osmangazi University, Eskişehir.

Zimmerman, B.J. (1989) 'A Social Cognitive View of Self-regulated Academic Learning', *Journal of Educational Psychology*, Vol. 81, No. 3, pp. 329–339. https://doi.org/10.1037/0022-0663.81.3.329

Zimmerman, B.J. and Kitsantas, A. (2014) 'Comparing Students' Selfdiscipline and Self-regulation Measures and Their Prediction of Academic Achievement', *Contemporary Educational Psychology*, Vol. 39, pp. 145–155. https://doi.org/10.1016/j.cedpsych.2014.03.004

Zumbrunn, S., Tadlock, J. and Roberts, E. D. (2011) *Encouraging self-regulated learning in the classroom: review of literature*, Metropolitan Educational Research Consortium (MERC).

Full research paper

USE OF SCREENCAST IN DISTANCE EDUCATION GIS LESSONS: STUDENTS' VIEWS

Salih Yıldırım

Marmara University, Turkey

salih.yildirim@marmara.edu.tr

ABSTRACT

This study aims to analyze the views of geography teacher students on the GIS course conducted with screencasts during the distance education process. Furthermore, the study reveals, along with the reasons, how the process of capturing and sharing screencasts with students can assist instructors. At the end of the term, 27 students studying in the second year of the geography teaching undergraduate program were asked about their opinions of the lesson. The data were collected through a structured evaluation form consisting of open-ended questions to evaluate the course. In the analysis of the comments obtained, MAXQDA software, which is frequently used in qualitative data analysis, was applied. The expressions in the answers were analyzed using the content analysis technique. Overall, the results prove that the beneficial aspects of the GIS course, which consists of applications and is conducted with screencast, outweigh the theory. Thanks to the screencast model, students can use their time efficiently and watch the videos repeatedly whenever and wherever they want. Nevertheless, the results reveal that extra methods are required to motivate students to transfer theoretical information and that screencasts have some disadvantages.

KEYWORDS

Distance education, GIS teaching, screencasts, user research

HOW TO CITE

Yıldırım S. (2021) 'Use of Screencast in Distance Education GIS Lessons: Students' Views', *Journal on Efficiency and Responsibility in Education and Science*, vol. 14, no. 4, pp. 247-257. http://dx.doi.org/10.7160/eriesj.2021.140404

Article history
Received
May 30, 2021
Received in revised form
July 26, 2021
Accepted
October 27, 2021
Available on-line
December 20, 2021

Highlights

- *Each student's ability to work according to their own learning pace or readiness is a critical advantage.*
- *Via screencasts, students can use their time efficiently and watch the videos repeatedly whenever and wherever they want.*
- *It is necessary to stay away from presentations containing only information, increase the visuals and keep the video short.*
- *The most crucial disadvantage is that students continue the learning process independently of the educator and their friends.*

INTRODUCTION AND LITERATURE

The outbreak of the COVID-19 pandemic in various countries at the end of 2019 has transferred traditional face-to-face learning to distance education platforms (Burgess and Sievertsen, 2020; Chen et al., 2020). Social distancing rule and school closing bring about change and reform in the education system of many countries (Bao, 2020). During the COVID-19 distance education process, screencasts are one of the favorite tools preferred in lessons. Furthermore, information and communication technologies (ICT) development encourages universities to experience alternative models such as online or distance education (Albăstroiu, Felea and Vasiliu, 2014; Kauppi et al., 2020). ICT plays a vital role in both the teacher's teaching process and the student's learning process. Teachers benefit from modern technologies in teaching the lesson and students in learning the lesson (Mitchell et al., 2018; Paraskeva, Bouta and Papagianni, 2008). Although certain problems arise in the process of adapting to modern technologies, university students adapt quickly to change (Tartavulea et al., 2020). Distance education offers students important opportunities to overcome space-time limitations, provide learning time flexibility and emphasize different learning styles (Onete et al., 2014).

In recent years, many instructors have been using screencasts in their lessons (Gormely and McDermott, 2011; Luongo, 2015). While screencast is an integral part of today's education system, it also plays a vital role in higher education. For instance, Stanford and Berkeley universities use screencasts

Electronic ISSN **1803-1617** Printed ISSN **2336-2375**

as an educational tool. Instructors may include screencast in distance learning, sometimes in all of their lessons, and sometimes as part of their blended learning (Brame, 2016). However, screencast software continues to become an essential part of our daily life. Although it has widespread use in many fields, videos created by screencast are frequently used, mainly in learning different computer-based software. Screencasts created and broadcast by other users have an essential place in solving the problems encountered during the use of various software.

Screencasts can also appear in the literature with the concepts of screencast and screen capture. Screencast means recording a computer screen simultaneously with sound, mouse movements, and clicks (Kawaf, 2019; Winterbottom, 2007; Snyder, Paska and Besozzi, 2014). Individuals with computer skills can use screencast software without any difficulty. In fact, the software can be downloaded for free, and its application requires very little technical knowledge (Oud, 2009; Ruffini, 2021). The screencast software was expensive in the early years of computers. On the other hand, it can be used free of charge with the development of open-source software in recent years. CamStudio and OBS (Open Broadcaster Software) studio are among the most used open-source software (Catalano, 2014).

Numerous studies in the literature prove that using a screencast in lessons has many advantages for teachers and students, regardless of whether they are inside/outside the classroom (Gormely and McDermott, 2011; Luongo, 2015; Peterson, 2007; Gambari and Hassan, 2017; Hasanudin and Fitrianingsih, 2018). While (Gormely and McDermott, 2011) found screencasts useful and exciting in an online education environment, (Peterson, 2007) emphasized that using screencasts can be effective while learning software. Because while learning the software, it is of great importance where the mouse is clicked and how the screen displays simultaneously. Luongo (2015) states that she received positive feedback from students in the classroom where he used screencasts in a face-to-face education environment. In addition, in her study, she mentions that instructors outside the classroom can apply this technique.

Instructors can use screencasts in software teaching, computer-based analysis, asynchronous communication with students, PowerPoint presentations, and solving extensive problems (Lloyd and Robertson, 2012). Students, on the other hand, have the opportunity to watch their weekly programs at any time they are suitable by stopping the recordings over and over again. It may be even more beneficial for students who have learning difficulties in lessons (Roache, 2006). This type of technology enables students to move at their own pace since they can pause the classes whenever and wherever they want, go back or watch as much as they need. It also converts screencasts into a digital format that guarantees the preservation of an instructor's live lecture (Kong, 2011). The effectiveness of screencast s in the classroom is directly related to the qualifications of the instructors (Harrison, 2020) and the creation of high-quality recordings. To create pedagogically qualified screencast recordings: (1) all of the theory, application, and comprehension activities should be included in the curriculum of the course; the instructor of the lesson should prepare (2) all-screencast s; (3) records must be relatively short; (4) instructors should speak clearly and comprehensively; (5) records should be checked repeatedly until they reach the desired quality (Brame, 2016; Ghilay and Ghilay, 2015; Rudow and Sounny-Slitine, 2015).

Screencasts are primarily used in lessons that require software applications such as GIS training. If I had taken the GIS course as a screencast throughout my undergraduate education, it would have been easier for me to learn the course. Because reaching the instructor's speed in conducting such a course plays a vital role in following effectively. Sometimes even missing a click of the instructor's mouse can make the lesson challenging to follow. Furthermore, as an instructor who has been taking the GIS course for years, I have often observed that students record the course with smartphones in addition to taking notes. Since the GIS discipline is closely related to technological developments, the teaching and learning models of the discipline are revised depending on these advancements, and it is known that GIS instructors are in search of new in the process (Argles, 2017; Bearman et al., 2016; Harvey and Kotting, 2011; Ooms et al., 2015; Mitchell et al., 2018). GIS education involves a delicate balance between theory and applications (Ooms et al., 2015). Higher education students are required to acquire theoretical knowledge and practice on GIS-related issues. However, it should be emphasized that practice is stressed for a more extended period compared to theory. It is essential to agree on a training platform where both would be ideal. For instance, it is easier to transfer the theory to students in face-to-face education compared to GIS education with screencast. Yet, screencasts also have significant advantages in terms of running applications. According to the advantages of screencasts, the background of the study is based on the view that GIS lessons, most of which are computer-based, can be conducted with screencasts.

Many researchers have focused on the use of screencasts in lessons for educational purposes. Some of them have highlighted screencasts can be used in computer-assisted lessons (Brame, 2016; Ghilay and Ghilay, 2015; Morris and Chikwa, 2014; Smith and Smith, 2012; Winterbottom, 2007). Since GIS applications are computer-aided, screencasts can be used in GIS lessons. Some studies describe GIS education as a discipline with a high potential to be developed by educators (Argles, 2017; Ooms et al., 2015; Vojteková et al., 2021; Zerger et al., 2002). After the outbreak of the pandemic in Turkey, but also worldwide, the form of education has changed. The education form has moved to the distance form of it, implemented by e-learning. Vojteková et al., (2021), pointed out the new method of distance GIS learning during the COVID-19 pandemic. According to above-mentioned research studies, it is good idea to develop alternative methods for the applied courses such as GIS education in the distance education period. For example, Vojteková et al., (2021), prepared videos for the GIS course with the JitsiMeet application. A similar alternative as in the example is also offered by this study. For this reason, the study aims to evaluate students' views by teaching GIS through screencasts and determining the most appropriate distance education format. Specifically, it is aimed to reveal the positive and negative points of the lesson and suggestions for the future GIS lessons by referring opinions of the students.

In the method section, the process of how the GIS lesson is conducted, how the screencasts are shared with the students, and how the data obtained from the interview is analyzed. In the results section, the qualitative data analysis is presented to the readers. The most important findings identified in this section are compared with previous researches in the discussion section. The conclusion section briefly summarizes the most important results identified and analyses the questions of the study. In addition, the discussion and conclusion sections offer some suggestions for GIS education and provide clues for further research.

MATERIALS AND METHODS

Technical Implementation: Screencasts Based GIS Instruction

This study was conducted with bachelor's degree sophomores of the geography-teaching department at Atatürk Education Faculty of Marmara University. During the GIS lesson, the videos created with screencast were shared with the students every week, and it was aimed for the students to learn the GIS theory and applications with the open-source QGIS software. Depending on the subject content, the duration of the screencasts varies between 3 and 20 minutes. While shorter screencasts are based on theoretical issues, longer screencasts are based on applications and reinforcement activities. The application and activities demonstrate how the instructor navigates the QGIS interface gradually. The author created high-resolution screencasts with the open-source OBS (Open Broadcaster Software) Studio software. There were 27 students enrolled in this class. The GIS course lasted 14 weeks, and all documents were shared with students via Google Classroom as a ZIP file. This platform served as a center for all educational activities of the course. Course documents cover PDF files of PowerPoint slides, all data, and videos used in education. For the lesson to be more efficient, videos related to theory and applications were created separately. After the theory and practice, comprehension and gaining momentum were included. This plays a vital role in reinforcing and summarizing the subject. The most critical goal in the recording process of videos is to make minimum editing after recording. Consequently, if any problems were encountered while checking the recordings, the videos were probably retaken.

The open-ended questions created to evaluate the course were directed to the students through the structured form. With this form, the opinions of geography teacher students regarding the GIS course were determined. The validity of the form was provided by submitting it to review by other field expert. In the end, questions were added to the evaluation form prepared by the researcher with the suggestions of the other field experts. A pilot study was conducted with two students to test the form's understandability before sending it to the students. The form was finalized in line with expert suggestions and students' opinions. Twenty-seven students participating in the assessment were coded as S1, S2,…S27 (Table 1). The researcher obtained the course data regarding the classes conducted in the fall semester of the 2020-2021 academic year via e-mail. It took approximately 15-20 minutes for the form to be answered by the students. Many students wrote more than one answer to the open-ended questions in the form. Questions directed to students are as follows:

1. What are your thoughts on the process of conducting the GIS course with a screencast?
2. What problems did you encounter during the process of conducting the GIS course with screencast?
3. What are your thoughts on the process of running the GIS course with the QGIS software?
4. Would you prefer this course to be conducted in face-to-face training, screencast, or hybrid? Explain with reasons.
5. What are your suggestions regarding the content and implementation of this course in the following years?

Code	Gender	Data Collection Date	Code	Gender	Data Collection Date
S1	M	02.17.2021	S15	F	02.20.2021
S2	M	02.13.2021	S16	M	02.19.2021
S3	F	02.18.2021	S17	F	02.17.2021
S4	M	02.18.2021	S18	M	02.17.2021
S5	F	02.22.2021	S19	M	02.22.2021
S6	M	02.18.2021	S20	M	02.19.2021
S7	F	02.16.2021	S21	F	02.13.2021
S8	M	02.16.2021	S22	M	02.18.2021
S9	M	02.16.2021	S23	M	02.19.2021
S10	F	02.14.2021	S24	F	02.16.2021
S11	M	02.19.2021	S25	M	02.18.2021
S12	F	02.22.2021	S26	M	02.14.2021
S13	M	02.13.2021	S27	F	02.19.2021
S14	M	02.16.2021			

Table 1: Characteristics of the students interviewed

As seen in table 1, 17 of the courses in which 27 students are enrolled are men, and 10 are women. The data were collected between 13.02.2021 and 22.02.2021. All of the participants are 18-23 who study at the department of geography teaching at the state university.

Content analysis technique was applied in analyzing the data obtained from students' expressions. The purpose of content analysis is to bring together similar expressions within the framework of specific themes and interpret them in depth (Patton, 2014; Yıldırım, A. and Şimşek, H., 2016). The first step in performing content analysis is the coding of data. Coding was done using MAXQDA, a widely-used qualitative data analysis software. First, data was collected with a Microsoft Word document, and answers were deciphered with MAXQDA software. In the deciphering process, the coding type called "coding within a general frame" was preferred. The three steps applied in terms of coding type are as follows: (i) detailed (explicit) coding was made, (ii) subcodes that mean the same with axis coding were brought together and integrated into themes, (iii) one or more statements of the students were directly quoted and given in quotation marks and interpreted.

After the data collection process, themes, categories, and codes were formed reading at least three times. The reliability rate was found according to the percentage agreement method of Miles and Huberman (1984). The reliability rate (82%) was calculated by dividing the number of agreements by the total number of agreements plus disagreements. The three steps applied to ensure the validity of the obtained data are as follows: (i) Interview data were repeatedly reviewed, (ii) participant expressions were discussed with a different researcher, (iii) a coding table was created. Simultaneously, in order not to be limited to the researcher's interpretation, the two researchers independently developed coding categories, and the results were compared. At this stage, a code table has been prepared for coding. While creating this table, sample code table was taken as a reference (Creswell, 2015). Consequently, coding was done only if the participants' statements were related to the definition in the code table. Table 2 illustrates the codes used in this study.

Categories	Codes	Number of Repetitions of Codes	Example
Advantages	Efficient Use of Time	19	*"I think the lesson time is used more efficiently in videos. Conducting the lesson shortly and concisely without encountering any problems does not distract our attention. It enables us to see the activities performed in the lesson in a more holistic way." (S10)*
	Watching Videos Again	18	*"Thanks to the lecture being given in the form of videos, I can understand an operation that I do not get by watching it over and over again." (S27)*
	The Elimination of the Boundary of Time and Space	14	*"Thanks to the screencasts, we can watch the lesson anytime and anywhere; this is a great benefit." (S23)*
	Adjusting the Video Speed	6	*"The fact that it is in the form of a video allows the student to watch the video in a shorter time by accelerating it according to his/her ability." (S2)*
Disadvantages	Students Failing to Ask Interactive Questions	13	*"I see it as a negative situation that I cannot ask the points I don't understand to the teacher and my friends." (S6)*
	Computer and Internet Related Problems	8	*"I can say that I lagged behind the narration in the video while loading the data into the program because the hardware of my computer was insufficient." (S25)*
	Too Much Lesson Time	4	*"The longer the lesson is in person, the better it is for us, but because we remain passive in front of the screen in online education, it sometimes becomes a problem to focus on videos for a very long time." (S24)*
	The Impermanence of What is Taught	3	*"The videos in distance education become more ,look then forget' version and we have to look constantly." (S9)*

Table 2: Examples of advantages and disadvantages of screencasts theme

RESULTS

The positive and negative aspects of the GIS course, which was conducted with a screencast throughout the semester, were considered from many angles. Based on the analyzes performed in line with the data obtained, the study findings were categorized under three broad themes: (i) the advantages/disadvantages of the screencast, (ii) the conduct of the lesson, and (iii) suggestions for future lessons.

Advantages of Screencast

The beneficial aspects of conducting the GIS course with screencast are the efficient use of time and watching the videos again. The time and space limits are eliminated, and the video speed can be adjusted. The categories, codes, and the number of repetitions of the codes in the theme are given in Table 2. Efficient use of time was determined as the most frequently repeated advantage (f = 19). With a planned and straightforward education, it can be ensured that the lesson time is used effectively and the students save time. Factors such as the courses being limited to screencasts, the disappearance of individual differences, and the absence of students who cannot catch up with the applications cause a decrease in lesson time. In lessons conducted face-to-face, the instructor's answer to each student who needs repetition and correction can extend the duration of the lesson. In addition to the quality of the videos, the long or short period may affect the efficient use of time. Below is one of the student statements to support the relevant findings:

"There are individual differences in teaching; not every student can learn at the same level and speed. Some students learn faster; some learn slower. In face-to-face education, many students with different individual learning speeds are given an average amount of time, thus preventing equal opportunity in education. With video lesson teaching, each individual can learn at their own pace and use time effectively. It is one of the most significant advantages of video lessons." S21

Another advantage of continuing the course with a screencast is that students can watch videos again. Thus, students have the opportunity to repeat the topics they think are missing or forgotten as many times as they desire. In addition, it should be noted that the instructor cannot repeat every lesson from the beginning to the end in a face-to-face education to meet all the demands of the student. One of the participant statements to support the findings is below:

"Sharing the lesson with us on a screencast helps us to repeat the lesson continuously and provides the opportunity to open and look at a topic we have forgotten." (S9)

One of the essential advantages of the screencast is that it allows students to learn whenever and wherever they want. Students' readiness may not be at the expected level on class days, or students may not be able to keep up with the pace of the educator. As can be understood from the expressions of the participants, the fact that technology offers such an option benefits students.

"Since the videos are recorded, they are not subject to a specific time limit. Therefore, although I do not have the opportunity to watch it at that time, I can watch it more comfortably at a different time." (S15)

One of the essential features of today's video playback tools is that the video speed can be adjusted according to the request and purpose. In other words, thanks to the feature, students have the opportunity to progress at their own pace. In this way, students can pass the subjects they think are sufficient or progress more slowly in the sections they have difficulty understanding. One of the participants' statements regarding this is as follows:

"There are some parts of the lesson that we think is not so important in the form of screencasts. Thus, I find it essential that videos can be watched at the desired speed." (S10)

Disadvantages of Screencast

In addition to the advantages of conducting the GIS course with a screencast, there are also disadvantages. The disadvantages are that students cannot ask interactive questions, computer and internet related problems, the duration of the lesson being too long, and impermanence of what is taught. The categories, codes, and the number of repetitions of the codes in the theme are given in Table 2. The most frequently repeated disadvantage ($f = 13$) was the students' inability to ask interactive questions. It is crucial that the students cannot simultaneously ask the instructor and their peers about their difficulty understanding. Being unable to ask questions and not interacting with other individuals in the classroom can be considered an obstacle to learning.
Furthermore, due to the lack of answers to the questions, the education process of the students may be interrupted. In this case, students may have to watch the screencasts repeatedly or skip without learning the subject. One of the participant statements to support the findings is below:

"In the training to be held face to face, we would learn the lesson by comparing it with many people. The lesson would be more fun. We could immediately ask the instructor about something we could not do or understand." (S8)

Another disadvantage of the course, which is conducted with screencasts, is the computer and internet related problems. The fact that the features of the computers such as working speed and screen quality are not at the desired level hinders the continuation of the course. It thus causes equal opportunity to be eliminated. However, poor internet speed can adversely affect the screencast's resolution and the efficient use of data. Mainly, the high raster data sizes can negatively affect the continuity of the course since it is closely related to both the quality of the computer components and internet speed. As can be seen from the following statement, some students do not have their computers or share their computers.

"We are three people studying university at home. I had a problem sharing the computer with my siblings. Our exams sometimes overlapped. And worse, the computer started freezing and shutting down for the last two weeks before the GIS exam." (S12)

Other disadvantages are the length of the lesson time and the lack of permanent teaching. The duration of the course hour may vary depending on the subject. Since it is not an interactive environment, some students stated that they could not focus on the lesson if duration of lesson gets longer. Again, for a similar reason, the lack of an interactive environment in the course may affect the permanence of the taught. Participant statements regarding the two disadvantages are as follows, respectively:

"It is important not to keep the lesson time longer than 1 hour. Since it is not interactive, the teacher's speech for more than 1 hour is not efficient." (S4)

"After watching and applying the lecture video, unfortunately, I realize that the information is not permanent. Maybe this is just an individual problem, I am not sure." (S20)

Conducting the Course

The students' expressions related to the course's theme, which has a comprehensive framework, are collected under three headings: the course content, completing the course with QGIS, and whether the course will be carried out face-to-face with a screencast or a hybrid education model.

Course Content

The expressions belonging to the course content category were classified as future utilization (*f* = 22), comprehension/acceleration activities (*f* = 22) and theory/practices (*f* = 15). Students express that they will benefit from the course in the future by recording screencasts, they will design maps based on the data shared in the course and the current data. When they become teacher, they want to teach geography lessons applying today's technological tools, and depending on this situation, the course may attract students. Student views on this are as follows:

"In the future, I can benefit from this software while designing maps with updated data and lecturing. According to the content of the subject to be covered, I can make the lesson better understood by using screencasts." (S1)

"When I become a teacher, I think of using GIS to increase the familiarity of students with the map and not to graduate without learning to prepare a map, albeit simple. I am sure students will be more interested in learning geography with today's technology. When I become a teacher, I would like to use recordings as a primary resource for remembering or rework GIS." (S17)

The comprehension and acceleration activities prepared to summarize the subject left positive impressions on the students. As can be seen from the following expressions, comprehension and acceleration activities play an essential role in reinforcing the issue, the permanence of what has been learned, and the preparation process for the exam.

"I think that having comprehension and acceleration activities at the end of the lesson has a significant effect on our learning both in terms of reinforcement of the lesson and preparation for exams." (S9)

"Thanks to the comprehension and acceleration activities at the end of the practice lesson, which the instructor also demonstrated in the screencast, I repeated myself and ensured the permanence of what I learned and transferred the information." (S21)

Firstly, screencasts consisting of theoretical information were prepared within the course scope, and then applications were included. Theoretical information was shared with the students as a PDF document. The statements of the two students regarding the theory and practices that are complementary to each other are as follows:

"Theoretical knowledge helps me make sense of the subject in my mind. Since the theory parts are recorded in PDF format, it is more permanent than a verbal narration for me. Therefore, I can progress by reading and understanding sentences one by one. The application part includes the theoretical knowledge being made by a guide. Thus, I find the opportunity to understand the missing places in theory or that I do not entirely understand." (S14)

"The theoretical information is straightforward and understandable. When we look at it, the information we encounter in some lessons must also be related to our class. There are no topics that we can say what its benefit is. The data is directly related to the content and objectives of the course. The applications are also very instructive in a way that theoretical knowledge is supported. Variations in the implementation phase can make the lesson more enjoyable." (S23)

Conducting the Course with QGIS

QGIS software was preferred for conducting the GIS course due to its open-source code and handy interface. The students' expressions belonging to the category were collected in 2 classes as positive thoughts and problems encountered. Detailed information about the classification is in Table 3.

Category	Codes	f	Category	Codes	f
Positive Factors	Downloading and Installing	24	Problems Encountered	The Interface is Not the Same	10
	Data Management	14		Viewing Data	7
	Ease of Use	9		Software Language	4
	Being Free of Charge	6		Subject Based	3
	Turkish Language Support	2		Internet Speed	2
	Appealing to Different Users	1		Software Shutdown Suddenly	1
	Updating	1		Software Opening Duration	1

Table 3: Categories, codes and number of repetitions for conducting the course with QGIS

Positive Factors

The positive opinions of the students regarding the conduct of the course with QGIS were determined. Particularly downloading and installing the software, data management, ease of use, and being free of charge have been defined as the most significant codes. Downloading and installing the software can be done effortlessly compared to much open-source software. Mainly the students with sufficient hardware and internet speed stated that they did not encounter any problems. One of the participant statements to support the findings is below:

"I did not encounter any problems with downloading and installing QGIS on my computer; it was effortless to install." (S16)

Another positive aspect of the course conducted with QGIS is the management of the data. Since data management has a structure that can differ from user to user, various features have been determined. These features can be listed as recording, viewing, storing, and analyzing data. Below are two of the expressions that determined the positive opinions of the participants on the related issues:

"For someone who is not very experienced in using a computer, data management and application interface, etc. may seem complicated at first, but I did not have a problem with these issues; I think it is an excellent program." (S11)

"The QGIS software did not cause any problems while storing the maps we created, adding raster or vector data to the program, and analyzing this data." (S17)

Ease of use is another positive finding obtained from teachers' expressions. More problems can be encountered, especially in the use of open-source software. Yet, as can be understood from the expressions, QGIS has an efficient, simple, and original design in the effective use of the interface while making applications or in map production. Two of the participant expressions in which positive thoughts were detected are below:

"In general, I enjoyed learning how to draw maps with this application since it has many features, and you can design products freely. When I saw other options while designing a map, noticing that this application will allow us to develop maps in almost all the ways we want, it will encourage me to try other buttons and create original products." (S2)

"I think the QGIS program makes the lesson better and easier to understand. A very useful program. I also found that as I spent time with the program, I learned new and more things." (S25)

The fact that the software is free, Turkish language support, appealing to different users, and being updated are other essential factors determined from the expressions of the participants. The fact that the software is free is a critical feature in increasing the number of users. The budget's being allocated for educational materials in our country is at a level that can be regarded as limited increases the importance of software even more. The fact that it provides Turkish language support significantly affects the use of the software by students. The software appeals to a wide range of users with its long-term and latest version options. Versions are updated in specific time frames. The students' views regarding all these positive factors are as follows:

"GIS software offers the maximum features a free software can." (S6)

"Turkish language support made it very easy for us to use the software." (S13)

"Thanks to its different versions, QGIS is also a perfect application that can appeal to different users." (S2)

"It's nice to see the app being updated." (S6)

Problems Encountered

In addition to the advantages of using the QGIS software, the problems faced by students are also not gone. A significant portion of the students stated that they encountered difficulties differentiating the interface and displaying the data. Other problems encountered were software language, subject-based, internet speed, sudden shutdown of the software, and opening time of the software. Today's software offers many alternatives for the user to design their interface practically and freely. Although the differentiation of the interfaces increases the practicality of the user, it makes it difficult to follow the lesson with screencasts. The fact that QGIS has many add-ons and manually adding these add-ons to the interface can negatively affect the process. In this sense, the views of a student who faced the problem are as follows:

"The program we use in this lesson may differ for every user. In particular, we had to add the tools we used in some projects manually. There were problems since we couldn't put everything we watched in the screencast into action immediately. Encountering these stages many times slows down learning." (S7)

Some of the students stated that they had problems viewing the data. Since a significant part of the data in GIS works with different extension files, such as the coordinate reference system, there is more than one extension file on the data insertion screen. Knowing which extension files will be added to the interface and selecting the correct extension file when dragging to solve the problem is essential. In addition, since the resolution of raster data is high, there may be a slowdown depending on the computer's performance and the internet. The views of a student who have encountered problems with data viewing are as follows:

"Sometimes it failed while uploading data to the computer. Because the data in the video could be loaded quickly and opened without any problems when loaded, but although I did the same as in the video, there were sometimes errors." (S8)

Although QGIS is Turkish-supported, it does not have the desired level of Turkish writing language functionality in many features such as tools, panels, plug-ins, warnings, etc. Students may also encounter subject-based problems. The difficulty level of some subjects requires higher-level skills compared to others. This situation makes it difficult to follow the course. The process of displaying high-sized raster data in the interface when adding base maps is directly proportional to the internet speed. In some unforeseen situations or primarily due to user error, QGIS may shut down suddenly.

Eventually, depending on the features of the computer, the first opening of the software requires a long time. Examples of student statements regarding the problems mentioned in the paragraph are given in Table 4.

Codes	Example
Software Language	*"The language of the version I used was mixed English and Turkish version. Since Since I do not speak English, I had a problem, but I solved it with the help of the internet." (S21)*
Subject Based	*"I had problems with join, area calculator, and spatial interrogation by condition. I solved these problems by watching the class videos again." (S1)*
Internet Speed	*"Some data may take a long time to display, depending on internet speed." (S5)*
Software Shutdown Suddenly	*"I run over the course because the application suddenly closed while I was using it." (S12)*
Software Opening Duration	*"It keeps a little longer while opening the application." (S22)*

Table 4: Students' statements about other problems encountered

Screencast or Face-to-Face or Hybrid?

The question "Would you prefer to conduct the course face-to-face or hybrid with a screencast" and their reasons were asked to the students. Of the 27 students, 14 preferred screencasts, seven preferred hybrid training, and six face-to-face training. The reasons behind the preferences also reveal that GIS training is more accepted with screencasts. We have mentioned the advantages and disadvantages of screencasts earlier. The benefits of screencasts are more effective in choosing this method more (Table 2). Factors such as the ability to watch the videos again, the efficient use of the lesson time, and the students' ability to choose the time and place are the factors behind the choice, as can be seen from the statements of the two students below.

"When the training is given to more than one person in face-to-face education, there is a constant interruption due to the students' diverse comprehension abilities during the course, and it disrupts the lesson. Going back to unclear topics repeatedly can also lead to confusing situations. When it is processed with a screencast, everyone can learn the parts they do not understand or are missing by watching the video again easier by watching the lecture video at any time. In this respect, I think it is more appropriate to conduct the lesson via screencast." (S13)

"I would definitely prefer it to be played with screencasts. If I were taking the GIS course face to face, I would feel stress and pressure on myself, as it would be more difficult for me to follow the teacher. Sometimes it can take a long time because I do not have prior knowledge to perform the most effortless operation on the computer, and I do not gain enough practicality. In face-to-face education, this situation could lead to incompatibility in the classroom. While the teacher was explaining a different process, my coming back might have slowed down the processing of the lesson. For such reasons, I prefer the GIS course to be conducted with videos. To be able to access lecture videos whenever I need them, to watch an action again, to reapply it, etc. It is advantageous for me. In this way, I can progress at my own pace and benefit from videos within the program I have determined." (S27)

Unquestionably, each model has its pros and cons. Hybrid education can make a difference because it involves the other two education models. Face-to-face teaching in Flipped Classroom is a method that is suitable for the blended learning hybrid education model. The student completes a specific part of the lesson at home and the rest in the classroom. A training model in which both face-to-face and screencast training will be applied may be ideal. Students both have the opportunity of in-class education. They can repeat the subjects they need to repeat whenever and wherever they want. The statements of the two students regarding this situation are as follows:

"I prefer this course to be hybrid. I think it would be much more effective teaching if acceleration activities are shared with students as homework or videos to watch. A more beneficial lesson will emerge with the combination of the positive aspects of video and face-to-face training." (S6)

"We would suffer a lot of time and subject in the face-to-face training because it was a computer lesson, but we would learn the lesson by comparing it with many people. The lesson would be more fun. We could immediately ask the teacher about something we could not do or understand. When we go home after receiving the face-to-face training, we sometimes forget the topics, so I also think that the narration is necessary, thanks to the screencast, to reinforce this lesson and to help us later.*"* (S8)

The reasons for students who prefer screencasts are directly proportional to the advantages of screencasts. The reasons for the students who prefer face-to-face education are also directly proportional to the disadvantages of screencasts. As can be seen from the student views below, the need to teach the course interactively comes to the fore.

"The GIS course is a course that needs to be progressed interactively. When the student makes any mistake, the educator's intervention in that error should be instant. Because a mistake made in the GIS course can cause the application to start over." (S4)

"It is certain that it will be better to conduct the course face to face with the instructor to find immediate solutions to more permanent and immediate problems." (S22)

Suggestions for Future Courses

Students' opinions were sought to improve the quality of the future lessons. The codes in the theme, the number of repetitions, and examples of the codes are in Table 5. Many applications were shared with the students in practice or comprehension and acceleration activities depending on the subject. Despite this, students suggest increasing the practices in the following lessons. Some of the students also made suggestions to reduce the share of theory in the lesson's teaching. Since the study group consists of students, the students associate the maps with the maps used in the high school geography lesson and recommend that the course content be enriched in this respect. With the presence of many applications, no homework was given in the course. It has been determined that if the homework is not given, the students will mostly work during the exam week, or the learned ones will not be permanent. Even though much data is shared during the course, students suggest that more diverse data should be shared. In addition to all these, a student indicates that the course should be taught by coding.

Codes	Number of Repetitions of Codes	Example
Increasing Applications	11	*"I think there should be more comprehension and acceleration activities within the course than theoretical knowledge." (S5)*
Mitigating the Theory	7	*"The effect of the theoretical information presented in the course is less than the applications. Reducing the theoretical knowledge in some subjects and giving more place to the applications will make it easier for them to teach the course." (S7)*
Subject-Based Applications	5	*"I would like to learn more in terms of the use of GIS in the teaching of which units at grade levels in the geography lesson curriculum." (S21)*
Assigning Homework	3	*"GIS is ultimately a programming course, and the more practice is, the better. My humble advice is that I think more permanent learning can be achieved by giving weekly practical assignments to students." (S17)*
More Data Sharing	2	*"It would be appropriate to share more data with students in a categorized way. Thus, the student will be able to make different applications using that data." (S15)*
Including Coding	1	*"I would like to learn more information based on coding in the course." (S10)*

Table 5: Codes, number of repetitions of codes and examples

DISCUSSION

There are GIS courses in many disciplines in higher education. A significant portion of these courses is conducted face-to-face, some online, and some with hybrid education. In addition to all these, it can be mentioned that there are researchers who have tried different educational models or who are looking for a model. The computer-based applications needed in conducting GIS courses are the common point of education models. It is known that screencasts in computer-aided lessons will benefit the education process (Brame, 2016; Ghilay and Ghilay, 2015; Morris and Chikwa, 2014; Smith and Smith, 2012). It indicates that screencast can be used effectively in GIS lessons. Zerger et al. (2002) mention that GIS education is a discipline with a very high potential to be developed by practitioners and educators due to its nature. Similarly, Argles (2017) used a screencast in some of the GIS training, and the students liked the videos. Vojteková et al. (2021) shared the videos they prepared for the GIS course with the JitsiMeet application which was programmed for their university during the COVID 19 process and students found the application highly useful. According to the findings of this research, this course, which is conducted with screencasts for one semester, has advantages and limitations. Despite some limitations, 14 out of 27 students who attended the course support the continuation of the course with a screencast in the coming years with their reasons.

Benefits

First of all, it should be stated that screencast has many advantages regardless of GIS training. However, we know that the power of screencast is even more prominent in computer applications (Ghilay and Ghilay, 2015; Smith and Smith, 2012). Conducting the lesson with this model when the video editing rules are followed and used effectively in computer applications enables students to use their time efficiently and watch the videos repeatedly whenever they want, in the place and speed they want (Vojteková et al., 2021). Even these advantages alone form the background of the course being taught with this model or being accepted by more students. In addition, the elimination of individual differences, that is, each student's ability to work according to their own learning pace or readiness is another critical advantage. When GIS applications are performed face to face in classrooms, not every student can keep up with the speed of the educator. Even a single click with the mouse missed by the student can disrupt the flow, which may cause the lesson's duration to be prolonged. The negativities mentioned above are eliminated with the screencast, especially realizing applications and increasing the number of applications becomes easier. During the course, the number of applications was increased by comprehension/gaining speed activities. This situation gained the appreciation of many students. Open source QGIS software was preferred in the course. The materials used in educational environments being free of charge, especially in developing countries such as Turkey (education supports and services, in general, differ considerably compared to developed countries), indicating a significant obstacle to the widespread use of the material be eliminated. In addition, the software has many users around the world. Being a user-friendly software, organizing the interface according to the user, being free of charge and simple data management increase the attractiveness of QGIS.

As an educator, I observed that students' success and their grades in exams increased compared to face-to-face lessons. Especially the students who were not good with computers were having problems in their GIS lessons. Since the applications were not registered, they asked me to repeat the lesson many times. We could do this a limited number of times. The readiness of the course videos means that the instructor devotes more time to the questions from the students regarding the course, gives feedback to the students, and provides personal guidance. It is also an excellent advantage for me to be able to use the current videos in future lessons. In this process, I also observed that students could make their map designs at a higher level. Moreover, it opens the door to reach individuals who have the motivation to learn GIS through registrations and different platforms. Although it is challenging to guide them as they are not enrolled in formal education, it is positive for my motivation that the education reaches more people.

Limitations, Possible Solutions and Further Research

In addition to the benefits of the screencast, there are also many limitations to be considered. However, as long as the instructors are willing to plan the future effectively and strategically throughout the term, limitations seem to be resolved (Rudow and Sounny-Slitine, 2015). The most crucial disadvantage is that students continue the learning process independently of the educator and other friends. When students face a problem, they may have difficulty asking the instructor or peers questions, interrupting the learning process (Ronchetti, 2010; Vojteková et al., 2021). The differences in the software interface are the possible causes of the problems experienced in the applications. Although users can differentiate the interface according to their wishes, a standard interface can be designed for solutions at specific points. It is necessary to take precautions against problems arising from the quality of computer equipment and the internet. The doors of the existing laboratory can be kept open to students with inadequate equipment or low internet speed. Since one of the students enrolled in the course did not have a computer, he used the GIS laboratory. Another disadvantage is that the theory-based lessons are also conducted with the screencast. In this regard, students should be extra motivated by using technological tools effectively (Ooms et al., 2015). For the theory to be successful, it is necessary to stay away from presentations containing only information, increase the visuals and keep the video time short. When we consider the disadvantage for the instructor, the problems that arise in the preparation of the screencast are the main problems. In this regard, it is necessary to have specific qualified equipment such as headphones. Besides, a quiet environment should be preferred for better recording. Because the emergence of independent sounds that disrupt the flow during recording may mean that the recording is retaken or divided.

Researchers can plan their future studies by (i) taking measures against the disadvantages of the screencast, (ii) simplifying the interface of the software used in screencasts and moving from a standard design, (iii) paying attention to increasing the visuality in the presentation of theoretical information or adopting the hybrid education model in the theory part of the course, (iv) taking strict measures against the problems that may arise during the recording process.

CONCLUSION

In this study, second-year GIS lecture series were delivered via screencasts in geography teaching department. The rationale behind the survey was to teach GIS with the help of screencasting and to determine the most appropriate formats for distance education. The results show that the students welcomed the conduct of the entire GIS lecture with screencast technology. Instructor qualifications and the desire of instructors to produce technically first-class screencast play an important role in the effectiveness of video recordings. Screencasts have the flexibility to learn independently and students can use their time efficiently and watch the videos repeatedly whenever and wherever they want. However, there is no doubt that the instructor will have more difficulty preparing the lesson with the screencast. Although recording and editing videos require more time than face-to-face training, using recordings prepared in the coming years also means saving time in this process. Most importantly, the instructor can ensure that students get the maximum efficiency from the lecture. Based on my own experiences and student opinions, it is clear that the lecture is more efficient with the screencast. In addition, the instructor can deliver screencasts to a much larger number of students through different platforms.

ACKNOWLEDGEMENT

This paper grew out of the 2nd Istanbul International Geography Congress, organized by Istanbul University, Geography Department and Turkish Geographical Society on June 17-18, 2021. It was benefited enormously by comments and suggestions from the reviewers of this journal.

REFERENCES

Albăstroiu, I., Felea, M. and Vasiliu, C. (2014) 'Geographic Information System-modern teaching method in business administration', *Amfiteatru Economic Journal*, Vol. 16, No. 37, pp. 770–783.

Argles, T. (2017) 'Teaching practical science online using GIS: a cautionary tale of coping strategies', *Journal of Geography in Higher Education*, Vol. 41, No. 3, pp. 341–352. https://doi.org/10.1080/03098265.2017.1315531

Bao, W. (2020) 'COVID-19 and online teaching in higher education: A case study of Peking University', *Human Behavior and Emerging Technologies*, Vol. 2, No. 2, pp. 113–115. https://doi.org/10.1002/hbe2.191

Bearman, N., Jones, N., André, I., Cachinho, H. A. and DeMers, M. (2016) 'The future role of GIS education in creating critical spatial thinkers', *Journal of Geography in Higher Education*, Vol. 40, No. 3, pp. 394–408. https://doi.org/10.1080/03098265.2016.1144729

Brame, C. J. (2016) 'Effective educational videos: Principles and guidelines for maximizing student learning from video content', *CBE—Life Sciences Education*, Vol. 15, No. 4, pp. 1–6. https://doi.org/10.1187/cbe.16-03-0125

Burgess, S. and Sievertsen, H. (2020) *Schools, skills, and learning: The impact of COVID-19 on education*, [Online], Available at: https://voxeu.org/article/impact-covid-19-education [21 Jul 2021].

Catalano, A. (2014) 'Improving distance education for students with special needs: A qualitative study of students' experiences with an online library research course', *Journal of Library & Information Services in Distance Learning*, Vol. 8, No. 1–2, pp. 17–31. https://doi.org/10.1080/1533290X.2014.902416

Chen, T., Peng, L., Yin, X., Rong, J., Yang, J. and Cong, G. (2020) 'Analysis of user satisfaction with online education platforms in China during the COVID-19 pandemic', *Healthcare*, Vol. 8, No. 3, 200. https://doi.org/10.3390/healthcare8030200

Creswell, J. W. (2015) *30 essential skills for the qualitative researcher*, Thousand Oaks, California: Sage Publications.

Gambari, A. I. and Hassan, S. A. (2017) 'Effects of instructional screencast on the performance of National Open University undergraduates in educational technology in Kwara State, Nigeria', *Bulgarian Journal of Science and Education Policy*, Vol. 11, No. 1, pp. 132–159.

Ghilay, Y. and Ghilay, R. (2015) 'Computer Courses in Higher-Education: Improving Learning by Screencast Technology', Journal of Educational Technology, Vol. 11, No. 4, pp. 15–26 https://doi.org/10.26634/jet.11.4.3148

Gormely, K. and McDermott, P. (2011) 'Do You Jing? How Screencasting Can Enrich Classroom Teaching and Learning', *Language and Literacy Spectrum*, Vol. 21, No. 1, pp. 12–20.

Harrison, T. (2020) 'How distance education students perceive the impact of teaching videos on their learning', *Open Learning: The Journal of Open, Distance and e-Learning*, Vol. 35, No. 3, pp. 260–276. https://doi.org/10.1080/02680513.2019.1702518

Harvey, F. and Kotting, J. (2011) 'Teaching mapping for digital natives: New pedagogical ideas for undergraduate cartography education', *Cartography and Geographic Information Science*, Vol. 38, No. 3, pp. 269–277. https://doi.org/10.1559/1523040638269

Hasanudin, C. and Fitrianingsih, A. (2018) 'Flipped classroom using screencast-o-matic apps in teaching reading skill in Indonesian language', *International Journal of Pedagogy and Teacher Education*, Vol. 2, pp. 151–158. https://doi.org/10.20961/ijpte.v2i0.25356

Kauppi, S., Muukkonen, H., Suorsa, T. and Takala, M. (2020) 'I still miss human contact, but this is more flexible—Paradoxes in virtual learning interaction and multidisciplinary collaboration', *British Journal of Educational Technology*, Vol. 51, No. 4, pp. 1101–1116. https://doi.org/10.1111/bjet.12929

Kawaf, F. (2019) 'Capturing digital experience: The method of screencast videography', *International Journal of Research in Marketing*, Vol. 36, No. 2, pp. 169–184. https://doi.org/10.1016/j.ijresmar.2018.11.002

Kong, S. C. (2011) 'An evaluation study of the use of a cognitive tool in a one-to-one classroom for promoting classroom-based dialogic interaction', *Computers & Education*, Vol. 57, No. 3, pp. 1851–1864. https://doi.org/10.1016/j.compedu.2011.04.008

Lloyd, S. A. and Robertson, C. L. (2012) 'Screencast tutorials enhance student learning of statistics', *Teaching of Psychology*, Vol. 39, No. 1, pp. 67–71. https://doi.org/10.1177/0098628311430640

Luongo, N. (2015) 'Missing the chalkboard: Using screencasting in the online classroom', *Computers in the Schools*, Vol. 32, No. 2, pp. 144–151. https://doi.org/10.1080/07380569.2015.1030968

Miles, M. B. and Huberman, A. M. (1984) *Qualitative data analysis: A sourcebook of new methods*, Beverly Hills: Sage Publications.

Mitchell, J. T., Roy, G., Fritch, S. and Wood, B. (2018) 'GIS professional development for teachers: Lessons learned from high-needs schools', *Cartography and Geographic Information Science*, Vol. 45, No. 4, pp. 292–304. https://doi.org/10.1080/15230406.2017.1421482

Morris, C. and Chikwa, G. (2014) 'Screencasts: How effective are they and how do students engage with them?', *Active Learning in Higher Education*, Vol. 15, No. 1, pp. 25–37. https://doi.org/10.1177/1469787413514654

Onete, B., Pleșea, D., Teodorescu, I. and Cîrstea, A. (2014) 'Evolutions and opportunities of business education in the context of educational reform from the digital age', *Amfiteatru Economic Journal*, Vol. 16, No. 37, pp. 746–758.

Ooms, K., De Maeyer, P., De Wit, B., Maddens, R., Nuttens, T., Van De Weghe, N. and Vervust, S. (2015) 'Design and use of web lectures to enhance GIS teaching and learning strategies: the students' opinions', *Cartography and Geographic Information Science*, Vol. 42, No. 3, pp. 271–282. https://doi.org/10.1080/15230406.2015.1011236

Oud, J. (2009) 'Guidelines for effective online instruction using multimedia screencasts', *Reference Services Review*, Vol. 37, No. 2, pp. 164–177. https://doi.org/10.1108/00907320910957206

Paraskeva, F., Bouta, H. and Papagianni, A. (2008) 'Individual characteristics and computer self-efficacy in secondary education teachers to integrate technology in educational practice', *Computers & Education*, Vol. 50, No. 3, pp. 1084–1091. https://doi.org/10.1016/j.compedu.2006.10.006

Patton, M. Q. (2014) *Qualitative research & evaluation methods: Integrating theory and practice*, 4th edition, Thousand Oaks, CA: Sage publications.

Peterson, E. (2007) 'Incorporating screencasts in online teaching', The International Review of Research in Open and Distributed Learning, Vol. 8, No. 3, pp. 1–4. https://doi.org/10.19173/irrodl.v8i3.495

Roache, J. (2006) 'Using screen capture technology to develop on-line course material', *Proceedings of Society for Information Technology and Teacher Education International Conference*, Chesapeake, pp. 519–520.

Ronchetti, M. (2010) 'Using video lectures to make teaching more interactive', *International Journal of Emerging Technologies in Learning*, Vol. 5, No. 2, pp. 45–48. https://doi.org/10.3991/ijet.v5i2.1156

Rudow, J. and Sounny-Slitine, M. A. (2015) 'The use of web-based video for instruction of GIS and other digital geographic methods', *Journal of Geography*, Vol. 114, No. 4, pp. 168–175. https://doi.org/10.1080/00221341.2014.977932

Ruffini, M. F. (2021) *Screencasting to Engage Learning*, [Online], Available at: https://er.educause.edu/articles/2012/11/screencasting-to-engage-learning [2 Mar 2021].

Smith, J. G. and Smith, R. L. (2012) 'Screen-capture instructional technology: A cognitive tool for designing a blended multimedia curriculum', *Journal of Educational Computing Research*, Vol. 46, No. 3, pp. 207–228. http://dx.doi.org/10.2190/EC.46.3.a

Snyder, C., Paska, L. M. and Besozzi, D. (2014) 'Cast from the past: Using screencasting in the social studies classroom', *The Social Studies*, Vol. 105, No. 6, pp. 310–314. https://doi.org/10.1080/00377996.2014.951472

Tartavulea, C. V., Albu, C. N., Albu, N., Dieaconescu, R. I. and Petre, S. (2020) 'Online Teaching Practices and the Effectiveness of the Educational Process in the Wake of the COVID-19 Pandemic', *Amfiteatru Economic*, Vol. 22, No. 55, pp. 920–936. https://doi.org/10.24818/EA/2020/55/920

Vojteková, J., Tirpáková, A., Gonda, D., Žoncová, M. and Vojtek, M. (2021) 'GIS Distance Learning during the COVID-19 Pandemic (Students' Perception)', *Sustainability*, Vol. 13, No. 8, 4484. https://doi.org/10.3390/su13084484

Winterbottom, S. (2007) 'Virtual lecturing: Delivering lectures using screencasting and podcasting technology', *Planet*, Vol. 18, No. 1, pp. 6–8. https://doi.org/10.11120/plan.2007.00180006

Yıldırım, A. and Şimşek, H. (2016) *Sosyal bilimlerde nitel araştırma yöntemleri*, Ankara: Seçkin Publication.

Zerger, A., Bishop, I. D., Escobar, F. and Hunter, G. J. (2002) 'A self-learning multimedia approach for enriching GIS education', *Journal of Geography in Higher Education*, Vol. 26, No. 1, pp. 67–80. https://doi.org/10.1080/03098260120110377

Full research paper

STATISTICS TEACHING PRACTICE AT CZECH UNIVERSITIES WITH EMPHASIS ON STATISTICAL SOFTWARE

Aneta Mazouchová[1]✉
Tereza Jedličková[2]
Lucie Hlaváčová[3]

[1]Department of Economic Statistics, Prague University of Economics and Business, Czech Republic

[2]Institute of Geology and Paleontology, Faculty of Science, Charles University, Czech Republic

[3]Department of Biology and Environmental Studies, Faculty of Education Charles University, Czech Republic

✉ anetamazouchova@gmail.com

ABSTRACT

This paper aims to reveal the beliefs of students and fresh university graduates about teaching statistics during their university studies with focus on using statistical software. The objective is to detect the approach of faculties to statistics education and to find out which didactic materials and teaching methods are mainly used. Students' opinions are captured by means of a questionnaire survey and analysed both quantitatively and qualitatively. The results show the increasing importance of quantitative research and the necessity of improving statistical thinking. Unfortunately, the teaching methods used in various statistical courses are outdated and unattractive for most students. They call for an active and modern approach. Teaching statistics with the statistical software support seems to be the right way to make statistics accessible to students. The recommendation is to take students' notions into account when preparing statistical courses.

KEYWORDS

Statistical courses, statistical software, student's demands, teaching statistics, university student, questionnaire survey

HOW TO CITE

Mazouchová A., Jedličková T., Hlaváčová L. (2021) 'Statistics Teaching Practice at Czech Universities with Emphasis on Statistical Software', *Journal on Efficiency and Responsibility in Education and Science*, vol. 14, no. 4, pp. 258-269. http://dx.doi.org/10.7160/eriesj.2021.140405

Article history
Received
December 12, 2020
Received in revised form
June 11, 2021
Accepted
November 10, 2021
Available on-line
December 20, 2021

Highlights

- *The research of this paper was conducted both quantitatively and qualitatively.*
- *The survey was carried out across all Czech universities and 763 respondents took part.*
- *Students consider the teaching methods of statistics at Czech universities to be outdated and too theoretical.*
- *According to the addressed students, Czech universities do not provide students with the opportunity to get acquainted with any statistical software.*
- *To be effective, the teaching methods of statistics have to adapt to students' needs, especially in the field of humanities.*

INTRODUCTION

University students frequently consider statistics as one of the most challenging subjects. They often even speak about the fear of statistics (Cobb, 1992; Ralston et al., 2016). The fear/anxiety of statistics is described both in foreign (e.g. Steinberger, 2020) and Czech studies (e.g. Hybšová, 2017; Widenská, 2014). Many studies examine students' attitudes toward statistics and address the issue of teaching statistics to eliminate the fear of statistics (e.g. Dempster and McCorry, 2009; Kurniawan et al., 2019; Ralston et al., 2016). However, such studies are unique in the Czech Republic and have the character of „best practice" rather than research (e.g. Hindls and Hronová, 2005; Skalská, 2019).

Teachers of statistics and other statistical subjects face such prejudice. In response, sufficient attention must be paid not only to the course design and the learning context, but also to the mind of individual students and motivating their learning process (Law, Lee and Yu, 2010). The subject must adapt didactically to a specific field of study to view statistics as a science that accompanies and helps students find answers for scientific questions (Hybšová, 2017). The development and increase of teaching quality help to greater popularity and understanding statistics as a scientific discipline.

Given the above mentioned, preparation of a quality statistical subject is demanding and affected by numerous factors, e.g. the field of study, the initial level of mathematical and statistical knowledge that differ among students. Scheduling of statistical subjects during studies plays an important role, as well. Students in higher study groups are more likely to be motivated in statistics courses because they can assume the need to use statistics when working on their thesis. According to Velleman and Moore (1996) it is essential to motivate students of introductory courses and to change their negative approach towards statistics mainly influenced by negative perceptions of statistics by population.

Selection and utilisation also affect the resulting quality (Hybšová, 2017). Despite numerous studies about innovations in the education of statistics (see, e.g. Huynh and Baglin, 2017; Billig and Waterman, 2014), its teaching in the Czech Republic remains very rigid. It is not uncommon that teachers prefer the "pen-and-paper" method and other didactic tools like statistical software (SW) remain unused. An example of the efficiency and usefulness of using interactive materials during the education process was verified and published by Aberson et al. (2002) and Law, Lee and Yu (2010) confirmed the efficiency and usefulness of electronic education in their paper. Using statistical software has motivational potential as its use eliminates the stress caused by statistical calculations and students focus more on the practical use of statistics (Hsu, Wang and Chiu; 2009). Stejskalová et al. (2019) describe the case studies of real-life students to consider an appropriate complement to the traditional method of teaching.

Students are afraid of calculations (Cobb, 1992). Therefore, using statistical software in teaching statistics offers a way to perform the feared statistical calculations. Thus, students focus more on practical significance (Hsu, Wang and Chiu; 2009). There are a number of paid as well as free statistical programmes that can be used during a course. The teacher needs to consider the choice of software to work with carefully. This selection determines statistical software that the students are the most likely to use when writing their thesis. If the teacher wants the students to use statistics after graduation, the choice should tend to software that does not require an expensive licence. Another reason for choosing a specific software is e.g. the existence of demonstrative tutorials. According to Poláčková and Jindrová (2010) multimedia presentations can make the teaching process of statistical classes more efficient. Therefore, students can understand how to work with and use statistical software faster and easier. These tutorials are also beneficial for teachers, who can spend more time explaining the statistical methods instead of focusing on the software. Tutorials can help less experienced information technology (IT) users to overcome potential problems with navigation through the statistical software.

The objective of this paper is to detect the approach of faculties to teaching statistics, to find out which didactic materials and teaching methods are mainly used and to highlight the main issues and concerns of students regarding statistics. The paper is divided into method description, results and discussion. The conclusion summarises the most important results and recommendations for further research.

This paper follows up on the conference paper of Mazouchová, Jedličková and Hlaváčová (2020).

METHODOLOGY AND RESEARCH QUESTIONS

This paper focuses on the use of statistical software in university studies. The respondents were students in at least their second year of study and fresh graduates from all Czech public universities. A self-administered survey was employed. In line with De Leeuw and Hox (2008), respondents received an email that invited them to respond to a web-based survey, using a clickable link for easy access included in a personalised email. Respondents were addressed through all public Facebook groups for students and graduates. The questionnaire included a sorting question about the year of study of the respondent. The respondents could continue answering the questionnaire if they were studying at least the second year of a bachelor programme or finished university less than 5 years ago.

We tried to obtain the largest possible sample to capture the overall situation in the Czech Republic. Within this sampling, it is predictable that the experience varies among different fields of study. This must be taken into account when interpreting the results.

The survey received a total of 763 fully completed questionnaires from students of 14 public universities. Women represented 79% of the respondents, men 21%. This trend is in line e.g. with Curtin, Presser and Singer (2000), Moore and Tarnai (2002) or Singer, van Hoewyk and Maher (2000). 39% of the questionnaires were responded by university graduates, 18% of respondents were students of a bachelor study programme, 36% studied in a master programme and 7% of the respondents studied in doctoral studies.

The first part of the questionnaire focused on finding out whether the respondents attended any statistics course during their university studies. The next part concentrated specifically on the use of statistical software in their classes and the student's ability to use the given software. Subsequently, the respondents were asked about the use of statistical methods when processing their final theses. Finally, respondents evaluated several statements related to statistics and its use during their university studies on the 5-point Likert scale.

We used the Hierarchical Cluster Analysis with Ward's Method to verify the structure of the responses to the sentiment questions. The arising clusters divided respondents into 4 groups (see Table 2). The clusters were frequent and demonstrated rather even distribution.

The dependence of the statements on sex and on student participation in at least one statistics course was investigated. For the purpose of comparison, the statements were recorded as follows: 1 = I totally disagree, 2 = I do not agree, 3 = I do not know, 4 = I agree, 5 = I completely agree. Means and standard deviation of groups were calculated. Mann-Whitney U tests were applied to compare two groups. P-value was evaluated at the 5% level of significance.

The final question of the survey asked respondents to freely comment on the given issue. 117 participants (15%) used this opportunity. Their open responses were qualitatively evaluated by means of content analysis. The initial analysis used the Open Coding in Hand method according to Hendl (2005) and Strauss and Corbin (1999). The response texts were searched for meaningful units, which were then marked. In accordance with Erlingsson and Brysiewicz, 2017) the meaningful units consisted of single words, word sequences or whole sentences. Such marked units received codes and were entered into tables, creating a clear list according to Lee and Fielding (2004).

Subsequently, the codes were adjusted, statements with the same content and different formulation were joined and marked with a modified code (Švaříček and Šeďová, 2007). Some codes were left on purpose in the long and detailed version to avoid losing data by shortening them. The listed codes were then analysed according to Švaříček and Šeďová (2007), the same topics found and codes grouped into categories and subcategories. Relations and key topics were identified.

Due to the fact that the respondents studied a wide range of study fields, this paper rather represents a survey. It is not advisable to generalise the conclusions using inductive statistical methods. The acquired data were evaluated only by means of descriptive statistics.

Five research questions were specified with respect to the defined objectives:

- What statistical software do university students know and are able to use?
- What problems do students face when using statistical software?
- Do university students process the statistical data for their thesis themselves?
- What factors affect opinions regarding statistics?
- What comments do students have in terms of statistical courses at Czech universities?

RESULTS

Initially, we asked respondents if they completed any statistics course during their studies. 31% of respondents did not complete any statistics subject while 34% of respondents completed one semester of statistics and 35% completed two and more semesters.

The following part of the questions was intended only for students who completed at least one statistics course ($n = 526$). The questions concentrated on the statistical software used in the classes. Respondents could choose from suggested answers with various types of statistical software (including MS Excel) or write their own response.

More than a half of the respondents (53%) used MS Excel, while 28% stated they used no statistical software at all. Some students also used Statistica (19%), SPSS (13%) and R (10%) in their classes. Jamovi (3%), SAS (2%), JASP (2%), Gretl (1%) and GraphPad (0,5%) remain virtually unused. Figure 1 shows software mentioned more than once in the responses.

The third part of the questionnaire concentrated on the use of statistical methods when processing a thesis. 72% of respondents stated that they used statistical methods to process their theses. 20% out of them admitted that the data for their theses were processed by a statistician, their supervisor or another person.

The main question of the survey asked what statistical programme respondents know and can use. The one known and used by almost all respondents was MS Excel, which, in their opinion, they can use well or at least partially. Other known programmes include Statistica, SPSS, R and SAS. Paradoxically, the JASP and Jamovi freeware software are the least known and used (see Figure 2).

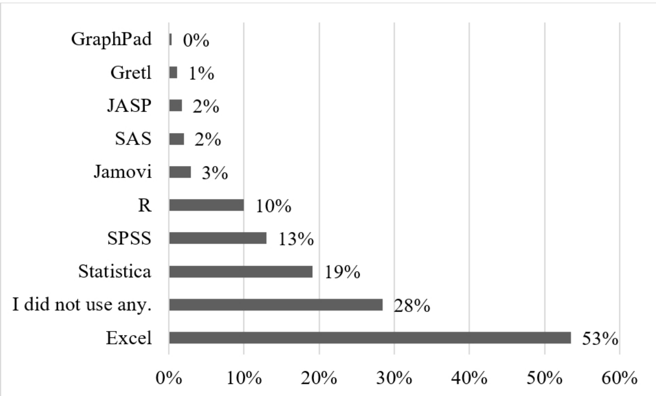


Figure 1: Statistical software used in lessons (*n* = 526), 2019–2020

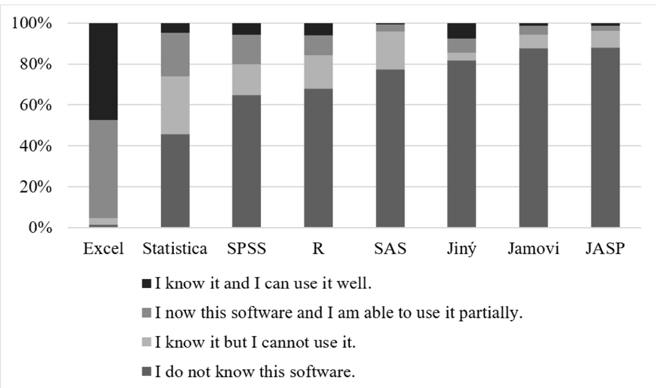


Figure 2 Answers to question "What software do you know and can use?" (*n* = 763), 2019–2020

The next part of the questionnaire asked the respondents to evaluate eight statements. They had a five-point scale available to express their sentiment. Table 1 shows the statements and answers. 55% of respondents claim that statistical software in English represents no problem. Comprehending statistics constitutes a more complicated issue for many students, with 55% of respondents stating that they do not understand statistics. Moreover, respondents have no preference between qualitative and quantitative research (38% prefer quantitative and 36% qualitative research). Almost a third of respondents (32%) state that their supervisor advised them on the appropriate methods for data processing.

Statement	I totally disagree (1)	I do not agree (2)	I do not know (3)	I agree (4)	I completely agree (5)
Statistical software in English is a problem.	25%	30%	15%	22%	7%
The problem is that I do not understand statistics.	10%	22%	13%	36%	19%
The problem is that I do not understand statistics and the software is in English.	17%	28%	18%	21%	15%
I have a block towards statistics. I do not understand it and I do not want to understand it.	23%	37%	16%	15%	8%
I prefer qualitative research to a quantitative one.	13%	25%	27%	23%	13%
My supervisor does not understand statistics.	27%	22%	37%	8%	4%
My supervisor required the use of the statistical method in the thesis.	23%	23%	16%	21%	17%
My supervisor suggested me the appropriate method of data processing.	24%	27%	18%	23%	9%

Table 1: Relative frequencies of statements (*n* = 763)

Cluster analysis of statements

Cluster analysis divided respondents into four clusters according to their answers to sentiment questions. Centroids were calculated for individual clusters (see Table 2).
Cluster 1 (*n* = 171) includes respondents who assessed a lot of the statements in a neutral manner being unable to judge them. However, they feel no block in terms of statistics. Their supervisor requires statistics as a part of thesis and the respondents assume that their supervisor understands statistics.
Cluster 2 (*n* = 163) groups respondents stating that they do not understand statistics and prefer qualitative research. Statistical software in English represents a problem for them. However, their thesis supervisor did not require statistics when working on the thesis or suggest any statistical methods.
Cluster 3 (*n* = 132) includes respondents who have no problem with statistical software in English. They do not understand statistics as such, do not wish to understand it and feel that they have a mental block in terms of statistics. They prefer qualitative research, although their thesis supervisor requested the use of statistics in the thesis and suggested specific statistical methods.
Cluster 4 (*n* = 228) groups respondents who stated negative sentiment towards all statements. They have no problem with statistical software in English. They understand statistics and prefer quantitative research. They assume that their thesis supervisor understands statistics, although no use of statistical data processing was required in the thesis. These respondents found their respective data processing methods themselves.

Statement / Cluster	1	2	3	4
It is a problem if statistical software is in English.	3	3	2	2
It is a problem that I do not understand statistics.	3	4	4	2
The combination of the fact that the software is in English and that I do not understand statistics represents a problem.	3	4	3	2
I feel a mental block towards statistics. I do not understand it and I do not wish to.	2	3	4	2
I prefer qualitative research to quantitative.	3	4	4	2
My thesis supervisor does not understand statistics.	2	3	3	2
My thesis supervisor required that my thesis includes the use of statistics.	4	2	4	2
My thesis supervisor suggested what statistical methods I should use to process the data.	3	2	4	2

Table 2: Final Cluster Centres

Relation between statements on selected factors

Table 3 shows the means and standard deviations for men and women. Compared to men (mean = 2.12), more women (mean = 2.66) consider as a problem that statistical software is in English (*p*-value < 0.001). The female group of respondents (mean = 3.37) also report having problems with statistics more often compared to the male group (mean = 3.06, *p*-value = 0.007). The combination of English and statistics is also a more significant problem for women (mean = 2.97) than men (mean = 2.48, *p*-value < 0.001). Women evaluate their relationship towards statistics as a block more often (mean = 2.56) then men (mean = 2.20, *p*-value = 0.002). Other differences of statements are not significant at the 5% level - groups of male and female respondents evaluate both these statements equally.

Statements	Men		Women		*p*-value
	Mean	Std. Deviation	Mean	Std. Deviation	
Statistical software in English is a problem.	2.12	1.185	2.66	1.273	<0.001***
The problem is that I do not understand statistics.	3.06	1.245	3.37	1.290	0.007**
The problem is that I do not understand statistics and the software is in English.	2.48	1.310	2.97	1.326	<0.001***
I have a block towards statistics. I do not understand it and I do not want to understand it.	2.20	1.157	2.56	1.266	0.002**
I prefer qualitative research to quantitative one.	2.88	1.221	3.00	1.232	0.302
My supervisor does not understand statistics.	2.38	1.143	2.41	1.111	0.774
My supervisor required use of statistical method in the thesis.	2.83	1.464	2.94	1.424	0.417
My supervisor suggested me appropriate method of data processing.	2.59	1.307	2.67	1.304	0.487

Table 3: Mann-Whitney U test comparing statements by gender (Men *n* = 162 and Women *n* = 601) * = significant at level 0.001, * = significant at level 0.01, * = significant at level 0.05**

Table 4 describes the means and standard deviations of students who completed at least one statistics course and those who did not. Students without any completed course of statistics tend towards qualitative research (mean = 3.21) in comparison with students who participated and completed at least one statistics course (mean = 2.89, *p*-value = 0.010). Those who completed of at least one statistics course are more confident in the supervisor's knowledge of statistics (mean = 2.34) than students with no completed statistics course (mean = 2.58, *p*-value = 0.014). Respondents who participated in some statistics course often pointed out their supervisor's request to use statistical methods in the final thesis (mean = 3.00) compared to the group of respondents without a completed statistics course (mean = 2.66, *p*-value = 0.029).

Statements	Completed at least one statistic course		Not completed any statistic course		*p*-value
	Mean	Std. Deviation	Mean	Std. Deviation	
Statistical software in English is a problem.	2.54	1.289	2.55	1.237	0.915
The problem is that I do not understand statistics.	3.30	1.319	3.34	1.180	0.534
The problem is that I do not understand statistics and the software is in English.	2.86	1.376	2.89	1.231	0.590
I have a block towards statistics. I do not understand it and I do not want to understand it.	2.47	1.283	2.53	1.153	0.496
I prefer qualitative research to a quantitative one.	2.89	1.253	3.21	1.132	0.010**
My supervisor does not understand statistics.	2.34	1.164	2.58	0.965	0.014**
My supervisor required the use of the statistical method in the thesis.	3.00	1.472	2.66	1.281	0.029*
My supervisor suggested me the appropriate method of data processing.	2.68	1.348	2.57	1.161	0.634

Table 4: Mann-Whitney U test comparing 2 groups – the absolvents of at least one statistics course (*n* = 526) and the respondents without completing any statistics course (*n* = 237) * = significant at level 0.001, * = significant at level 0.01, * = significant at level 0.05**

To better understand the effect of teaching statistics with the support of statistical software, students who completed at least one statistics course (*n* = 526) were divided into two groups: respondents who used statistical software in their courses, and respondents who did not use statistical software in their courses. All statements differ between those two groups (see table 5). Students with a completed statistics course supported with statistical software have fewer problems with the English language (*p*-value = 0.005), with understanding statistics (*p*-value < 0.001) and also with the combination of English and statistics (*p*-value < 0.001). These respondents also feel less antipathy towards statistics (*p*-value < 0.001) and prefer quantitative research (*p*-value 0.006). The students who completed at least one statistics course are also more confident in the supervisor's knowledge of statistics (*p*-value = 0.026). Their supervisor required more often the use of statistical methods in the thesis (*p*-value = 0.001) and suggested appropriate methods of data processing (*p*-value = 0.007).

Statements	Without statistical software		With statistical software		p-value
	Mean	Std. Deviation	Mean	Std. Deviation	
Statistical software in English is a problem.	2.86	1.303	2.46	1.273	0.005**
The problem is that I do not understand statistics.	3.75	1.197	3.17	1.328	0.000***
The problem is that I do not understand statistics and the software is in English.	3.31	1.352	2.75	1.357	0.000***
I have a block towards statistics. I do not understand it and I do not want to understand it.	2.88	1.275	2.37	1.268	<0.001***
I prefer qualitative research to quantitative one.	3.21	1.282	2.82	1.233	0.006**
My supervisor does not understand statistics.	2.57	1.197	2.28	1.149	0.026*
My supervisor required use of statistical method in the thesis.	2.59	1.466	3.10	1.456	0.001**
My supervisor suggested me appropriate method of data processing.	2.37	1.344	2.75	1.339	0.007**

Table 5: Mann-Whitney U test comparing 2 groups – absolvents of statistical course using statistical software (*n* = 107) and absolvents of statistical course without statistical software (*n* = 419) * = significant at level 0.001, * = significant at level 0.01, * = significant at level 0.05**

Students' comments on the statistics courses

The final part of the questionnaire gave respondents the opportunity to freely comment on the topic of statistics classes at university and the used software. Content analysis divided the responses into seven categories (Table 6 and 7). Most commonly, students commented on the need for statistics classes, the character of the classes in statistics courses, including the methods and teaching forms used, as well as the teacher's personality.

Most respondents see the *need for statistics* regarding writing the final thesis, arising either from their own needs or the supervisor's demands. The notion occurred several times that to correctly understanding and interpreting expert publications requires at least basic knowledge of statistics. On the other hand, the majority of humanities students stated that they do not need statistics very much and that, in their opinion, it is given too much space in their study.

Students also assessed the general *character of statistics* classes. A part of students perceives statistics classes in a positive way, considering them of good quality and being satisfied with the course. More students expressed their dissatisfaction, however. Complaints were expressed about the insufficient contents of statistics classes. Courses lacked the basics and principles of statistics and students missed the explanation of the process of using individual statistical methods. Completing their courses did not provide the participants with sufficient knowledge for practice. Others mentioned disproportionate demands towards students. Respondents often considered classes too complicated, detailed and often too strict as many ended their study because of statistics. Some students stated that statistics spoils the whole study. Organisational issues represent another commonly mentioned negative, mainly chaotic organization of courses, i.e. the courses are not continual. Students lack the space to master statistics and practice it. Some students also stated that they did not encounter statistics in their studies at all.

Methods and forms of classes constitute another category of statements. Students believe that statistics should be taught in a more entertaining form and simpler way not to deter students in their first contact with the discipline. Respondents consider the current conception of classes for non-mathematicians, or humanities students to be rather unsuitable. Some stated that they would welcome classes with the use of SW to bring the discipline nearer to students as they completed a course with no use of SW whatsoever, or used to a limited extent. If SW was used in a course, respondents criticised classes teaching them to merely click a button to see the result. This brings no practical benefit in their opinion.

In general, respondents consider the approach to statistics problematic when everything is only seen from the theoretical point of view and explained by putting forward definitions with little sense without specific examples. Respondents demand a practical approach, active practice of illustrative examples of various statistical methods with clear connection to real life. The fourth category concerns the *teacher's personality*. Respondents mainly evaluated whether the teacher could explain given issues and the attitude to students. Students often demand teachers to be able to both understand and explain statistics. Teachers often settle for students memorizing statistical procedures without context. They rarely require a true understanding of statistics and an ability to interpret results. Respondents often pointed out that a teacher's attitude towards students should be patient. Some students came across a teacher with the attitude that whoever does not understand should not be there. Teachers assume students' perfect knowledge of secondary-school mathematics even when studying psychology and other humanities. Others criticised their supervisor for not understanding statistics too much, although the supervisor or the study system demands statistics in the final thesis. Moreover, few people are willing to help students with statistics.

The fifth category includes statements concerning the *issues in statistics classes*. Humanities students consider statistics a complicated discipline as they are used to studying in a different way than statistics requires. Students often admitted to problems with understanding what test or method to use in particular situations and identifying it from results or interpreting results. Respondents also mentioned inconsistent use of Czech statistical terminology. Each teacher uses a different one and students have to deal with it, sometimes even in English.

The survey focused on *statistical software*, which constitutes another frequently mentioned area and another category. Respondents tend to believe that statistical software is necessary for statistics classes. Besides the fact that using the software can save students a lot of time, knowing how to use particular software proves necessary for applying knowledge in practice. Students mostly learned to know the Statistica, Excel, R, and SPSS software in classes. Respondents praise the quality of the SPSS tutorials available on Youtube, as well as online courses for the R software. Those who had statistics also complained that they only studied paid software that they could not use after the study.

Trust in the field of statistics comprises the final, relatively narrow category. Some respondents consider statistics a strange field that contradicts human understanding. Moreover, some think that there are as many results as there are experts.

Categories		Codes
NEED FOR STATISTICS	Too little of statistics	I'd appreciate an obligatory selected course of statistics
		Too little statistics, e.g. in comparison with abroad
	Too much of statistics	Statistics is overrated, MS Excel formulas would suffice
		The quantitative approach is too much promoted in humanities
	Reasons for knowledge of statistics	Necessary in the final thesis
		Final thesis supervisor demands it
		Understanding and interpreting research papers correctly
		Suitable tool for practising logical thinking
THE CHARACTER OF STATISTICS CLASSES	Satisfaction with the course	Quality classes
		Satisfied, the course was beneficial
	Course unsatisfactory in contents	Missing the explanation of statistics fundamentals
		The process of using individual methods was not clarified
		Low quality of classes
	Disproportionate demands on students	Complicated and detailed classes
		Disproportionate time investment into learning statistics
		Strict – many students ended because of it; it's frustrating
	Organisation deficiencies	Chaotically arranged classes, not continual
		Did not take a statistics course during study
METHODS AND FORMS OF CLASSES	Too much theory, too little practice	Everything solved theoretically
		Dictating definitions, the pen-and-paper method
		The classes should be more practical
		No specific practical examples
	Ability to captivate	Need for a more fun form of classes
		Need for a fool-proof way of teaching
		Classes unsuitable for non-mathematicians
		Need for teaching statistics with statistical software

Table 6: Categorized responses from free comments with regard to the given issue – part I

Categories		Codes
TEACHER’S PERSONALITY	Ability to interpret statistics	Should both understand and be able to explain
		Should demand understanding not only drilling methods
	Attitude to students	Arrogant attitude
		Need for patience
		Automatically assumed perfect knowledge of secondary school mathematics
	External help	Does not understand statistics either
		Not many can help with statistics
ISSUES IN STATISTICS CLASSES	Personal issues	A student of humanities is used to studying differently
		Unclear when to use which test, method
		Problem interpreting results
	Field issues	Inconsistent statistical terminology
STATISTICAL SOFTWARE	Need for implementing SW to classes	SW is necessary for statistics classes
		Necessary for jobs in analytical positions
		Necessary for applying knowledge in practice
		Would save students’ time
	SW used in classes	Specific examples of SW
		Demand to use free SW
	Recommended SW	Specific examples of SW
		Reasons for recommending
DISTRUST IN THE FIELD	Field characteristics	Strange field
		Contradicts human understanding
	Trust issues	As many results as there are experts
		No trust in statistical results of others

Table 7: Categorized responses from free comments with regard to the given issue – Part II

DISCUSSION

Statistics and its applications find use in all study fields. According to the results, 38% of thesis supervisors demand the use of statistics from their students. It can be assumed, though, that if there is only one statistically oriented course during their study, students are not likely to be able to interpret the results for their theses. This corresponds to the fact that 20% of respondents that used statistics in their thesis did not process the data themselves. They rather leave the processing on a statistician, the supervisor or someone else. Considering that students usually had only one semester of statistics and are not sure about the calculations, it is understandable that they leave data processing to someone else. Moreover, the mentioned 20% of respondents is a lower estimate as not everyone admits to delegating the data processing.

The survey showed that if students encounter statistics courses in the minimum extent, the conception of the courses proves crucial.

The results show that teaching statistics at universities is still implemented by means of traditional teaching methods, which, according to Mustafa (1996), are ineffective. The main problem is that students fail to establish a clear link between statistics and its uses in the real world and ask themselves: What am I going to use this for? Nonetheless, university studies anticipate the use of statistics, mainly for the thesis. Sufficient knowledge of different possibilities of statistical data processing influences the design of a thesis and gives students new possibilities not only for evaluating data but also for creating research tools and their distribution. Before each statistic course, individual faculties should analyse students’ attitudes and the level of their current knowledge of statistics. Only if the needs of specific students are considered, the teaching of statistical courses may be maximally effective. Suggestions from teacher preparation efforts include a stronger focus on teaching and learning statistics and preparing teachers to use e.g. dynamic statistical software tools (e.g. Lee and Hollebrands, 2008, 2011; Pfannkuch and Ben-Zvi, 2011; Pratt, Davies and Connor, 2011).

Moreover, the use of statistics in the current world grows in importance as the development in contemporary society is connected to the data boom. Teaching statistics should reflect this fact. Data and research method knowledge can help graduates not only in the labour market, but also to assess the socioeconomic situation, for example. Since media widely use numbers and statistics as arguments, basic knowledge of statistics and its application represents a necessary skill of a university educated person.

Statistics courses for humanities should be fully adjusted to students to be effective. A student should become the focal point in designing the course (Harris et al., 2007). The qualitative analysis shows that a large part of students finds statistics very distant in its logic. The way of introducing statistics to such students is of the highest importance. Statistics represents a discipline that can induce enthusiasm with correctly chosen methods, or at least show students the way to facilitate their work.

On the other hand, it can also completely dissuade students. Unfortunately, this is often the case due to poorly designed courses in a majority of statistical courses. Current trends in didactics implement various activation methods. However, such methods rarely penetrate statistics courses, according to students' statements. Students also point out that theoretical statistics comprises the prime aspect in their courses. Contrary to the described practice, Şahin, Ökmen and Kılıç (2020) highlight the advantages of student-centred methods and their positive impact on the students' results.

Students' sentiments towards statistics play an important role as well. Students do not trust statistics, which is connected to prejudice. University courses should eliminate such prejudice. Showing students the use of statistics in everyday life and their future career constitutes the main tool. Chew (2007) states that courses should employ good practice examples and proposes a model of creating examples with an adequate cognitive load. However, students complain in the survey about overly abstract theoretical classes and insufficient explanation of statistics fundamentals along with non-conceptual definition memorising. The presented research showed that women have a larger block in terms of statistics than men. They are similarly not comfortable with the combination of statistics and the English language used in various types of statistical software. In addition, these barriers, combined with a persistent gender gap in statistical reasoning (e.g. Martin, Hughes and Fugelsang, 2017), significantly disadvantage women in the study of statistics.

The results further show that respondents with a completed statistics course tend to have a more positive approach to quantitative research methods. Two aspects can be causing this. Either the fact that one completes a statistics course improves one's attitude towards statistics and students are not afraid to use it in practice. Or natural sciences impact the results significantly. Statistics and quantitative methods are more taught in this branch and students with a more positive approach to statistics and mathematics enrol.

Chew (2007) further emphasises that the classroom atmosphere must enable and encourage elaborative processing and examples. Although Fajčíková and Fejfarová (2019) point out that evaluating teachers does not impact the course quality evaluation by university students, the qualitative research part of this paper shows that students who took a statistics course assess the given teacher's personality and approach to the course. Despite respecting the teacher's expertise, the students often miss a teacher's ability to explain statistics comprehensibly. Moreover, many students do not feel good in statistics classes and consider the statistics teacher's behaviour to be arrogant.

It is alarming that in the era when using digital technologies represents a daily occurrence almost one third of the respondents who completed statistical subjects during their university studies did not get familiar with any statistical software, not even with MS Excel. The importance of using statistical software is emphasized by numerous researchers involved in the teaching of statistics (e.g. Velleman and Moore, 1996; Pratt, Davies and Connor, 2011; Davidson et al., 2019). Their conclusions correspond to our findings, i.e. students who participated in a statistics course supported with using statistical software have less problems understanding statistics, using English in statistics and are more aligned with the supervisor's requirements for using statistical methods in their final thesis.

Moreover, Davidson et al. (2019) point to the development of statistical freeware such as JASP. The advantage of free software is undoubtedly that students can use them even after graduation because they do not need to buy a license. Some students pointed this out in the qualitative part of the research as well. However, the results show that the JASP and Jamovi freeware are the least known among Czech students and in lessons used only occasionally (3% and 2%, respectively). Both software brands offer a huge variety of statistical procedures that cover syllabi of basic statistics courses. Poláčková and Jindrová (2010) point out the use of statistical software with available instruction tutorials which can help the less experienced IT users, who usually have more troubles with navigation in a statistical programme. The opinions of some respondents are in line with the research as the respondents prefer statistical software with tutorials available online.

To use the software in lessons proves both practical and motivational. Students are usually afraid of calculations in statistics (Cobb, 1992), as confirmed by the presented results. When using software, the calculations are done by a computer and students can focus on input conditions of statistical methods and on the interpretation of results. Respondents mention that using the software can save a student a lot of time and facilitate work to ones who are not so strong in mathematics. This is more practical than being stressed by calculations.

Moreover, a student is more focused on using statistics in the given field and its practical significance (Hsu, Wang and Chiu, 2009). However, students complained about the problem of interpreting results in the qualitative part of the presented research. This illustrates the impossibility of teaching with software without a thorough course concept.

Any teaching aid may be successful only if the teacher realises the advantages of its implementation and identifies with them (Fullan and Stiegelbauer, 1991). Therefore, teachers should be convinced of the advantages and suitability of the use of the given statistical software brand. The results above demonstrate that statistical software represents a tool for students to help overcome the initial fear of discipline and facilitate their work. These aspects represent a strong argument for including statistical software in classes and should be considered in statistical course preparation.

The cluster analysis grouped respondents into four clusters according to their responses to the sentiment questions about using statistics. Respondents in each cluster exhibit different potential. They express different needs in terms of statistics classes and the use of statistical software in the courses. Cluster 1 members may be characterised as undecided, cluster 2 respondents as eternal statistics adversaries, cluster 3 members as temporary statistics adversaries, and cluster 4 as statistics fans. The group of the *undecided respondents* includes respondents with a high probability to get convinced of the benefits of statistics. They feel no mental block against statistics and can admit the importance of the field. They are not completely convinced and adhere to neutral sentiment in their statements. Suitably built classes would draw statistics nearer to Cluster 1 members. As Hsu, Wang and Chiu (2009) described teaching about statistical software might play a significant role in the process. However, the cluster 2 of *eternal statistics adversaries* gives the impression that even completing a customised statistics course would not convince them of the advantages. They do not comprehend statistics, never needed it and never will. They probably see no benefit in statistics. Therefore, working with statistical software might represent a further burden and turn-off for them. Griffith et al. (2012) provided a similar description of people who view statistics negatively. The group considers the difficulty and complexity to be the largest issue of statistics. Besides, they described its non-use in future careers and stated that they dislike mathematics as well (Griffith et al., 2012).

The so-called *temporary statistics adversaries* in Cluster 3 do not incline towards statistics and will never pursue it on their own. As they may have come across statistics due to their thesis supervisors' requirements, they may even admit the importance of statistics. A suitably prepared statistics course would help the group to get rid of the mental block against the field. Since they state that they do not wish to understand statistics, working with statistical software might provide an opening for using statistics. The fact that the software does a lot of the work for them might convince them. They do not need to deal with the "tyranny of the computable" as Cobb (2007) describes the need to understand necessary calculations. This group might also be persuaded by real-life data use in classes. Gould (2010) claims that a modern statistics course must adjust to the fact that students' first exposure to data occurs outside academia – in the newspaper, on the internet etc. The classes should use students' knowledge of some diagrams and basic statistical terms and use their own experience as a starting point, describing the statistics surrounding us and showing them its practical use. Holmes (2003) supports this idea by stating that statistics classes should consider a student's real-life and world perception.

Cluster 4 includes so-called statistics fans, i.e. people who have some command of statistics and understand its importance. This group shows the potential to develop their statistical literacy, e.g. by using statistical software in classes. This might broaden their knowledge and skills through new statistical tests and better assessment of the use of individual tests.

Considering the curriculum, its range and conception of a potential statistics course prove relevant for all clusters. Each group exhibits different needs. Group 2 (eternal adversaries) will probably perceive even the introduction to statistical methods as a necessary evil. The content of the course will be crucial for the other three groups, however. Ideally, completing such a course should overlap into initial research phases, i.e. the research design, compiling the research tool, selecting suitable methods etc. (Justice, Zieffler and Garfield, 2017; Holmes, 2003). Deciding on the statistical software, working with it and assessing its use potential plays an essential role in all these groups, although the role differs in each group (see above).

The main limitation of this research is a wide range of study fields of respondents, as was stated in part of Methodology and research questions. However, this research serves mainly as the first probe to point out the problem. The authors also plan to do a more detailed examination on a smaller sample of respondents.

CONCLUSION

The paper's main aim is to determine the approach to teaching statistics at universities and to give an overview of the mainly used didactic materials and teaching methods. To understand the results thoroughly, students were asked to comment freely on the statistics courses and the used statistical software.

Although current teaching practice works with an array of activation methods, the survey results indicate that these methods have not penetrated statistics courses at Czech universities. The design of statistics courses remains rigid and does not reflect current needs of students. Theoretical presentation of definitions prevails over specific examples and practical verification of statistical methods. Thus, students have virtually no opportunity to master individual statistical operations. In the context of such inconsistent statistics teaching practice, any attempt to bring the discipline nearer to students and facilitate their work is desired. Using statistical software in classes, ideally, such software that remains available to students after completing the course represents a possible solution.

The results can be an impulse for improvement not only of teaching statistics but also of teachers' approach towards students. The importance of quantitative research increases as well as the importance of the question of how to teach the current generation necessary statistical thinking. Traditional teaching methods are not effective enough in the present dynamic world and teaching requires modernisation and an innovative approach. The results show that these changes cannot be done without statistical software and teachers open to changes. In the future, the authors plan a more detailed analysis, which will offer specific solutions for various fields of study regarding the cognitive load theory.

ACKNOWLEDGEMENT

This work has been supported by Charles University Research Centre No. UNCE/HUM/024.

REFERENCES

Aberson, C. L., Berger, D. E., Healy, M. R. and Romero, V. L. (2002) 'An Interactive Tutorial for Teaching Statistical Power', *Journal of Statistics Education*, Vol. 10, No. 3. https://doi.org/10.1080/10691898.2002.11910682

Billig, S. H. and Waterman, A. S. (ed.) (2014) *Studying service-learning: Innovations in education research methodology*, New York, NY: Routledge.

Chew, S. L. (2007) 'Designing effective examples and problems for teaching statistics', in Dunn, D. S., Smith, R. A. and Beins, B. (ed.) *Best practices for teaching statistics and research methods in the behavioral sciences*, pp. 73-91, Mahwah, NJ: Erlbaum.

Cobb, G. (1992) 'Teaching statistics', in Steen L. A. (ed.) *Heeding the call for change: Suggestions for curricular action*, Vol. 22, pp. 3–43, Washington, DC: Mathematical Association of America.

Cobb, G. W. (2007) 'The introductory statistics course: A Ptolemaic curriculum?', *Technology innovations in statistics education*, Vol. 1, No. 1. https://doi.org/10.5070/T511000028

Curtin, R., Presser, S., and Singer, E. (2000) 'The effects of response rate changes on the index of consumer sentiment', *Public Opinion Quarterly*, Vol. 64, No. 4, pp. 413–428. https://doi.org/10.1086/318638

Davidson, H., Jabbari, Y., Patton, H., O'Hagan, F., Peters, K., and Cribbie, R. (2019) 'Statistical Software Use in Canadian University Courses: Current Trends and Future Directions', *Teaching of Psychology*, Vol. 46, No. 3, pp. 246–250. https://doi.org/10.1177/0098628319853940

De Leeuw, E. D. and Hox, J. J. (2008) 'Self-administered questionnaires: mail surveys and other applications', in De Leeuw, E. D., Hox, J. J. and Dillman, D. (ed.) *International handbook of survey methodology, pp. 239–263,* New York, NY: Routledge. https://doi.org/10.4324/9780203843123

Dempster, M. and McCorry, N. K. (2009) 'The Role of Previous Experience and Attitudes Toward Statistics in Statistics Assessment Outcomes among Undergraduate Psychology Students', *Journal of Statistics Education*, Vol. 17, No. 2. https://doi.org/10.1080/10691898.2009.11889515

Erlingsson, Ch. and Rysiewicz, P. (2017) 'A hands-on guide to doing content analysis', *African Journal of Emergency Medicine*, Vol. 7, No. 3, pp. 93–99. https://doi.org/10.1016/j.afjem.2017.08.001

Fajčíková, A. and Fejfarová, M. (2019) 'Evaluation of the Quality of Teaching from the Perspective of University Students', *Journal on Efficiency and Responsibility in Education and Science*, Vol. 12, No. 2, pp. 34–40. https://doi.org/10.7160/eriesj.2019.120201

Fullan, M. G. and Stiegelbauer, S. (1991) *The New Meaning of Educational Change*, New York: Teachers College Press.

Gould, R. (2010) *'Statistics and the modern student', International Statistical Review*, Vol. 78, No. 2, pp. 297–315. http://dx.doi.org/10.1111/j.1751-5823.2010.00117.x

Griffith, J. D., Adams, L. T., Gu, L. L., Hart, C. L. and Nichols-Whitehead, P. (2012) 'Students' attitudes toward statistics across the disciplines: a mixed-methods approach', *Statistics Education Research Journal*, Vol. 11, No. 2, pp. 45–56. https://doi.org/10.52041/serj.v11i2.328

Harris, C. M., Mazoué, J. G., Hamdan, H., and Casiple, A. R. (2007) 'Designing an online introductory statistics course', in Dunn, D. S., Smith, R. A. and Beins, B. (ed.) *Best practices for teaching statistics and research methods in the behavioral sciences*, pp. 93–108. Mahwah, NJ: Erlbaum.

Hendl, J. (2005) Kvalitativní výzkum: základní metody a aplikace [Qualitative Research. Basic Methods and Applications], Prague: Portál.

Hindls, R., and Hronová, S. (2005) 'Jak výuka odrazuje nestatistiky od statistiky' [How teaching discourages non-statisticians from statistics], *Statistika*, Vol. 42, No. 2, pp. 168–172.

Holmes, P. (2003) '50 Years of Statistics Teaching in English Schools: Some Milestones', *Journal of the Royal Statistical Society: Series D (The Statistician)*, Vol. 52, No. 4, pp. 439–463. https://doi.org/10.1046/j.1467-9884.2003.372_1.x

Hsu, M. K., Wang, S. W. and Chiu, K. K. (2009) 'Computer attitude, statistics anxiety and self-efficacy on statistical software adoption behavior: An empirical study of online MBA learners', *Computers in Human Behavior*, Vol. 25, No. 2, pp. 412–420. https://doi.org/10.1016/j.chb.2008.10.003

Huynh, M. and Baglin, J. (2017) 'Teaching statistics through data investigations in Australian secondary schools: An island-based pilot project', *International Journal of Innovation in Science and Mathematics Education*, Vol. 25, No. 1, pp. 49–63.

Hybšová, A. (2017) *Statistická gramotnost studentů učitelství přírodopisu/biologie v České republice [Biology pre-service teacher's statistical literacy in the Czech republic]*, [PhD thesis], Prague: PedF UK.

Justice, N., Zieffler, A., and Garfield, J. (2017) 'Statistics Graduate Teaching Assistants'beliefs, Practices And Preparation For Teaching Introductory Statistics', *Statistics Education Research Journal*, Vol. 16, No. 1, pp. 294–319.

Kurniawan, D. A., Astalini, A., Darmaji, D., and Melsayanti, R. (2019) 'Students' Attitude towards Natural Sciences', *International Journal of Evaluation and Research in Education*, Vol. 8, No. 3, pp. 455–460. http://doi.org/10.11591/ijere.v8i3.16395

Law, K. M. Y., Lee, V. C. S. and Yu, Y. T. (2010) 'Learning motivation in e-learning facilitated computer programming courses', *Computers & Education*, Vol. 55, No. 1, pp. 218–228. https://doi.org/10.1016/j.compedu.2010.01.007

Lee, R. and Fielding, N. (2004) 'Tools for Qualitative Data Analysis', in Hardy, M. and Bryman, A. (ed.) *Handbook of Data Analysis*, London: Sage, pp. 529–546. https://dx.doi.org/10.4135/9781848608184.n23

Lee, H. S. and Hollebrands, K. (2008) 'Preparing to teach mathematics with technology: An integrated approach to developing technological pedagogical content knowledge', *Contemporary Issues in Technology and Teacher Education*, Vol. 8, No. 4., pp. 326–341.

Lee, H. S. and Hollebrands, K. F. (2011) 'Characterizing and developing teachers' knowledge for teaching statistics', in Batanero, C., Burrill, G. and Reading, Ch. (ed.) *Teaching statistics in school mathematics-Challenges for teaching and teacher education: A joint ICMI/IASE study*, Dordrecht: Springer, pp. 359–369. https://doi.org/10.1007/978-94-007-1131-0

Martin, N., Hughes, J. and Fugelsang, J. (2017) 'The roles of experience, gender, and individual differences in statistical reasoning', *Statistics Education Research Journal*, Vol. 16, No. 2, pp. 454–475.

Mazouchová, A., Jedličková, T. and Hlaváčová, L. (2020) 'Using statistical software at Czech universities', *Proceedings of the 17th International Conference Efficiency and Responsibility in Education (ERIE 2020)*. Prague, pp.181–188.

Moore, D. L., and Tarnai, J. (2002) 'Evaluating nonresponse error in mail surveys', in Groves, R. M., Dillman, D. A., Eltinge, J. L., and Little, R. J. A. (ed.) *Survey Nonresponse*, New York, NY: John Wiley & Sons, pp. 197–211.

Mustafa, R. Y. (1996) 'The challenge of teaching statistics to non-specialists', *Journal of statistics education*, Vol. 4, No. 1. https://doi.org/10.1080/10691898.1996.11910504

Pfannkuch, M. and Ben-Zvi, D. (2011) 'Developing teachers' statistical thinking', in Batanero, C., Burrill, G. and Reading, Ch. (ed.) *Teaching statistics in school mathematics-Challenges for teaching and teacher education: A joint ICMI/IASE study,* Dordrecht: Springer, pp. 323–333. http://dx.doi.org/10.1007/978-94-007-1131-0_31

Poláčková, J. and Jindrová, A. (2010) 'Innovative Approach to Education and Teaching of Statistics', *Journal on Efficiency and Responsibility in Education and Science*, Vol. 3, No. 1, pp. 14–27.

Pratt, D., Davies, N., and Connor, D. (2011) 'The role of technology in teaching and learning statistics', in Batanero, C., Burrill, G. and Reading, Ch. (ed.) *Teaching statistics in school mathematics-Challenges for teaching and teacher education: A joint ICMI/IASE study*, Dordrecht: Springer, pp. 97–107. http://dx.doi.org/10.1007/978-94-007-1131-0_13

Ralston, K., MacInnes, J., Crow, G., and Gayle, V. J. (2016) *We need to talk about statistical anxiety: A review of the evidence around statistical anxiety in the context of quantitative methods pedagogy*, London: NCRM.

Şahin Ş., Ökmen B. and Kılıç A. (2020) 'Effects of Teaching the Learning Psychology Course in Different Ways on the Student's Success and Attitudes', *Journal on Efficiency and Responsibility in Education and Science*, Vol. 13, No. 3, pp. 113–129. http://dx.doi.org/10.7160/eriesj.2020.130302

Singer, E., van Hoewyk, J., & Maher, M. P. (2000) 'Experiments with incentives in telephone surveys', *Public Opinion Quarterly*, Vol. 64, No. 2, pp. 171–188. http://dx.doi.org/10.1086/317761

Skalská, H. (2019) 'Výuka statistiky pro informatiku a management v éře datové vědy' [Statistics education for informatics and management in the era of data science], *Informační bulletin České statistické společnosti.* Vol. 30, No. 4, pp. 1–9.

Steinberger, P. (2020) 'Assessing the Statistical Anxiety Rating Scale as applied to prospective teachers in an Israeli Teacher-Training College', *Studies in Educational -Evaluation*, Vol. 64, 100829. https://doi.org/10.1016/j.stueduc.2019.100829

Stejskalová I., Komárková L., Bednářová M. and Štrach P. (2019) 'Student Adoption of a Non-Traditional Teaching Method in Accounting: How Previous Experience Impedes Willingness to Change', *Journal on Efficiency and Responsibility in Education and Science*, Vol. 12, No. 1, pp. 1–11. http://dx.doi.org/10.7160/eriesj.2019.120101

Strauss, A. L. and Corbin, J. (1999) *Základy kvalitativního výzkumu: postupy a techniky metody zakotvené teorie* [Basics of Qualitative Research. Techniques a Procedures for Developing Grounded Theory], Brno: Sdružení Podané ruce.

Švaříček, R. and Šeďová, K. (2007) *Kvalitativní výzkum v pedagogických vědách* [Qualitative research in pedagogical sciences], Prague: Portál.

Velleman, P. F., and Moore, D. S. (1996) 'Multimedia for teaching statistics: Promises and pitfalls', *The American Statistician*, Vol. 50, No. 3, pp. 217–225. https://doi.org/10.2307/2684658

Widenská, E. (2014) 'Efficiency of Practicing with Materials Using ICT and Paper Ones in Mathematics', *Journal on Efficiency and Responsibility in Education and Science*, Vol. 7, No. 2, pp. 37–43. https://doi.org/10.7160/eriesj.2014.070203

Full research paper

A JOINT ASSESSMENT OF REASONING ABOUT GENERAL STATEMENTS IN MATHEMATICS AND BIOLOGY

Libuše Samková[1]✉
Lukáš Rokos[2]
Lukáš Vízek[3]

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

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

[3]Department of Mathematics, Faculty of Science, University of Hradec Králové, Czech Republic

✉ lsamkova@pf.jcu.cz

ABSTRACT

This contribution belongs to a larger empirical study that focuses on issues related to the implementation of inquiry-based learning and formative assessment in science and mathematics education, while it also refers to the issue of STEM education. Here, we discuss the two topics from the perspective of professional preparation of primary school teachers. We employ an educational tool called Concept Cartoons and perceive it as a common diagnostic tool for investigating modes of reasoning about general statements in arithmetic, geometry and biology. The presented qualitative exploratory empirical study maps and codes various kinds of reasoning that can be identified with the tool and investigates possibilities of a joint coding procedure. As a result, it provides a conversion table between various modes of reasoning in the three subject domains. The arisen code categories cover the field of generic examples, including the initial stages so that they can be used for scaffolding the process of learning the foundations of deductive reasoning. The joint approach to reasoning in mathematics and biology shows how argumentation and formative assessment can be understood equally and developed simultaneously in both school subjects. It helps us to see how the two school subjects can be integrated didactically.

KEYWORDS

Argumentation, biology education, Concept Cartoons, formative assessment, future primary school teachers, mathematics education

HOW TO CITE

Samková L., Rokos L., Vízek L. (2021) 'A Joint Assessment of Reasoning about General Statements in Mathematics and Biology', *Journal on Efficiency and Responsibility in Education and Science*, vol. 14, no. 4, pp. 270-287. http://dx.doi.org/10.7160/eriesj.2021.140406

Article history
Received
June 4, 2021
Received in revised form
August 3, 2021
Accepted
September 1, 2021
Available on-line
December 20, 2021

Highlights

- *Establishing a framework for joint assessment of reasoning about general statements in arithmetic, geometry and biology, with a conversion table.*
- *Providing a formative assessment background for scaffolding the process of learning the foundations of deductive reasoning.*
- *Promoting opportunities for integrating mathematics and biology education within professional preparation of primary school teachers.*

INTRODUCTION

In recent decades, inquiry-based teaching and formative assessment have belonged among educational frameworks that have been discussed in relation to possible enhancement of the form and outcomes of teaching practices. Inquiry-based teaching has been considered as a means of providing a sustainable connection between up-to-date scientific research and everyday school practice (McComas, 2002; Minner, Levy and Century, 2010), and formative assessment has proved its ability to provide a suitable evaluative approach to student outcomes and development of student learning during inquiry-based school activities (Dolin and Evans, 2018). Due to some particular timing of the initial recent interest in inquiry-based pedagogy (the first US and European curricular documents involving inquiry referred only to the school subject of science: National Research Council, 1996; Rocard et al., 2007), the inquiry has often been perceived solely in relation to science teaching and learning. However, the inquiry can be implemented in various school subjects including mathematics (Artigue and Blomhøj, 2013; Dorier and Maass, 2014). In our paper, we present an approach that goes beyond the initial framework of science education and perceives inquiry in a comparable manner both in science and mathematics school subjects.

In recent years, we have been analysing an educational tool called *Concept Cartoons* (Keogh and Naylor, 1999) and its possible use in the professional preparation of future teachers, especially in investigating various aspects of teacher knowledge in mathematics. We focused on Concept Cartoons as a tool for assessing reasoning on arithmetic topics (e.g., Samková and Tichá, 2017), in qualitative as well as mixed research designs (Samková, 2019). To systematically map the terrain, we naturally proceeded to a question whether and how it is possible to use Concept Cartoons to assess reasoning about mathematical statements. Combined with the aim to provide a common approach to professional preparation of future teachers that would address both science and mathematics education, we raised a question whether and how it is possible to do the assessment of reasoning jointly in science and mathematics subjects.

Such an arrangement resulted in an exploratory qualitative empirical study that is presented in this paper. As the main participants in the study, we involved future primary school teachers. We draw on the fact that primary school teachers are generalists in our country (i.e., are intended to teach both science and mathematics). Thus, we can pay comparable attention to their argumentation skills in both school subjects. To obtain results that would be applicable in professional preparation of primary school teachers at our university, we proceeded from the structure of our five-year professional preparation program that in its first three years approaches the study of content and didactics in mathematics and biology simultaneously. Other science subject domains (physics, chemistry) are approached later in the program and to a much lesser extent. Thus, we choose biology as the domain of our interest in science.

The presented research study is of an iterative nature. It started as a study that was conducted by the first author and focused on modes of argumentation related to reasoning about general statements in arithmetic (Samková, 2020); for the rest of the paper, we will call this study *an arithmetic study*. Following the findings of the arithmetic study, the first author contacted the second author and together, we started an *arithmetic-biology study*. We conducted explorations covering general statements not only in arithmetic but also in biology, especially in marine zoology. But the coding process did not lead to a conceptually coherent structure of code categories that would be joint for both subjects. We were faced with the decision to either collect additional data or modify the perspective. At the same time, the first author conducted an independent study with the third author, where we explored reasoning about general statements in geometry (Vízek and Samková, 2021); we will call this study *a geometry study*. The course of the coding process in the geometry study indicated that the geometric perspective might be crucial for gaining the desired coherence in the arithmetic-biology study. Therefore, we combined the arithmetic-biology study with the geometry study to create a *mathematics-biology study* that is the main study of this paper. The arithmetic study, the arithmetic-biology study and the geometry study serve as sub-studies to the main study.

This paper has been developed as an extension of conference contributions that introduced separately the arithmetic study (Samková, 2020) and the geometry study (Vízek and Samková, 2021). It brings the two mathematical sub-studies into a common context, adds a biological dimension to the focus, and provides a conversion table between modes of reasoning about general statements in arithmetic, geometry and biology.

The text is organised as follows: at the beginning, it presents the background of the presented research, the three sub-studies and the main study. Then it discusses the findings and captures the implications for further research.

Argumentation and general statements in mathematics

When mathematics learners develop, their view of a proper response to questions whether and why a mathematical statement is true develops as well. The first attempts of such a response usually have the form of a *reference to an authority* (e.g., "It is true because you told us yesterday."), later on to one or more *confirming examples* – particular cases in which the statement holds. The confirming examples are welcome when the statement is existential (because they confirm the desired existence) but not so welcome when the statement is general. In the latter case, the disadvantage of confirming examples is that sometimes they can be easily used to mistakenly show the veracity of a false statement. For instance, the general statement that multiplying one positive number by another positive number always produces a bigger number looks true when you multiply 2 by 3, or 5 by 2.

Quite a different role in argumentation is played by *counter-examples* – particular cases in which the statement does not hold although they meet all the prerequisites of the statement. Such counter-examples are welcome when the statement is general since one counter-example is enough to disprove the statement. For instance, the statement that multiplying one positive number by another positive number always produces a bigger number may be disproved by a counter-example consisting in multiplying 2 by 1.

Various types of argumentation in assessing and proving mathematical statements were systematised by Harel and Sowder (1998) under the name of *proof schemes*. In their interpretation, 'a person's proof scheme consists of what constitutes ascertaining and persuading for that person' (ibid: 244). Their multi-level classification of proof schemes includes:

- *external conviction proof schemes* (doubts are removed by the ritual of the argument presentation, by the word of authority, or by the symbolic form of an argument);
- *empirical proof schemes* (doubts are removed by quantitatively evaluating the conjecture by one or more specific cases – *inductive proof scheme*, or by reasoning from an illustration or a geometric figure, regardless of possible transformations – *perceptual proof scheme*);
- *analytical proof schemes* (doubts are removed by means of logical deductions: *transformational proof schemes* are based on operations on objects and anticipations of results of the operations, *axiomatic proof schemes* are based on some prior results, axioms and definitions).

In some later sources (e.g., Harel and Sowder, 2007), the name of *deductive proof schemes* is used instead of *analytical proof schemes*, to highlight the role that deductive reasoning plays in this type of proof schemes: 'Deductive reasoning is a mode of thought commonly characterized as a sequence of propositions where one must accept any of the propositions to be true if he or she has accepted the truth of those that preceded it in the sequence.' (ibid: 811). Harel and Sowder also relate their proof schemes to other taxonomies, e.g., the taxonomy given by Balacheff (1988), and refer to a *generic example* as a sample of a transformational proof scheme. In their interpretation, a generic example is a 'justification by an example representing salient characteristics of a whole class of cases' (Harel and Sowder, 2007: 810).

Primary school teachers and their reasoning about general statements in mathematics

Although Harel and Sowder's research leading to the typology of proof schemes in mathematics was conducted generally, with college students as respondents and with special attention paid to mathematics major students, the typology might be applied to any stage of mathematical education, including the primary school level. Axiomatic proof schemes might be too formal for the primary school level, but all the other types of proof schemes might naturally appear there (Komatsu, 2010).
A typology that might be considered an alternative to the Harel and Sowder's system of proof schemes and is more apposite for the case of future primary school teachers, was introduced by Simon and Blume (1996). According to data collected from future elementary school teachers during a mathematics content course, Simon and Blume propose a list of five levels of responses attempting to justify mathematical statements (ibid: 17):

- 'Level 0 – Responses identifying motivations that do *not* address justification.
- Level 1 – Appeals to external authority.
- Level 2 – Empirical demonstrations.
- Level 3 – Deductive justification that is expressed in terms of a particular instance (generic example).
- Level 4 – Deductive justification that is independent of particular instances.'

Future primary school teachers have often got deeply rooted misconceptions about argumentation and proving: they tend to rely on an external authority as the basis of their conviction (ibid), believe that it is possible to affirm the validity of a generalisation through a few examples, randomly selected examples or big-number examples (Martin and Harel, 1989; Stylianides and Stylianides, 2009). Some of them do not understand the role of counter-examples in refuting general statements and consider one counter-example not being enough (Zazkis and Chernoff, 2008). Similar findings were also reported for secondary school students (Galbraith, 1981) and college students (Selden, 2012). Such a state may be persistent and, later in school practice, may affect how primary school teachers notice essential mathematical reasoning forms when justification and generalisation appear in the classroom (Melhuish, Thanheiser and Guyot, 2020).

Argumentation in science

In science education, argumentation is considered a core skill for the development of students' scientific literacy, critical thinking, and reasoning (Berland and Reiser, 2009; Lazarou, Sutherland and Erduran, 2016). The ability of students to form and understand arguments related to scientific phenomena and the processes behind them is the basis of scientific literacy. The quality of students' argumentation allows the teacher to recognise how the students understand the issue (Cullen et al., 2018).
The use of reasoning in science subjects is a complex and systematic approach that can include a lot of activities (Erduran and Jiménez-Aleixandre, 2007). One of the approaches used in science subjects is the process of evaluating and justifying claims mediated by Concept Cartoons (Naylor, Keogh and Downing, 2007). The arguments provided by students and teachers in science classrooms can support previously discussed knowledge, put it in a new context, and use the creation of various artefacts through which students could support their arguments (Furtak et al., 2010). Scientific reasoning and argumentation had a positive effect on students' achievement (Dofner et al., 2018) and this approach could be used to improve students' scientific knowledge. Ping, Halim and Osman (2020) found that students were able to provide explanations when using the argument, which included evidence supporting their opinion.

The role of general statements in inquiry-based education and formative assessment

Inquiry-based education has been recently perceived as a means of providing a sustainable connection between up-to-date scientific research and everyday school practice (McComas, 2002; Minner, Levy and Century, 2010). Thus, it consists of classroom activities involving students that observe, pose questions, reason, search for information, collaborate, collect data and interpret them, discuss obtained results (Dorier and Maass, 2014). Such an environment is naturally rich in generalisations (Bulková, Medová and Čeretková, 2020) and general statements of various appearances and validity are frequently voiced in the classroom.
As for the teachers, the inquiry-based environment requires teachers' feedback that would support students' learning (Dolin and Evans, 2018). The feedback typically addresses four different aspects of the learning situation: the task, the process to complete the task, student's self-regulation and student's persona (Hattie and Timperley, 2007). In this paper, we focus mainly on the second aspect and its part that relates to the components of the solution procedure already given by the student (e.g., whether the procedure is correct, suitably expressed, explained and interpreted). Namely, we focus on feedback that the teacher provides to the student during the process of generalisation and its interpretation: we plan to investigate the quality of responses that the teacher provides to assess the validity of a general statement given by the student.

MATERIALS AND METHODS

Our study aims to answer two research questions, the first one serving as preparatory for the second one:

RQ1: What kinds of reasoning about general statements in biology and mathematics can be observed in future primary school teachers when using Concept Cartoons as a diagnostic instrument?

RQ2: What are the possibilities of joint assessment of reasoning about general statements in biology and mathematics?

The study is of an exploratory qualitative design since the phenomenon of a possible joint approach to reasoning about general statements in mathematics and biology through Concept Cartoons has not been studied before. To explore and describe the nature of the phenomenon, collected data are analysed qualitatively, using open coding and constant comparison (Miles, Huberman and Saldaña, 2014).

The research is iterative; it consists of three consecutive sub-studies and the main study. The first and third sub-studies focus just on the first research question narrowed to the context of mathematics. The second sub-study and the main study cover both research questions to their full extent.

Participants

The three sub-studies were conducted with participants studying the second or third year of a five-year full-time master degree program for future primary school teachers. This program is mostly frequented by prospective teachers that came to university directly from the upper-secondary school. In the Czech Republic, where the study takes place, primary school teachers are generalists; they are supposed to teach all school subjects at the primary school level (students from 6 to 11 years of age).

The first sub-study had 28 participants (labelled as *group 1*), the second sub-study 49 participants (*group 2*) and the third sub-study had 29 participants (*group 3*; 26 of them being the same as in group 1). In all three sub-studies, data were collected at a compulsory course focusing on mathematical content preparation of future primary school teachers. The participants from groups 1, 2 and 3 are called *standard participants*.

The main study reprocessed data collected during the three sub-studies and, additionally, included nine participants studying the last year of a full-time master degree program for future lower-secondary school teachers (students from 11 to 15 years of age). In the Czech Republic, these teachers are specialists; each of them specialises in two school subjects of their own choice. We included five future lower-secondary school teachers specialising in mathematics (labelled as *group 4*) and five specialising in biology (*group 5*); one of them specialised in both mathematics and biology (i.e., belonged to both groups). All of them were selected from future lower-secondary school teachers working on a diploma thesis in mathematics or biology specialisation and, for the purpose of data collection, contacted individually. The participants from groups 4 and 5 are called *additional participants*.

As mentioned above, the main study worked with data collected from all standard and additional participants, i.e., from 28 + 49 + 29 – 26 + 5 + 5 – 1 = 89 different participants. To identify individual participants, we used codes consisting of a letter referring to a group of participants (S for groups 1 and 3, R for group 2, D for groups 4 and 5)[1] and a randomly assigned number. For instance, R29 identifies a participant from group 2.

Diagnostic instrument – common aspects

As a diagnostic instrument, we used *Concept Cartoons*. Concept Cartoons had been initially developed by Keogh and Naylor (1999) as a tool to motivate and support the learning of students during elementary science lessons. Later, they were also implemented in other school subjects, including mathematics (Dabell, Keogh and Naylor, 2008). Each Concept Cartoon has a form of a picture with a school or out-of-school everyday situation related to a given curricular content and several children discussing the situation through a bubble dialogue. Our previous research showed how Concept Cartoons might be used as a tool for diagnosing various types of teacher knowledge (e.g., Samková, 2019).

In this study, we employ three Concept Cartoons related to the topic of general statements, one located in arithmetic, one in biology and one in geometry. Although the original sets of Concept Cartoons exist in mathematics (Dabell, Keogh and Naylor, 2008) as well as in science (Naylor and Keogh, 2010), none of the original Concept Cartoons suited precisely our purpose. We were looking for Concept Cartoons based on widely spread misconceptions related to Czech primary school curriculum that could be presented in the form of general statements. Moreover, we needed the statements to allow specifying many confirming examples as well as counter-examples. In the bubbles, we expected other statements that might serve as hints to the presented issue. Among the original Concept Cartoons, we did not find any with these attributes. Therefore, the arithmetic Concept Cartoon for our study was created by a modification of the content of all bubbles in an original Concept Cartoon (Dabell, Keogh and Naylor, 2008: 3.2), and the biology and geometric Concept Cartoons arose independently of the original sets. The three Concept Cartoons and the circumstances of their origin will be introduced in detail within the description of the sub-studies.

THE COURSE OF THE RESEARCH

Data collection and data analysis – common aspects

During all three sub-studies and the main study, data collection always consisted of assigning a worksheet with the diagnostic Concept Cartoon(s). Then we asked the respondents to respond to (each of) the pictures in a written form: to decide which children in the picture are right and which are wrong and justify the decision. The participants worked on the worksheet individually.

During data analysis, we first registered which bubbles were chosen by individual respondents as correct and under which additional conditions. Afterwards, we openly coded all the material (the written responses given by the respondents as well as the diagnostic Concept Cartoons), looking for various

1 The distribution of reference letters follows external rules that the first author has for systematic handling of all her data. This system enables to differentiate between different study groups identified by university, study program, and year of commencement. Groups 1 and 3 in this paper are subsets of the same study group, thus the same letter for identification of their members.

aspects related to reasoning, justification and argumentation in mathematics and science. Then we applied the method of constant comparison – from the overall perspective, from the perspective of individual bubbles across all participants and from the perspective of individual participants across all bubbles. At the end of the process, the aim was to have assigned exactly one code category to each of the participants and each of the bubbles included in the research. For instance, in the first sub-study, we worked with 28 participants and data analysis included four bubbles of the arithmetic Concept Cartoon. Therefore, we intended to finish the analytic process with exactly $28 \cdot 4 = 112$ assignments.[2]

From the perspective of the general scheme of qualitative data analysis shown in Figure 1 (the diagram starts at the top), the situation with assignments of code categories turned out differently in different sub-studies. During the first and third sub-studies, we successfully ran several rounds of the constant comparison process (the middle-central part of the diagram) and finished the analytic process by providing the intended one-to-one assignments that led to establishing a conceptually coherent structure of codes (the lower-right corner of the diagram). During the second sub-study, the constant comparison process was not successful and we ended in the lower-left corner of the diagram. The idea of combining all three sub-studies together to create a basis for the main study enriched the second sub-study by desired additional data. It enabled to continue the analytic process (i.e., to return to the top part of the diagram). The detailed overall scheme of data handling and types of assigned code categories during individual stages of the research is shown in Figure 2.

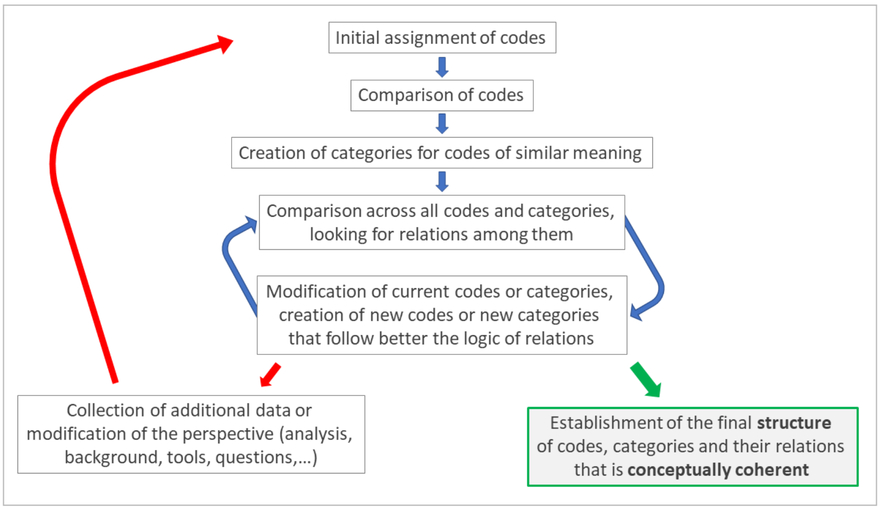


Figure 1: The general scheme of qualitative data analysis (source: own illustration)

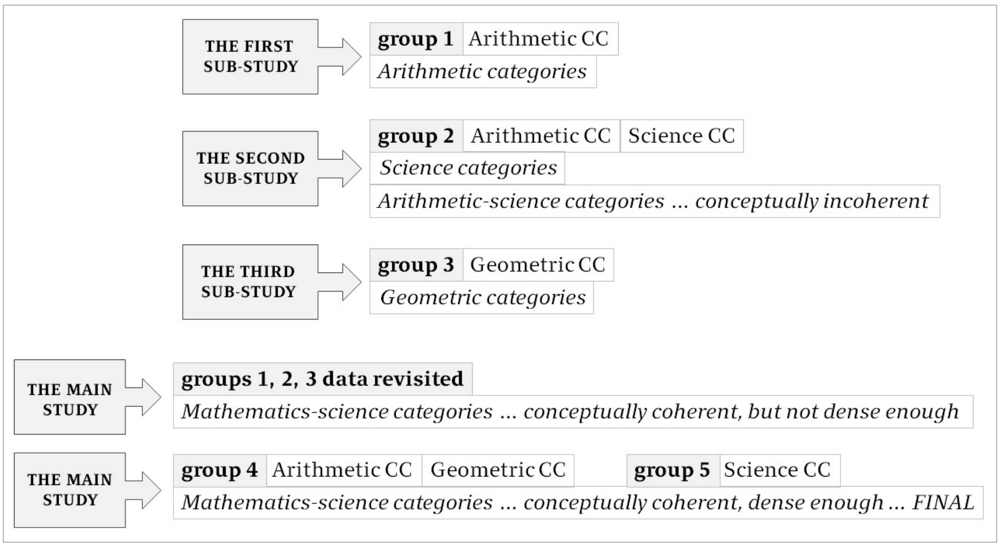


Figure 2: The overall scheme of data handling and types of assigned code categories during individual stages of the research, CC = Concept Cartoon (abbr.)

2 In the Czech Republic, where the research study takes place, the proper symbol for multiplication in school mathematics is given as " · ", i.e., a middle dot. To keep the authenticity of data excerpts presented in this paper and to maintain the coherence of the symbolic notations throughout the paper, we use this symbol also for multiplication that appears in the texts outside data excerpts.

THE ARITHMETIC SUB-STUDY (THE FIRST SUB-STUDY)

The arithmetic sub-study focused on argumentation related to general statements in arithmetic, namely to statements about the properties of results of the multiplication operation. We proceeded from the fact that these properties are different within the domains of natural numbers, rational numbers and integers. Thus, such an environment has great potential for exploring various confirming examples, counter-examples and conditions of validity.

Diagnostic instrument – particularities

In the Concept Cartoon, we presented an informal description of the operation of multiplication and a bubble dialogue comprising of four general statements on multiplication based on four common misconceptions about the operation: "multiplication makes a bigger number", "multiplication by a fraction makes a smaller number", "multiplication by a negative number makes a smaller number" and "multiplication by zero makes a smaller number" (Figure 3). The domain set of the numbers in focus is not specified in the picture intentionally; it might equal the set of natural numbers as well as integers, rational numbers (fractions, decimal numbers) or real numbers.

The misconceptions hidden behind individual bubbles are partially related to critical moments in learning when an earlier way of thinking fails to account sufficiently for new ideas (cf. Confrey and Kazak, 2006); they refer respectively to

- multiplication as a repeated addition of a positive integer which gives a result that is bigger than the repeated number;
- multiplication of a positive integer by a positive proper fraction which gives a result that is smaller than the positive integer;
- multiplication of a positive integer by a negative integer which gives a negative result that is smaller than the positive integer;
- multiplication of a positive integer by zero, which gives a zero result that is smaller than the positive integer.

All misconceptions more or less relate to *natural number bias* (Alibali and Sidney, 2015; van Dooren, Lehtinen and Verschaffel, 2015), a tendency to use considerations and procedures learned in the domain of natural numbers and transfer them unjustifiably to the domain of integers or rational numbers. One of the misconceptions combines natural number bias with a similar tendency related to the transfer from the domain of proper fractions to the domain of all fractions (Stevens et al., 2020).

As for the veracity of the statements in bubbles, none of them is correct since for each of them there exist counter-examples (numbers for which the statement is not valid). However, it is possible to specify under what conditions the statements are valid:

- Jan: for all pairs of positive numbers bigger than 1, for all pairs of negative numbers, for positive numbers multiplied by positive numbers bigger than 1, for negative numbers multiplied by zero, for negative numbers multiplied by positive numbers smaller than 1;
- Emil, Bara: for all positive numbers;
- Tom: for positive numbers multiplied by positive fractions smaller than 1, for negative numbers multiplied by positive fractions bigger than 1, for positive numbers multiplied by negative fractions.

The most general statement is the statement in the upper-left bubble (Jan). It allows variations in multiplier as well as multiplicand, and none of them is specified in the bubble. The other three statements might be considered hints to Jan's bubble, since they point indirectly out some of the types of multipliers for which the most general statement might not be valid: number zero (Emil), negative numbers (Bara), fractions (Tom).

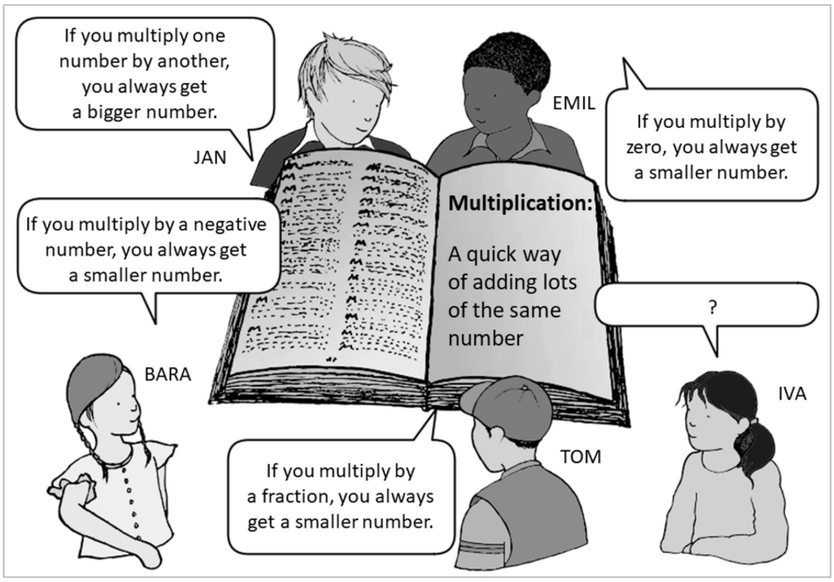


Figure 3: A Concept Cartoon on multiplication; (source of the text in the book, the template of the book and the template of children with empty bubbles: Dabell, Keogh and Naylor, 2008: 3.2)

Electronic ISSN 1803-1617 Printed ISSN 2336-2375

Results of the arithmetic sub-study

During the data analysis process, 28 · 4 = 112 assignments were made and eight different code categories appeared as a result of the analytic process. Each of the categories was labelled by an acronym. The list of all relevant acronyms with their short description and frequency in data is shown in Table 1. In the further text, the meaning of the acronyms is explained and accompanied by data excerpts, in the same order as in Table 1.

acronym	description	bubbles				
		Jan	Emil	Bara	Tom	all
CEX	one or more counter-examples	14	5	6	8	33
GE	generic examples, no justification	1	2	6	1	10
GED	generic examples, with justification	0	1	2	2	5
COX	conditions of validity indicated, no justification	7	2	4	4	17
COD	conditions of validity indicated, with justification	3	5	6	3	17
OF	over-fixation to previous learning	0	5	0	3	8
VR	vague response	3	8	4	5	20
XR	no response	0	0	0	2	2
altogether		28	28	28	28	112

Table 1: List of relevant acronyms for arithmetic code categories and their frequency in data, *n* = 28, 2016 (source: own calculation)

According to the results of the analysis, most of the participants based their reasoning just on one or more counter-examples, without any additional reasoning (code category CEX):

S32 Jan: No! 1 · 358 = 358 → stays the same.

Several of the counter-examples were accompanied by a commentary that linked them to a wider group of numbers with the same behaviour within the given context, i.e., these counter-examples might be considered as generic examples (code category GE):

S33 Bara: She considered neither zero nor negative numbers
(−8) · 0 = 0 is not smaller
(−3) · (−3) = 9 is not smaller

Other responses consisted in counter-examples that were generic examples accompanied by other deductive arguments (code category GED), more or less correct:

S29 Bara: When I multiply a negative number by a negative number, I get a number that is positive, so bigger, −3 · (−3) = 9. When I multiply zero by a negative number, I get zero – the same number.

Some participants did not apply counter-examples, instead they just indicated conditions under which the given statement holds, with a proper deductive justification (code category COD) or without (code category COX). Some of them were able to cover properly the conditions for Emil and Bara, but none of them covered completely the conditions for Jan or Tom:

S26 Jan: It would be true if we were working within the set of natural numbers.

Eight of the 28 participants displayed erroneous considerations due to their over-fixation to a particular learning context (natural numbers) or a particular didactic model (e.g., a fraction as a part of a whole); they were assigned the code category OF:

S10 Tom: Yes – a fraction is a part of a whole → the result will be always smaller.

Twelve of the 28 participants provided vague or unclear responses to some or all of the bubbles (code category VR). These responses could not be considered as properly justified arguments:

S33 Emil: If you multiply by 0, you always get 0.

Among the 112 responses, 2 were blank and 32 belonged to the categories GE, GED, COD that refer to generic examples and other deductive reasoning.

Emerging concerns

After the arithmetic sub-study, the above-mentioned findings emerged a concern whether similar code categories would also appear when letting the participants reason about a Concept Cartoon with general statements in a different school subject, namely in biology. We addressed this concern in the arithmetic-biology sub-study.

THE ARITHMETIC-BIOLOGY SUB-STUDY (THE SECOND SUB-STUDY)

The arithmetic-biology sub-study intended to focus on argumentation related to general statements together in arithmetic and biology. For this purpose, we created a new Concept Cartoon situated in the context of biology, added it to the Concept Cartoon from the arithmetic sub-study and conducted a new sub-study with both of them.

Diagnostic instrument – biology – particularities

For the biology Concept Cartoon, we chose a zoology topic, namely the topic of large marine animals and their classification. To ensure a similar environment for argumentation as in the arithmetic case, we focused again on possible misconceptions related to critical moments in learning when an earlier way of thinking fails to account sufficiently for new ideas. As pointed out by the research about student thinking in such a context, students often form their knowledge and ideas according to their own experience with the environment and on the basis of intuition (Lazarowitz and Lieb, 2006). A common misconception consists in sorting organisms into incorrect taxonomic classes according to their habitat and/or similarity in a body structure (Kattmann, 2001), for instance, in determining large marine mammals as fish based on the fact that they live in the water (Trowbridge and Mintzes, 1988; Berthelsen, 1999). A similar misconception consists in the inclusion of turtles and reptiles among amphibians (Yen, Yao and Chiu, 2004) or invertebrates (Braund, 1997).

Kubiatko and Prokop (2007) performed a study with 468

Slovak lower-secondary school students in which they monitored various misconceptions related to mammals. An often-occurring misconception was the claim that the penguin is a mammal, justified by the fact that penguins live in the sea just like large marine mammals (cf. Prokop, Prokop and Tunnicliffe, 2007). Only half of the respondents reported that baby whales feed on milk, even though most of them correctly classified whales as mammals. More than two thirds of the respondents displayed a misconception about dolphin breathing, as they reported that dolphins breathe through the gills because they live in an aquatic environment. In connection with the students' difficulties in the classification of whales, Kubiatko and Prokop also provide a possible linguistic explanation for this issue: the Slovak translation of the word *whale* has a meaning of *big fish*. The same situation is in the Czech language.

Based on the above-mentioned findings, we created a Concept Cartoon displaying various improper ways of affiliating animals according to their habitat and/or similarity in body structure or function (Figure 4):

- classifying all big animals living in the water as fish, since they inhabit the same environment as fish;
- including an animal among fish based on the fact that it has gills;
- including an animal among fish because it reproduces by laying eggs.

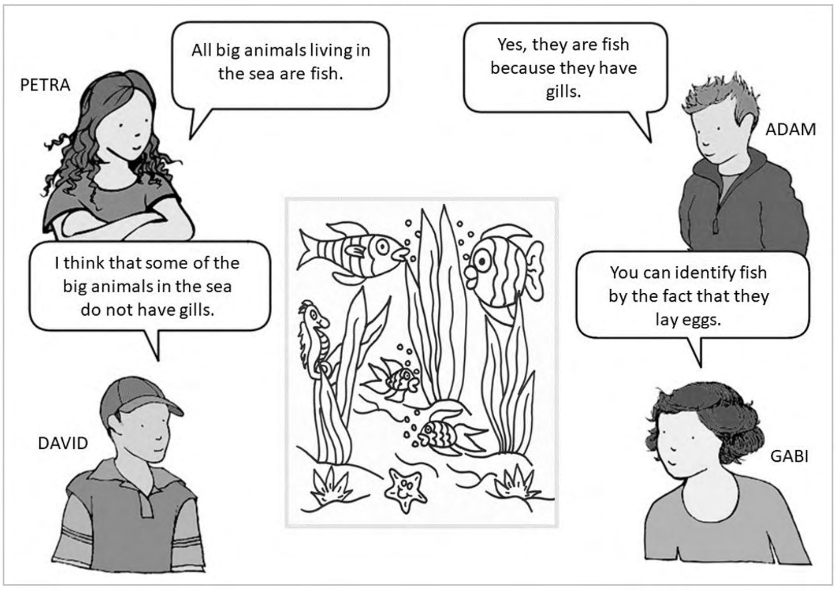


Figure 4: A Concept Cartoon on large marine animals; (source of the template of children with empty bubbles: Dabell, Keogh and Naylor, 2008: 2.10; source of the central picture: Clipart Library, 2019)

Three of the bubbles are not correct, and it is possible to specify many confirming examples as well as counter-examples for their statements:

- Petra: there are large fish living in the sea (e.g., giant oarfish – *Regalecus glesne*) but the largest marine animals are mainly mammals (e.g., whales, dolphins), cartilaginous fishes (e.g., sharks) or cephalopods (e.g., giant squid); with this bubble, the question also emerges of how big must an animal be to be considered a large marine animal;
- Adam: we can identify fish by having gills but most large marine animals are representatives of mammals and thus breathe through the lungs;
- Gabi: most fish lay their eggs, and it is their main way of reproduction; however, there are exceptions, e.g., the family of breeding fish – *Poeciliidae*, who give birth to live babies.

The fourth bubble (David) contains the only statement in the Concept Cartoon that is not a general statement. It is an existential statement that is correct.

Similarly to the arithmetic case, the top-left bubble (Petra) shows the most general statement. The other three bubbles can be considered hints to Petra's bubble, since they indirectly indicate some partial characteristics of the counter-examples related to Petra's bubble. Again, none of the presented characteristics is complete.

Data analysis – particularities

For data analysis, to obtain as similar environments as possible for arithmetic and biology, we decided to analyse just two bubbles, one from each of the Concept Cartoons. We chose the most similar bubbles from the perspective of general statements: Jan's bubble in the arithmetic Concept Cartoon (Figure 3) and Petra's bubble in the biology Concept Cartoon (Figure 4). Both are the most general bubbles in the picture, and the other bubbles serve as their support.

Results of the arithmetic-biology sub-study – the arithmetic part

During the data analysis process, 49 assignments were made for the arithmetic Concept Cartoon (one assignment for each participant – we analysed just Jan's bubble). These assignments covered six of the eight code categories from the arithmetic sub-study (Table 1, excluding GED, COD), plus three new categories for erroneous responses (Table 2).

Electronic ISSN 1803-1617 Printed ISSN 2336-2375

acronym	description	Jan
DC	decimal numbers error	4
ON	number one error	1
DN	different numbers error	1
altogether		6

Table 2: List of acronyms for new arithmetic code categories and their frequency in data, *n* = 49, 2020 (source: own calculation)

acronym	description	single	multiple	altogether
NC	names of classes	9	1	10
NO	names of orders	1	0	1
NS	names of species	2	0	2
NCS	names of classes and their species	15	11	26
NOS	names of orders and their species	0	1	1
NCO	names of classes and their orders	0	1	1
NCS&NOS	names of orders and their species, names of classes and their species	0	1	1
NCOS	names of classes and their orders and their species	1	1	2
altogether		28	16	44

Table 3: List of acronyms for biology code categories for counter-examples and their frequency in data, *n* = 49, 2020 (source: own calculation)

acronym	description	
BA	big-animals error	2
DM	dolphins not being mammals	1
altogether		3

Table 4: List of acronyms for new biology code categories for errors and their frequency in data, *n* = 49, 2020 (source: own calculation)

Code category DC is for the opinion that decimal numbers are numbers between 0 and 1:

R43 Jan is not right, for instance, when I multiply by a decimal number ($6 \cdot 0.5 = 3$), a smaller number comes out.

Code category ON for the opinion that multiplying by 1 gives a smaller number:

R34 Jan is not correct because when we multiply by 1 or 0, the result is smaller.

Code category DN for the opinion that a product of two different numbers gives a bigger number:

R22 $2 \cdot 5 = 10$ when multiplying two different numbers, we get a bigger number.

Results of the arithmetic-biology sub-study – the biology part

For the biology Concept Cartoon, the analytic process was not successful. For 44 of the 49 respondents, we came in data across argumentation aspects that had not appeared in the arithmetic context: counter-examples that consisted of various combinations of names of classes, names of orders and names of species. In these cases, we were not able to eliminate the assignments to one per respondent. Instead, for these 44 respondents, we concluded the analytic process with two assignments per respondent: one assignment for the type of counter-example(s) provided by the respondent (namely, for the particular combination of classes, orders and species), and the other for the number of counter-examples provided by the respondent (single vs multiple), see Table 3.

Among the 44 respondents with counter-examples, 28 presented single counter-examples of various types:

R13 Petra is not right, because mammals can also live in the sea. (code category NC)

R37 Petra – not all large animals in the sea are fish, they are cetaceans. (cat. NO)

R8 Other big animals live in the sea, not just fish, e.g., a whale. (cat. NS)

R25 Not all large animals are fish, for instance, a dolphin is a mammal. (cat. NCS)

R39 Petra is not right – there are also mammals in the sea (cetacean – whale). (cat. NCOS)

The remaining 16 participants provided multiple counter-examples:

R29 Petra is not right, we have also cartilaginous fish and mammals. (cat. NC)

R49 There are not only fish in the sea but also, for example, mammals (dolphin), octopus – cephalopods, turtle – kind of reptiles. (cat. NCO&NCS)

R20 Petra is not right, because cetaceans, e.g., blue whale, dolphin, orca, are big marine animals but mammals. (cat. NCOS)

The rest of the respondents (5 of 49) gave either no response (1 case; code category XR, the same category as in arithmetic), vague response (1 case; code category VR, the same category as in arithmetic) or erroneous answers based on conceptual mistakes. The erroneous answers resulted in establishing two new code categories (Table 4).

Code category BA is for exemplifying marine animals that are not big (e.g., seahorses, shrimps):

R31 Jellyfish and seahorses also live there.

Code category DM for not including dolphins among mammals:

R35 Some animals in the sea do not have gills, such as the dolphin, and mammals also live in the sea.

Due to the inability to eliminate the assignments related to biology counter-examples to one per respondent and relate them to the existing arithmetic code categories, we were faced with the necessity to either change the perspective from which we observed our data or collect additional data to help clarifying the situation. That means, we ended in the situation from the lower-left corner of the diagram in Figure 1. For some time, the analytics process stayed open, unfinished. The resolution came with the geometry sub-study.

THE GEOMETRY SUB-STUDY (THE THIRD SUB-STUDY)

The geometry sub-study also proceeded from the arithmetic sub-study but was conducted independently of the arithmetic-biology sub-study. We got the advantage of the newly established cooperation between the first and third authors and took the opportunity to work on the topic of geometry in which Concept Cartoons have not been properly investigated yet. We employed a Concept Cartoon with data collected many years ago but never researched.

Diagnostic instrument – particularities

The geometric Concept Cartoon discusses the topic of recognising a rectangle. In this context, we deal with two possible ways of comprehending quadrilaterals. The first one considers each quadrilateral as a unique object. For instance, a square is not understood as a special case of rectangle, i.e., rectangles are considered having different length of adjacent sides. De Villiers (1994) calls this approach a *partition classification*. Such an approach is typical for geometric concepts at the primary school level but is not suitable for higher levels of schooling since it makes it difficult to study common properties and relationships between various objects. The secondary school curriculum thus employs a *hierarchical classification* (ibid) that understands specific concepts as subsets of the more general ones. In that sense, squares are considered special cases of rectangles; rectangles are considered special cases of parallelograms, etc. The transfer between partition and hierarchical classifications is a rich source of critical moments in learning when earlier ways of thinking fail to account sufficiently for new ideas.

According to educational research, students often face difficulties with the classification of quadrilaterals: they formulate statements with superfluous (Miler, 2019) or insufficient (Fujita and Jones, 2007) information. Similar difficulties are also observed in future teachers (Tuset, 2019).

The Concept Cartoon in Figure 5 covers various properties of rectangles: “having opposite sides equal in length”, “having all angles right”, “having diagonals that bisect each other” and “having diagonals equal in length”. To be able to identify the rectangle properly, we need a precise identification of the rectangle by assigning properties that are necessary and sufficient for determining the object. Within the framework of hierarchical classification of rectangles, the definition may consist of e.g., in the combination of properties “having diagonals that bisect each other” and “having diagonals equal in length”. Within the framework of partition classification, the two properties would be accompanied by the property “having different lengths of adjacent sides”.

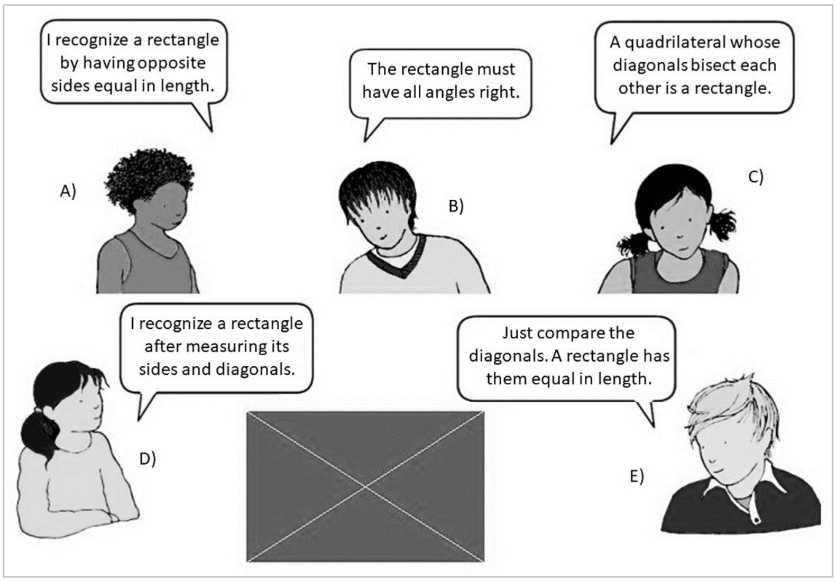


Figure 5: A Concept Cartoon on recognising a rectangle; (source of the template of the children with bubbles: Dabell, Keogh and Naylor, 2008: 1.3; source of the texts in bubbles: Roubíček, 2014)

For this sub-study, we concentrated on bubbles A, C and E of the Concept Cartoon. These three bubbles contain statements that can be all reformulated as implications:

(A) If a quadrilateral has opposite sides equal in length, then it is a rectangle.

(C) If a quadrilateral has diagonals that bisect each other, then it is a rectangle.

(E) If a quadrilateral has diagonals equal in length, then it is a rectangle.

None of the above statements is true since the assumptions do not contain enough information to identify a rectangle unequivocally. Specifically, there exist counter-examples for which the three statements are not valid: A, C and E are not valid for all parallelograms outside of rectangles (e.g., a rhomboid), E is not valid also for some quadrilaterals outside of parallelograms (e.g., an isosceles trapezoid). However, in all three cases, the converse implication is true, i.e., all properties mentioned in the assumptions are actual properties of rectangles. Moreover, it is possible to add additional conditions to the assumptions that would make the statements true; the additional conditions are hinted in the other bubbles.

From the perspective of argumentation, such an arrangement of the content of bubbles can challenge respondents' skills in conditional reasoning, namely their ability to distinguish properly between assumptions and conclusions of a conditional statement that is informally worded. The informal type of conditional statements often appears in primary school classrooms (Hadar, 1977).

Results of the geometry sub-study

During the analytic process, 29 · 3 = 87 assignments were made (3 bubbles for each participant) and eleven relevant code categories were revealed in data (see Table 5). This time, while creating acronyms for code categories, we newly distinguished between three-letter and two-letter ones. The three-letter acronyms were intended for code categories referring to responses that might be considered a constructive basis for further learning of proper argumentation. The two-letter acronyms were intended for code categories referring to blank responses or responses that were incompatible with proper argumentation. The meaning of the individual acronyms is explained in detail in the further text, in the same order as in Table 5.

acronym	description	bubbles			
		A	C	E	A+C+E
GSX	name of a set of counter-examples, no particular example, no justification	6	10	11	27
GEX	name of a set of counter-examples, a particular example given	0	1	0	1
GSD	name of a set of counter-examples, with justification	0	1	1	2
AWX	awareness of the existence of counter-examples, no particular example, no justification	0	4	5	9
AWC	awareness of the existence of counter-examples, conditions of validity indicated but insufficiently	3	0	0	3
COX	conditions of validity indicated, no justification	2	0	0	2
XR	no response	4	7	8	19
AG	agreement with a bubble, no other comments	2	0	3	5
VR	vague response	1	0	0	1
EK	erroneous knowledge of concepts and their properties	0	3	1	4
CI	converse handling of the implication	11	3	0	14
altogether		29	29	29	87

Table 5: List of relevant acronyms for geometry code categories and their frequency in data, *n* = 29, 2014 (source: own calculation)

Data excerpts

According to data analysis, 44 of the assignments of code categories belonged to three-letter acronyms. Among them, the most frequent mode of reasoning consisted of naming a set of counter-examples but not providing a particular counter-example nor justification (code category GSX):

S9 (E) No, it could be, for example, a trapezoid.

Only one of the responses provided a name of a set of counter-examples accompanied by a particular counter-example that was in the form of an illustrative picture (code category GEX):

S9 (C) No, they can bisect in various quadrilaterals (for example, in a square: *a picture of a square with diagonals*).

One of the participants provided the name of a set of counter-examples accompanied by an attempt of a deductive justification (code category GSD):

S31 (C) If we draw both diagonals, 4 isosceles triangles will be formed. But watch out: the diagonals also bisect in a square, a rhombus and a rhomboid.

Some of the participants were aware that the statements in bubbles might not be always true but did not provide any (sets of) counter-examples or additional conditions of validity (code category AWX):

S13 (C) Not only a rectangle.

Others, who were aware that the statements in bubbles might not be always true, provided additional conditions of validity but these conditions were not sufficient (code category AWC):

S6 (A) Yes, that is true. However, we must add that the adjacent sides should not be of the same length. Then it would be a square.

Several participants provided sufficient additional conditions but without justification (code category COX):

S9 (A) No, the other condition is that it must have right angles. Otherwise, it could be a parallelogram.

The remaining 43 assignments of code categories belonged to two-letter acronyms, i.e., to code categories referring to blank responses or responses that were incompatible with proper argumentation. 19 of them referred to cases when the participant did not provide any response to a bubble (code category XR), five to cases when the participant provided only incorrect agreement with a bubble without further comments (code category AG).

One assignment referred to a response that was too vague to be clear (code category VR):

S10 (A) Because a = a'; b = b'.

Some of the participants presented erroneous knowledge of geometric concepts and their properties (code category EK):

S15 (C) A rectangle is a regular quadrilateral – its diagonals do not bisect each other.

Twelve of the participants handled the implications conversely, i.e., confused assumptions with conclusions (code category CI):

S25 (C) True, a rectangle has the diagonals of the same length, they bisect each other and mark the centre of the rectangle.

A reflection on the relation between arithmetic and geometry sub-studies

The results of the geometry sub-study complemented the results of the arithmetic sub-study. Having the current study situated to a different field (geometry) and using a different type of statements in bubbles (implications that are not true but the converse implications are true), only four code categories from the arithmetic sub-study reappeared here: XR, VR, GE and COX. However, the category GE has been renamed as GEX due to the newly established three-letter rule. The remaining code categories are quite new: GSX and GSD refer to the initial stages of GEX and GED where the respondent provides a name of a set of counter-examples but does not provide a particular sample; AWX refers to initial stages of COX or GEX; AWC refers to initial stages of COX; EK refers to erroneous responses similar as OF but with a different source of the error (over-fixation to previous concepts in OF vs weak knowledge of new concepts in EK); CI refers to erroneous knowledge of logical aspects; and AG refers to incorrect agreements without further comments (it might refer to initial stages of VR, EK or CI).

Emerging concerns

The above-mentioned reflection on the relation between the arithmetic and geometry sub-studies emerged a hope that the collection of geometric data might become the desired missing piece allowing us to conclude the still unfinished analytic process of the arithmetic-biology sub-study. For this purpose, we combined the three sub-studies together, revisited all their data and this is how the main study came about.

THE MATHEMATICS-BIOLOGY STUDY (THE MAIN STUDY)

Data analysis

During the main study, we revisited data from the arithmetic-biology sub-study and observed them from the perspective of the findings of the geometry sub-study. Primarily, we focused on possible interconnections between biology and geometry code categories. During the constant comparison process, we also revisited data from the arithmetic and geometry sub-studies. Similarly, like in the arithmetic-biology sub-study, we chose one bubble from each of the three Concept Cartoons for data analysis. From the geometric Concept Cartoon, we selected the C bubble. During the analytic process, we followed the convention of distinguishing between three-letter code categories referring to responses that might be considered a constructive basis for further learning of proper argumentation and two-letter code categories referring to blank responses or responses incompatible with proper argumentation.

Results of the main study – three-letter categories

This time, the analytic process was successful and we made $28 + 2 \cdot 49 + 29 = 155$ assignments in data collected during the first, second and third sub-studies, i.e., in all standard participants' data related to the three selected bubbles (Jan, Petra, C). These assignments covered seven existing three-letter categories (AWX, CEX, GSX, GSD, GEX, COX, COD). Since data were revisited during the main study, some of the excerpts were assigned a different category in the main study than in the sub-study. For instance, one of the responses from the arithmetic sub-study and one of the responses from the biology part of the arithmetic-biology sub-study that had been originally assigned the VR category were labelled later as the AWX category:

S18 Jan: You can get also a smaller number, or the same.

R48 Petra is wrong, not all the big animals in the sea are fish.

We considered four of the three-letter categories *relevant* for the joint mathematics-biology approach (i.e., for the second research question): AWX, GSX and GEX since they appeared in all three Concept Cartoons, and CEX that appeared in arithmetic and biology ones.

The other three-letter categories did not appear across the subjects (COX and COD were just in arithmetic, GSD just in geometry). So that we included additional participants (future lower-secondary school teachers who specialize in mathematics and/or biology) to see whether they could enrich the categories also in other subjects. During this additional analytic process, we made $2 \cdot 5 + 5 = 15$ assignments but none such enrichment came into being, so that we labelled the categories COX, COD, GSD as *irrelevant* for the joint approach.

For details on the relevant three-letter code categories, their frequency in data (from standard as well as additional participants = 155 + 15 = 170 assignments) and relations to the code categories from the sub-studies, see Table 6. The code categories in Table 6 are sorted by the quality of the argument, from the weakest (AWX) to the strongest (GEX).

Electronic ISSN 1803-1617 Printed ISSN 2336-2375

final acronym	description	frequency in data	previous acronyms
AWX	awareness of the existence of counter-examples, no particular counter-example given, no justification	7	AWX, VR (some)
CEX	a counter-example (i.e., an example for which the statement is not valid), detached, no justification	30	CEX, NS
GSX	a generic example in the form of a name or characteristics of a set of counter-examples, no particular element of the set given, no justification	32	GSX, NC, NO, NCO
GEX	a generic example in the form of a name or characteristics of a set of counter-examples, a particular element of the set given, no justification	54	GE, GEX, NCS, NOS, NCOS
altogether		123	

Table 6: Final list of relevant three-letter acronyms for mathematics-biology code categories and their previous appearances, frequency in data from all participants (170 assignments), *n* = 89, 2014-2021 (source: own calculation)

final acronym	description	frequency in data	previous acronyms
XR	no response	10	XR
AG	agreement with a bubble, no other comments	1	AG
VR	vague response	3	VR (some)
EC	erroneous knowledge of mathematical or science concepts and their properties	13	EK, OF, DC, ON, DN, BA, DM
EL	erroneous knowledge of logical aspects	4	CI
altogether		31	

Table 7: Final list of two-letter acronyms for mathematics-biology code categories and their previous appearances, frequency in data from all participants (170 assignments), *n* = 89, 2014-2021 (source: own calculation)

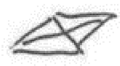
acronym	arithmetic	biology	geometry
	If you multiply one number by another, you always get a bigger number.	All large animals living in the sea are fish.	A quadrilateral whose diagonals bisect each other is a rectangle.
AWX	S18: You can get also smaller number, or the same.	R48: Not all big animals in the sea are fish.	S13: Not only a rectangle.
CEX	S31: $7 \cdot (-3) = -21$	R8: There are other big animals in the sea than just fish, e.g., a whale.	S25 (part): 
GSX	S4: Numbers might be smaller (e.g., after multiplying a negative number).	R26: There are also mammals in the sea.	S6: Yes, but also a square.
GEX	R25: That is not true because when multiplying negative numbers, we can get a number that is smaller: $8 \cdot (-8) = -64$.	R43: There are also mammals in the sea (e.g., dolphin).	S9: No, they can bisect in various quadrilaterals, for example, in a square: 

Table 8: A conversion table between data excerpts from arithmetic, biology and geometry across all relevant three-letter mathematics-biology code categories

Results of the main study – two-letter categories

As for the two-letter categories, the assignments from the main study covered three existing two-letter categories (XR, VR, AG). Moreover, two completely new categories were established to generally cover all the previous categories on various conceptual (EC) and logical (EL) errors. For details on two-letter code categories, see Table 7.

The sum of occurrences from Tables 6 and 7 equals 123 + 31 = 154. The remaining 170 – 154 = 16 assignments belong to the irrelevant three-letter code categories (COX, COD, GSD) that are not listed in the tables.

Results of the main study – the conversion table

To illustrate better the interrelations between code categories across all subjects, we prepared a conversion table that compares data excerpts from arithmetic, biology and geometry across all relevant three-letter code categories (Table 8). Since the CEX category was not represented in geometric data, we drew on the fact that GEX = CEX + GSX and, as an example of CEX in geometry, we selected a part of an excerpt labelled as GEX.

Using the final tables of acronyms and the conversion table (Tables 6 to 8), we may study the modes of argumentation in different subject domains (arithmetic, biology, geometry) from the perspective of individual respondents. We applied this approach to data from the arithmetic-biology sub-study, with the following results: 48 of the 49 respondents were assigned relevant code categories both in arithmetic and biology, of them 17 were assigned the same code category in both subjects, 19 were assigned two different three-letter categories, 3 were assigned a two-letter code category in biology but a three-letter code category in arithmetic, and 9 were assigned a three-letter code category in biology but a two-letter code category in arithmetic. Among the 19 respondents with different three-letter code categories, 7 provided stronger arguments in mathematics (e.g., CEX in mathematics and AWX in biology), and 24 provided stronger arguments in biology (e.g., CEX in mathematics and GSX in biology).

DISCUSSION

This study enriched the findings about possible use of Concept Cartoons in the professional preparation of primary school teachers by focusing on Concept Cartoons from the perspective of argumentation in both mathematics and science. In the context of argumentation about general statements, we brought an approach that is joint for elementary mathematics (arithmetic, geometry) and science (biology). The main part of the study arose from three sub-studies located in arithmetic, in arithmetic and biology, and in geometry. Although the arithmetic sub-study was the first one from which the other two gradually devolved, eventually it was the geometry sub-study that turned out to be the key one for understanding the interrelations between reasoning in arithmetic and biology that appeared in our data.

Regarding the first research question

As an answer to the first research question, "*What kinds of reasoning about general statements in biology and mathematics can be observed in future primary school teachers when using Concept Cartoons as a diagnostic instrument?*" we may state that the results themselves are promising from the perspective of participants' knowledge as well as from the perspective of applicability of Concept Cartoons as a diagnostic instrument.

As for participants' knowledge displayed in responses to the three Concept Cartoons, the findings are partially in accordance with previous research studies. Like in (Simon and Blume, 1996), some of the participants provided responses that could not be considered justifications since being vague or unclear (category VR) but these responses accounted for only 3 of the 170 assignments made in the main study. Unlike in (ibid), none of the participants presented a response that could be considered an external conviction proof scheme. Some of the participants showed erroneous knowledge of mathematical or science concepts and their properties (category EC; 13 of 170 assignments in the main study), mainly related to critical moments in learning (Confrey and Kazak, 2006; Lazarowitz and Lieb, 2006). Unlike in (Trowbridge and Mintzes, 1988), none of the misconceptions in biology was based on determining large marine mammals as fish based on the fact that they live in the water. None of the participants confirmed the findings of Zazkis and Chernoff (2008) since they widely used counter-examples but did not hesitate to refute the statements on their basis. Like in (Martin and Harel, 1989), some of the counter-examples in arithmetic were big-number examples (e.g., S32/Jan). In mathematics as well as biology, none of the participants tried to affirm the validity of a general statement through several confirming examples.

From the perspective of the learning process in argumentation, we distinguished between three- and two-letter acronyms for code categories. The three-letter ones (139 of the 170 assignments) referred to responses that might have been considered a constructive basis for further learning of proper argumentation, the two-letter ones (31 of the 170 assignments) referred to blank responses or responses that were not compatible with proper argumentation. Similarly, like in Buchbinder and Cook (2018), some of the responses incompatible with proper argumentation were caused by erroneous knowledge of logical aspects (category EL), namely by confusing assumptions and conclusions in an informally worded implication (Hadar, 1977). Others were caused by erroneous knowledge of concepts and their properties (category EC), for instance by natural number bias (Alibali and Sidney, 2015) or by taxonomy determined according to the habitat (Kubiatko and Prokop, 2007).

As for the Concept Cartoons in the role of a diagnostic instrument, the willingness to respond to the bubbles was high among the participants: only 10 of the 170 responses were blank (category XR). A wide range of proof scheme types and subtypes appeared in data, densely filling the content-related code categories (AWX, CEX, GSX, GEX, GSD, COX, COD, EC; 152 of the 170 assignments). From the perspective of Simon and Blume's (1996) levels of future teachers' argumentation, most of the three-letter assignments could be considered generic examples (level 3; categories GSX, GEX; 86 of the 170 assignments). They were followed by empirical demonstrations (level 2; category CEX; 30 of the 170 assignments), and instance-independent deductive justifications (level 4; categories GSD, COD; 4 of the 170 assignments). Some of the three-letter responses were too incomplete to be clearly related to a particular level (categories AWX, COX; 14 of the 170 assignments).

The three Concept Cartoons differed not only in the subject domains where they were situated but also in other aspects, and these differences influenced collected data. For instance, the geometric Concept Cartoon was the only one presenting informally worded implications that were not true but their converse versions were. Therefore, it provided an opportunity to investigate whether the respondents distinguished properly between assumptions and conclusions; some of the

Electronic ISSN 1803-1617 Printed ISSN 2336-2375

respondents failed in it (category CI; 14 of the 87 assignments in the geometry sub-study). On the other hand, the arithmetic Concept Cartoon was the only one able to stimulate the respondents to indicate conditions of validity and justify them (category COD; 17 of the 112 assignments in the arithmetic sub-study). Such differences in data can direct possible further research concerning argumentation and Concept Cartoons, e.g., at focusing on relations between detailed characteristics of Concept Cartoons and modes of argumentation revealed in data when using the Concept Cartoons for collecting data.

Regarding the second research question

As an answer to the second research question, "*What are the possibilities of joint assessment of reasoning about general statements in biology and mathematics?*" we may also state that the results are promising. The combination of the three Concept Cartoons from arithmetic, geometry and biology showed that there is a possibility to apply a joint approach to the assessment of reasoning in the three subject domains. As illustrated in the conversion table (Table 8), there were four three-letter categories in our data that had a common description for all three subject domains (Table 6). Their respective excerpts appeared in all three domains (AWX, CEX, GSX, GEX; 123 of the 170 assignments). These categories covered the field of instance-dependent deductive justifications (level 3 according to Simon and Blume, 1996), including its initial stages (AWX, CEX and GSX might refer to initial stages of GEX), and were mutually related (GEX = CEX + GSX).

The field of instance-independent deductive justifications (level 4 according to Simon and Blume, 1996) was represented in data only rarely (categories GSD, COD; 4 of the 170 assignments) and separately (GSD just in geometry, COD just in arithmetic). Even the involvement of additional participants who were future lower-secondary school teachers specialising in mathematics and/or biology did not help increase these two categories' occurence. Even if we considered not only the main study but also the sub-studies, there would be only one other category in data that might be considered of the level 4: code category GED with 5 assignments. This category appeared in the part of the arithmetic sub-study that was not common to the main study. However, the GED category also did not appear in other subject domains.

The current study did not solve the issue of a joint assessment of instance-independent deductive reasoning (level 4 according to Simon and Blume, 1996) due to its sparse occurrence in data. The question is whether the next step in a systematic joint approach to the level-4 terrain should consist in approaching domain specialists (mathematicians, biologists) or in changing the Concept Cartoons for data collection.

From the perspective of individual participants, the joint set of code categories provided by this research can be used for investigating modes of reasoning about general statements comparably in mathematics and biology. During the arithmetic-biology sub-study, individual participants provided modes of reasoning that differed in arithmetic and biology. Such a difference might have been caused by different levels of participant's content knowledge in the two subject topics but also by different approaches to argumentation that the particular participant might have applied within the two subject contexts. Additional questions and/or additional diagnostic tasks would be needed to identify the specific cause of the difference. However, the second cause reveals opportunities for a joint development of argumentation skills: stemming from the comparison of arguments provided in the two subjects and building on the approach applied in the subject with stronger arguments.

Regarding the implications for argumentation and formative assessment

Since generic examples can be considered the first step of deductive reasoning (Harel and Sowder, 2007; Simon and Blume, 1996) and data collected within our study covered the field of generic examples, it is possible to use the categories arisen from data analysis for scaffolding the process of learning the concept of generic examples. Consequently, the categories can also be used in learning the foundations of deductive reasoning. The possibility to label individual steps of the learning process by code categories then may allow teachers to orient themselves better in the course of formative assessment and also in actual levels of students' knowledge that is to be addressed by this assessment. In that sense, it may indirectly help the teachers promote their noticing of various reasoning forms (Melhuish, Thanheiser and Guyot, 2020).

More generally, we showed that there is a possibility for argumentation and formative assessment to be understood equally in mathematics and biology. Such a finding complements our recent study that has resulted in a joint communication model for inquiry and formative assessment in mathematics and biology (Rokos and Samková, 2020).

Regarding the implications for professional preparation of primary school teachers

The joint coding procedure for biology and mathematics offers a diagnostic tool for assessing future primary school teachers' modes of reasoning comparably in both school subjects. This arrangement provides a common background for teacher educators in mathematics and teacher educators in biology that allows them to collaborate on the planning as well as assessment of their respective teacher training courses. In that sense, the procedure enriches subject didactic of mathematics as well as subject didactic of biology (Kubiatko, 2021).

The presented research participants were future primary school teachers in a professional preparation program that simultaneously approaches the study of content and didactics in mathematics and biology. Within this environment, it is possible to intensively utilize a joint approach to argumentation and reasoning in the two school subjects. As mentioned above, any differences identified between the modes of future teachers' reasoning in the two subjects might provide an opportunity for the development of future teachers' argumentation skills and thus positively affect their professional knowledge (e.g., by enhancing the quality of their feedback; Hattie and Timperley, 2007).

The joint approach to reasoning about general statements in mathematics and biology is valuable for future primary school teachers not only as an example related to formative assessment within inquiry-based science and mathematics education practices (Dolin and Evans, 2018) but also as an example related to other educational frameworks utilizing the integration of mathematics and biology school subjects. For instance, to the STEM framework that integrates science, technology, engineering and mathematics education (Moore, Johnson and Glancy, 2020). In this sense, our paper contributes to the educational research on STEM by addressing the objection given by English (2016) that the STEM-related research mostly focuses on STEM integration just from the point of view of individual school subjects. We focus jointly on biology and mathematics, and from a general perspective of argumentation, so that we allow to approach the two school subjects commonly despite their different contents and different approaches to didactics (Hallström and Schönborn, 2019). Unlike some of the research on mathematics within STEM education, we do not perceive mathematics just as a computational tool for other STEM subjects (Valovičová et al., 2020) but as an equivalent subject identity. Such an arrangement forms a suitable ground for supporting and developing STEM-oriented primary (or elementary) school teacher education (Corp, Fields and Naizer, 2020).

CONCLUSION

In this study, we focused on future primary school teachers and their modes of argumentation in mathematics and science, namely in arithmetic, geometry and biology. For each of the subject fields, we prepared a Concept Cartoon focusing on general statements, collected data using them, and then conducted qualitative data analysis looking for displays of various levels of argumentation and for possibilities to assess these displays jointly in mathematics and science. Our effort resulted in a set of code categories for various levels of instance-dependent deductive reasoning (including its initial stages) and in a conversion table that provided the framework for investigating and comparing modes of argumentation of individual respondents across the three subject domains (arithmetic, geometry, biology). Such a joint approach offers the opportunity to understand and develop respondents' knowledge in a deeper way by approaching the same concept (argumentation) jointly within three different subject domains. From a broader perspective, we entered the issue of a joint approach to argumentation in mathematics and biology in relation to inquiry-based education as an attempt to perceive inquiry in a comparable manner in both subjects. Since inquiry-based education is naturally rich in generalisations and in formulating and discussing general statements, our study's findings provide a highly applicable framework in the inquiry-based environment. For future research, it might be valuable to gain a more general perspective by extending the approach to other science subject domains (physics, chemistry) or studying the approach also in relation to other STEM subjects (technology, engineering). These additional subjects and subject domains might be represented in primary school content and/or professional preparation of primary school teachers to a lesser extent than mathematics and biology, however, the joint approach to argumentation might also be investigated in relation to secondary school education.

ACKNOWLEDGEMENT

This article was supported by the Technology Agency of the Czech Republic under Grant 'Learning Hyperspace for Formative Assessment and Inquiry Based Science Teaching', project No. TL02000368.

REFERENCES

Alibali, M. W. and Sidney, P. G. (2015) 'Variability in the natural number bias: who, when, how, and why', *Learning and Instruction*, Vol. 37, pp. 56–61. http://doi.org/10.1016/j.learninstruc.2015.01.003

Artigue, M. and Blomhøj, M. (2013) 'Conceptualizing inquiry-based education in mathematics', *ZDM Mathematics Education*, Vol. 45, No. 6, pp. 797–810. http://doi.org/10.1007/s11858-013-0506-6

Balacheff, N. (1988) 'Aspects of proof in pupils' practice of school mathematics', in Pimm, D. (ed.) *Mathematics, Teachers and Children*, London: Hodder & Stoughton, pp. 216–238.

Berland, L. K. and Reiser, B. J. (2009) 'Making sense of argumentation and explanation', *Science Education*, Vol. 93, No. 1, pp. 26–55. http://doi.org/10.1002/sce.20286

Berthelsen, B. (1999) 'Students' naive conceptions in life sciences', *MSTA Journal*, Vol. 44, No. 1, pp. 13–19.

Braund, M. (1997) 'Primary children's ideas about animals with and without backbones', *Education 3-13*, Vol. 25, No. 2, pp. 19–24. http://doi.org/10.1080/03004279785200191

Buchbinder, O. and Cook, A. (2018) 'Examining the mathematical knowledge for teaching of proving in scenarios written by pre-service teachers', in Buchbinder, O. and Kuntze, S. (ed.) *Mathematics Teachers Engaging with Representations of Practice*, Cham: Springer, pp. 133–154. http://doi.org/10.1007/978-3-319-70594-1_8

Bulková, K., Medová, J. and Čeretková, S. (2020) 'Identification of crucial steps and skills in high-achievers' solving complex mathematical problem within mathematical contest', *Journal on Efficiency and Responsibility in Education and Science*, Vol. 13, No. 2, pp. 67–78. http://doi.org/10.7160/eriesj.2020.130202

Clipart Library (2019) *Ocean floor black and white*, [Online], Available: http://clipart-library.com/coloring/kTMegybTj.gif [14 Nov 2020].

Confrey, J. and Kazak, S. (2006) 'A thirty-year reflection on constructivism in mathematics education in PME', in Gutiérrez A. and Boero P. (ed.) *Handbook of Research on the Psychology of Mathematics Education: Past, Present and Future*, Rotterdam: Sense, pp. 305–345.

Corp, A., Fields, M. and Naizer, G. (2020) 'Elementary STEM teacher education: Recent practices to prepare general elementary teachers for STEM', in Johnson C. C., Mohr-Schroeder, M. J., Moore, T. J. and English, L. D. (ed.) *Handbook of Research on STEM Education*, New York: Routledge, pp. 337–348. http://doi.org/10.4324/9780429021381-32

Cullen, S., Fan, J., van der Brugge, E. and Elga, A. (2018) 'Improving analytical reasoning and argument understanding: a quasi-experimental field study of argument visualization', *npj Science of Learning*, Vol. 3, No. 21. http://doi.org/10.1038/s41539-018-0038-5

Dabell, J., Keogh, B. and Naylor, S. (2008) *Concept Cartoons in Mathematics Education*, Sandbach: Millgate House Education.

Dofner, T., Förtsch, C., Germ, M. and Neuhaus, B. J. (2018) 'Biology instruction using generic framework of scientific reasoning and argumentation', *Teaching and Teacher Education*, Vol. 75, pp. 232–243. http://doi.org/10.1016/j.tate.2018.07.003

Dolin, J. and Evans, R. (ed.) (2018) *Transforming assessment: Through an interplay between practice, research and policy*, Cham: Springer. http://doi.org/10.1007/978-3-319-63248-3

van Dooren, W., Lehtinen, E. and Verschaffel, L. (2015) 'Unraveling the gap between natural and rational numbers', *Learning and Instruction*, Vol. 37, pp. 1–4. http://doi.org/10.1016/j.learninstruc.2015.01.001

Dorier, J.-L. and Maass, K. (2014) 'Inquiry based mathematics education', in Lerman S. (ed.) *Encyclopedia of Mathematics Education*, Dordrecht: Springer, pp. 300–304. http://doi.org/10.1007/978-94-007-4978-8_176

English, L. D. (2016) 'STEM education K-12: perspectives on integration', *International Journal of STEM Education*, Vol. 3, No. 3. http://doi.org/10.1186/s40594-016-0036-1

Erduran, S. and Jiménez-Aleixandre, M. P. (ed.) (2007) *Argumentation in science education: Perspectives from classroom-based research*, Dordrecht: Springer. http://doi.org/10.1007/978-1-4020-6670-2

Fujita, T. and Jones, K. (2007) 'Learners' understanding of the definitions and hierarchical classification of quadrilaterals: Towards a theoretical framing', *Research in Mathematics Education*, Vol. 9, No. 1, pp. 3–20. http://doi.org/10.1080/14794800008520167

Furtak, E. M., Hardy, I., Beinbrech, C., Shavelson, R. J. and Shemwell, J. T. (2010) 'A framework for analyzing evidence-based reasoning in science classroom discourse', *Educational Assessment*, Vol. 15, No. 3–4, pp. 175–196. http://doi.org/10.1080/10627197.2010.530553

Galbraith, P. L. (1981) 'Aspects of proving: A clinical investigation of process', *Educational Studies in Mathematics*, Vol. 12, No. 1, pp. 1–28. http://doi.org/10.1007/BF00386043

Hadar, N. (1977) 'Children's conditional reasoning', *Educational Studies in Mathematics*, Vol. 8, No. 4, pp. 413–438. http://doi.org/10.1007/bf00310946

Hallström, J. and Schönborn, K. J. (2019) 'Models and modelling for authentic STEM education: reinforcing the argument', *International Journal of STEM Education*, Vol. 6, No. 22. http://doi.org/10.1186/s40594-019-0178-z

Harel, G. and Sowder, L. (1998) 'Students' proof schemes: Results from exploratory studies', in Schoenfeld, A. H., Kaput, J., Dubinski, E. and Dick, T. (ed.) *Issues in Mathematics Education (Research in Collegiate Mathematics Education III, Vol. 7)*, Providence: AMS, pp. 234–283. http://doi.org/10.1090/cbmath/007/07

Harel, G. and Sowder, L. (2007) 'Toward comprehensive perspectives on the learning and teaching of proof', in Lester, F. (ed.) *Second Handbook of Research on Mathematics Teaching and Learning*, Charlotte: NCTM, pp. 805–842.

Hattie, J. and Timperley, H. (2007) 'The power of feedback', *Review of Educational Research*, Vol. 77, No. 1, pp. 81–112. http://doi.org/10.3102/003465430298487

Kattmann, U. (2001) 'Aquatics, flyers, creepers and terrestrials – Students' conceptions of animal classification', *Journal of Biological Education*, Vol. 35, No. 3, pp. 141–147. http://doi.org/10.1080/00219266.2001.9655763

Keogh, B. and Naylor, S. (1999) 'Concept Cartoons, teaching and learning in science: an evaluation', *International Journal of Science Education*, Vol. 21, No. 4, pp. 431–446. http://doi.org/10.1080/095006999290642

Komatsu, K. (2010) 'Counter-examples for refinement of conjectures and proofs in primary school mathematics', *Journal of Mathematical Behavior*, Vol. 29, No. 1, pp. 1–10. http://doi.org/10.1016/j.jmathb.2010.01.003

Kubiatko, M. (2021) 'Subject didactics: relevant issues', *Problems of Education in the 21st Century*, Vol. 79, No. 3, pp. 340–342. http://doi.org/10.33225/pec/21.79.340

Kubiatko, M. and Prokop, P. (2007) 'Pupils' misconceptions about mammals', *Journal of Baltic Science Education*, Vol. 6, No. 1, pp. 5–14.

Lazarou, D., Sutherland, R. and Erduran, S. (2016) 'Argumentation in science education as a systemic activity: An activity-theoretical perspective', *International Journal of Educational Research*, Vol. 79, pp. 150–166. http://doi.org/10.1016/j.ijer.2016.07.008

Lazarowitz, R. and Lieb, C. (2006) 'Formative assessment pre-test to identify college students' prior knowledge, misconceptions and learning difficulties in biology', *International Journal of Science and Mathematical Education*, Vol. 4, No. 4, pp. 741–762. http://doi.org/10.1007/s10763-005-9024-5

Martin, W. G. and Harel, G. (1989) 'Proof frames of preservice elementary teachers', *Journal for Research in Mathematics Education*, Vol. 20, No. 1, pp. 41–51. http://doi.org/10.2307/749097

McComas, W. F. (ed.) (2002) *The nature of science in science education*, Boston: Kluwer. http://doi.org/10.1007/0-306-47215-5

Melhuish, K., Thanheiser, E. and Guyot, L. (2020) 'Elementary school teachers' noticing of essential mathematical reasoning forms: justification and generalization', *Journal of Mathematics Teacher Education*, Vol. 23, pp. 35–67. http://doi.org/10.1007/s10857-018-9408-4

Miles, M. B., Huberman, A. M. and Saldaña, J. (2014) *Qualitative data analysis: A methods sourcebook*, 3rd edition, Thousand Oaks, CA: SAGE.

Minner, D., Levy, A. and Century, J. (2010) 'Inquiry-based science instruction – what is it and does it matter? Results from a research synthesis years 1984 to 2002', *Journal of Research in Science Teaching*, Vol. 47, No. 4, pp. 474–496. http://doi.org/10.1002/tea.20347

Moore, T. J., Johnson, A. C. and Glancy, A. W. (2020) STEM integration: A synthesis of conceptual frameworks and definitions', in Johnson C. C., Mohr-Schroeder, M. J., Moore, T. J. and English, L. D. (ed.) *Handbook of Research on STEM Education*, New York: Routledge, pp. 3–16. http://doi.org/10.4324/9780429021381-2

National Research Council (1996) *National science education standards*, Washington, DC: National Academy Press.

Naylor, S. and Keogh, B. (2010) *Concept Cartoons in Science Education*, 2nd Edition [CD-ROM], Sandbach: Millgate House Education.

Naylor, S., Keogh, B. and Downing, B. (2007) 'Argumentation and primary science', *Research in Science Education*, Vol. 37, No. 1, pp. 17–39. http://doi.org/10.1007/s11165-005-9002-5

Ping, I. L. L., Halim, L. and Osman, K. (2020) 'Explicit teaching of scientific argumentation as an approach in developing argumentation skills, science process skills and biology understanding', *Journal of Baltic Science Education*, Vol. 19, No. 2, pp. 276–288. http://doi.org/10.33225/jbse/20.19.276

Prokop, P., Prokop, M. and Tunnicliffe, S. D. (2007) 'Effects of keeping animals as pets on children's concepts of vertebrates and invertebrates', *International Journal of Science Education*, Vol. 30, No. 4, pp. 431–449. http://doi.org/10.1080/09500690701206686

Rocard, M., Csermely, P., Jorde, D., Lenzen, D., Walberg-Henriksson, H. and Hemmo, V. (2007) *Science education now: a renewed pedagogy for the future of Europe*, Luxembourg: Office for Official Publications of the European Communities, [Online], Available: http://www.eesc.europa.eu/resources/docs/rapportrocardfinal.pdf [4 Jun 2021].

Rokos, L. and Samková, L. (2020) 'Coding classroom talk from the perspective of formative assessment and inquiry-based education: a communication model for mathematics and science lessons', in *Proceedings of ICERI2020 Conference*, Seville, pp. 1170–1179. http://doi.org/10.21125/iceri.2020.0318

Roubíček, F. (2014) *The set of four geometric Concept Cartoons for assessing future primary school teacher's knowledge*, [Internal material, unpublished].

Samková, L. (2019) 'Investigating subject matter knowledge and pedagogical content knowledge in mathematics with the Concept Cartoons method', *Scientia in educatione*, Vol. 10, No. 2, pp. 62–79. http://doi.org/10.14712/18047106.1548

Samková, L. (2020) 'Observing how future primary school teachers reason about general statements', in *Proceedings of the 17th International Conference on Efficiency and Responsibility in Education (ERIE 2020)*, Prague, pp. 263–271.

Samková, L. and Tichá, M. (2017) 'On the way to observe how future primary school teachers reason about fractions', *Journal of Efficiency and Responsibility in Education and Science*, Vol. 10, No. 4, pp. 93–100. http://doi.org/10.7160/eriesj.2017.100401

Selden, A. (2012) 'Transitions and proof and proving at tertiary level', in Hanna, G. and de Villiers, M. (ed.) *Proof and Proving in Mathematics Education*, Dordrecht: Springer, pp. 391–420. http://doi.org/10.1007/978-94-007-2129-6_17

Simon, M. A. and Blume, G. W. (1996) 'Justification in the mathematics classroom: A study of prospective elementary teachers', *Journal of Mathematical Behavior*, Vol. 15, No. 1, pp. 3–31. http://doi.org/10.1016/s0732-3123(96)90036-x

Stevens, A. L., Wilkins, J. L. M, Lovin, L. H., Siegfried, J., Norton, A. and Busi, R. (2020) 'Promoting sophisticated fraction constructs through instructional changes in a mathematics course for PreK-8 prospective teachers', *Journal of Mathematics Teacher Education*, Vol. 23, No. 2, pp. 153–181. http://doi.org/10.1007/s10857-018-9415-5

Trowbridge, J. E., and Mintzes, J. J. (1988) 'Alternative conceptions in animal classification: A cross-age study', *Journal of Research in Science Teaching*, Vol. 25, No. 7, pp. 547–571. http://doi.org/10.1002/tea.3660250704

Tuset, G. A. (2019) 'Preservice teachers' geometrical discourses when leading classroom discussions about defining and classifying quadrilaterals', in *Proceedings of CERME11*, Utrecht, pp. 3523–3530.

Valovičová, L'., Ondruška, J., Zelenický, L'., Chytrý, V. and Medová, J. (2020) 'Enhancing computational thinking through interdisciplinary STEAM activities using tablets', *Mathematics*, Vol. 8, No. 12, 2128. http://doi.org/10.3390/math8122128

de Villiers, M. (1994) 'The role and function of a hierarchical classification of quadrilaterals', *For the Learning of Mathematics*, Vol. 14, No. 1, pp. 11–18.

Vízek, L. and Samková, L. (2021) 'Observing how future primary school teachers reason about quadrilaterals', in *Proceedings of the 18th International Conference on Efficiency and Responsibility in Education (ERIE 2021)*, Prague, pp. 159–167.

Yen, C.-F., Yao, T.-W. and Chiu, Y.-C. (2004) 'Alternative conceptions in animal classification focusing on amphibians and reptiles: A cross-age study', *International Journal of Science and Mathematics Education*, Vol. 2, No. 2, pp. 159–174. http://doi.org/10.1007/s10763-004-1951-z

Zazkis, R. and Chernoff, E. J. (2008) 'What makes a counterexample exemplary?', *Educational Studies in Mathematics*, Vol. 68, pp. 195–208. http://doi.org/10.1007/s10649-007-9110-4