ERIES JOURNAL

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 PhDr. Michaela Cocca. 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 17
issue 1

2024

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 (CZU), CZ 165 00 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, 2024

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.

In the first issue of 2024 (Vol. 17, No. 1), which you hold, we are pleased to present eight articles from authors in Argentina, the Czech Republic, Indonesia, the Philippines, Turkey, and Ukraine.

The first article, entitled "The Relationship Between High School Students' Perceptions of Role Modeling and Self-Regulation for Science," written by Melek Karaca, Oktay Bektaş, and Seyide Eroğlu, aims to develop a model to show the relationship between high school students' role-modeling perceptions and self-regulation skills. In the study, the predictive correlation design of the quantitative research method was used on a sample of 362 high school students studying at a public high school in Turkey. The results revealed that role models were important in developing high school students' self-regulation skills. Consequently, the authors stress the necessity for high school teachers to use strategies to develop students' self-regulation skills, ensure active participation, and support students effectively until they reach a sufficient affective level.

The second article, "Internal Image of Czech Tertiary Business Schools and Its Influence on The Interest of New Applicants for Study" by Jana Pavelková, Jana Turčínková, and Jakub Šácha presents findings about important factors influencing the overall students' satisfaction with university life, as well as their willingness to share positive references. The study took place in the Czech Republic, where students attended university business schools. The authors collected data via an online questionnaire with students with bachelor's, master's, and doctorate degrees (n = 274), complemented by ten in-depth interviews. The authors observed a close positive correlation between students' satisfaction and the willingness to recommend a university, where the quality of student life and the school's reputation were identified as the most important factors.

The third article, "Grade Point Average: The Relationship with Results of Entrance Assessment, Learning Motivation, Achievement Motivation, and Perception of Teacher Leadership" written by Nataliia Sereda, Svitlana Reznik, Tetiana Solodovnyk, Zhanna Bogdan and Oleksandr Romanovskyi, analyzed the relationship between the GPA of graduates of social majors at National Technical University «Kharkiv Polytechnic Institute» (Ukraine) with the results of the entrance assessment, learning motivation, achievement motivation, and perception of teacher leadership. For this purpose, the authors used correlation and regression analyses using responses from 502 graduate students at the Department of Pedagogy and Psychology of Social System Management between 2018 and 2020. The Results of the presented study demonstrate a significant correlation between the level of academic success and the level of educational and cognitive motivation of the graduates.

In the fourth article, "Efficiency Assessment on Codified Knowledge Products: An SFA Approach", Gustavo Ferro and Nicolás Gatti applied a Stochastic Frontier Analysis to assess cost efficiency in the production of codified knowledge outputs of National Innovation Systems in 82 countries. The authors used panel data including 1,189 observations for 23 years (1996-2019). The analysis revealed that 20 out of 82 countries explain more than 92 percent of the financial resources devoted to research and development, 88 percent of the researchers, 82 percent of documents and published documents, and nearly 95 percent of patent publications and grants. The United States, China, Switzerland, and the Netherlands were the best-evaluated states. On the other hand, poor efficiency was observed in the case of Brazil, Argentina, and Mexico.

The fifth article, "Pupils' School Performance and Their Cognitive Abilities to Solve Problems" written by Eva Hejnová, Petr Eisenmann, Lucie Loukotová, and Jiří Přibyl investigates the interrelationship between pupils' school grades in Czech language (native), mathematics, and physics and pupils' cognitive predispositions to problem-solving in science and mathematics diagnosed by the Lawson Classroom Test of Scientific Reasoning and the Culture of Problem-Solving test. The sample consisted of 180 students aged 14-15 from eight classes from six lower secondary schools and one class from an eight-year secondary grammar school in the Czech Republic. The results show that pupils with better grades in the monitored subjects achieve better results in both tests. It also turns out that there are generally statistically insignificant differences between the results of pupils assessed by grades 1 or 2 and between those assessed by grades 3 or 4.

The sixth article, "Collaborative Learning Based on Sophisticated Thinking Laboratory (STB-LAB) and Gather Town as Gamification Tool for Blended Laboratory on Science Undergraduate Student's" by Rena Denya Agustina, Riki Purnama Putra and Milla Listiawati, assesses the effectiveness of blended laboratory implementation using STB-LAB and Gather Town as a gamification tool, utilizing the Assessment Based on Teaching and Learning

Trajectory (AABTLT) with Student Activity Sheets (SAS). For this purpose, 122 students from the second semester of bachelor's studies in Physics and Biology education participated in the analysis. The results demonstrated that successful execution of STB-LAB syntax in blended laboratory activities leads to an increasing effect on collaborative skills when using the STB-LAB model.

The seventh article, "Individual Interest of Students in Physical Education and School Engagement in Fostering Physical Culture Inside the Campus" by Joseph Lobo and Genesis Dimalanta, focuses on determining the significant difference between the sex and institution of students concerning individual interest in physical education and school engagement fostering physical culture inside the campus. The sample consists of 1,659 undergraduate students from two colleges in the Pampanga region in the Philippines. The analysis indicates that the level of individual interest and school engagement has no significant difference with respect to gender and institution. However, individual interest has a significant relationship and direct effect on school engagement in Physical Education to foster Physical culture among students.

The last article, "Development of Computer-Based Chemical Five-Tier Diagnostic Test Instruments: A Generalized Partial Credit Model," by Achmad Rante Suparman, Eli Rohaeti, and Sri Wening, focuses on developing a five-tier chemical diagnostic test based on a computer-based test with 11 assessments categories with an assessment score from 0 to 10. A total of 20 items produced were validated by education experts, material experts, measurement experts, and media experts. The validation results were tested on 580 students from 19 schools in three regions of Indonesia and analyzed using the Generalized Partial Credit Model (GPCM) Item Response Theory (IRT) type. The analysis results show that all the items meet the requirements to be said to be valid for the model.

We would like to thank all authors who have submitted their manuscripts to the ERIES Journal, and special thanks to all reviewers for their effort in revising them. We hope all our readers will find this first issue of 2024 appealing, creating opportunities for future research. We also hope that the published articles will positively contribute to the field of efficiency and responsibility in education as it has been in recent years.

Sincerely

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

CONTENT

Full research paper

THE RELATIONSHIP BETWEEN HIGH SCHOOL STUDENTS' PERCEPTIONS OF ROLE MODELING AND SELF-REGULATION FOR SCIENCE

Melek Karaca[1]✉
Oktay Bektaş[1]
Seyide Eroğlu[2]

[1]Erciyes University, Turkey

[2]Ministiry of National Education, Turkey

✉ melekkaraca38@gmail.com

ABSTRACT

Role models are key to the development of self-regulation skills. The study aimed to develop a model that shows the relationship between high school students' role-modeling perceptions and self-regulation skills. In the study, the predictive correlation design of the quantitative research method was used. The study sample consisted of 362 high school students studying at a public high school and agreeing to participate in the research. The research data were collected with the "Perceived Role Models Scale" and the "Self-regulation Scale for Science". As a result, a model was developed showing that the reproduction, retention, and motivation dimensions of high school students' role modeling perceptions were statistically significant predictors of science self-regulation. This result proved that role models were important in developing high school students' self-regulation skills. Therefore, it could be argued that the development of self-regulation that begins at an early age is a lifelong process. The results of the study can be a reference point for research that will examine the relationship between the sub-dimensions of self-regulation and role modeling perception in depth.

KEYWORDS

Role model, self-regulation, science education, high school, structural equation model

HOW TO CITE

Karaca M., Bektaş O., Eroğlu S. (2024) 'The Relationship Between High School Students' Perceptions of Role Modeling and Self-Regulation for Science', *Journal on Efficiency and Responsibility in Education and Science*, vol. 17, no. 1, pp. 10-11. http://dx.doi.org/10.7160/eriesj.2024.170101

Article history
Received
March 10, 2023
Received in revised form
May 23, 2023
Accepted
September 4, 2023
Available on-line
March 31, 2024

Highlights

- *There is a significant relationship between high school students' role models and their self-regulation skills.*
- *Role models are important for the development of self-regulation skills in high school students.*
- *The sub-dimensions of the role model for high school students are also important for the development of self-regulation skill.*

INTRODUCTION

The framework of social cognitive theory includes understanding and changing human behavior. The theory asserts that individuals learn by using their cognitive characteristics as well as observing their environment (Bandura, 1986). Perhaps the greatest contribution of social cognitive theory to understanding human behavior and personality is indirect or observing learning, which is one of the basic principles of the theory. Bandura argues that learning is not limited to classical or operant conditioning. Learning can occur by observing, reading, or hearing other people's actions (Rutledge, 2000). According to Bandura, observers learn five things from the model: cognitive, affective, psychomotor skills, and value and belief systems. By observing others, individuals can learn new cognitive skills, such as decision-making and problem-solving, and psychomotor skills, such as dancing and riding a bicycle. Individuals' beliefs and value judgments can be strengthened or weakened as a result of observing the model. Observers can gain new values, beliefs, and ways of thinking as a result of modeling. They can also learn how to react to new situations encountered, as well as how the environment and property can be used. By observing the way, the model expresses their emotions, observers can learn to explain their emotions similarly.

The social cognitive theory argues that people can self-regulate their emotions, thoughts, and behaviors. Starting from this idea, the concept of self-regulation becomes an important concept. In the self-regulation process, mechanisms of control and management are put into action. The individual perceives his environment as the main source to realize his purpose (Locke & Latham, 2006).

Albert Bandura's social cognitive theory has led to the emergence of the concept of self-regulation. Based

on the therapy studies conducted by Bandura in 1977 with individuals having various phobias, the concept of self-regulation was developed (Berry & West, 1993). This is a cognitive and affective structure that includes skills such as self-regulation, symbolization, learning from others, strategy planning, self-regulation, and evaluation. This system functions as self-regulation by providing individuals with the ability to change their environment and control themselves (Pajares, 1996).

Learners with advanced self-regulation skills are individuals who have internalized the constructivist approach, who can analyze cognitive, behavioral, and environmental factors well in the process of reaching their own goals, identify situations that can help them reach their goals, adapt to changing conditions, and take an active role in the process. From this point of view, determining the effective and related situations in an individual's development of self-regulation skills and consciously organizing educational environments such as social environment and school according to these variables will contribute positively to the increase in the number of self-regulated learners.

Contrary to popular opinion, the process of acquiring self-regulation is not an internal process isolated from the social environment. The "self" used here refers to a process that requires personal initiative, perseverance, and harmony rather than an internal situation isolated from the social environment. According to social cognitive theory, individuals can learn through modeling by observing their environment (Bandura, 1977). Learning by modeling occurs as a result of interaction with role models that more individuals take as an example and identify with (Adesola et al., 2019). As a result of this interaction, people convey their attitudes, values, perspectives, and thoughts to each other and gain skills and competence (Rutledge, 2000). Role models are people whose specific goals, behaviors, and strategies are modeled by individuals. Also, role models are people who have a profound and significant impact on a person's life (Pell et al., 2022). Selected role models can be effective in the formation of the individual's cognitive, affective, or psychomotor skills (Dix et al., 2010).

Learning from role models may not always occur consciously. On the other hand, individuals living in the same environment may or may not see the same people as role models. The basis of this preference lies in individual perception. In terms of social learning, the model must be perceived correctly by the learner as well, and the right role model must be chosen. For example, while individuals at the beginning of their academic careers determine positive, close, and full-fledged role models, those in the middle or end of their careers tend towards specific and more negative role models (Gibson, 2003). In addition to the social status, psychological status, socioeconomic level of the role model, individual goals, and experiences obtained from the individual's previous life, reinforcements will affect the perception of role modeling (Hurd et al., 2011). Individuals with a developed perception of role modeling consciously prefer role models that will guide and assist them, especially in the process of achieving academically determined goals (Hackett et al., 1989). They do not always approve of all the features of the role model they have chosen as a result of this preference. Individuals with a developed perception of role modeling decide whether or not to take that behavior as an example after critically evaluating the behavior they observe according to their internal criteria (Hackett et al., 1989). Individuals choose, perceive, evaluate, interpret, decide, and implement the learned behavior when necessary. In summary, there is a dynamic relationship between individuals and the environment (Elkjaer, 2004).

Role models are an important resource for the individual's self-regulation development (Mueller et al., 2011; Urban et al., 2010). Exposure of learners to the right role model, knowingly or unknowingly, contributes to the development of self-regulation skills (Wirthwein et al., 2020). Therefore, if the relationship between the theoretically envisaged role model and self-regulation is understood more thoroughly, the models devised to develop self-regulation skills can be enriched. Following Zimmerman's self-regulated learning development model (2002), learners begin to learn by modeling and imitation, so they are still dependent on feedback from the environment. We can say that the students who can manage their self-learning have higher development characteristics than others. In these phases, some learners independently have metacognitive thinking about which strategy to use when, and how (Zimmerman, 2002). The purpose of this study is to reveal the importance of role models in meeting this need.

There are field studies examining the role model and self-regulation together (Acar et al., 2022; Augustine et al., 2022; Cai et al., 2022; Fung & Chung, 2021; Karaca & Bektas, 2021; Kurtoviç et al., 2021; Leslie, 2021; Ringoot et al., 2021; Vitiello et al., 2022; Xie & Li, 2022; Zielinska et al., 2022). Augustine et al. (2022) showed that family adjustment behavior enriched with self-regulation helps reduce anxiety problems. Karaca and Bektas (2021) applied the data collection tools used in this study to secondary school students. In the structural equation model the authors developed, they proved that the reproduction and retention dimensions of the perception of role modeling are important predictors of the development of students' self-regulation for science. In addition to these models, field studies state that self-regulation skills should be supported from an early age (Thomas et al., 2022; Xie & Li, 2022). Dignath and Büttner (2008) report that younger learners get more out of self-regulation training compared to high school and college students. On the other hand, it should not be overlooked that self-regulation is a lifelong process, even if the pace of development decreases (Kuhn, 1999). Learning is not limited to the school environment; the world has become a small village, and students need to develop their inner characteristics to make the right choices. Nevertheless, in our education system, affective features are not as important as cognitive features. Therefore, the purpose is to show high school students that it is important to set the right role model so that students can develop their self-regulation skills, regardless of their age. From this point of view, this study aims to develop a model that proves the relationship between high school students' role-modeling perceptions and self-regulation. Structural Equation Model (SEM) has a wide framework that takes into account the effect of latent variables, through which we can examine relational situations in depth and in multiple ways (Raykov & Marcoulides, 2006).

One of the SEM strategies is the model development strategy, in which the selected model is developed and supported by the data, considering that it best explains the relationship structure between the variables (Schermelleh-Engel & Moosbrugger, 2003). This study tried model probabilities until the model that expresses the possible relationship between self-regulation and role modeling perception most significantly and the latent variables that constitute them are obtained.

MATERIALS AND METHODS

Research Design

This study was conducted using the predictive correlation design of the quantitative research method. In predictive correlation studies, the relationships between variables are examined, and one of the variables is tried to be explained by the other (Fraenkel & Wallen, 1996). In this study, this design was preferred because high school students' perceptions of role modeling and self-regulation variables for science explain each other.

Population and Sample

The accessible population of the study was high school students studying in state high schools in Kayseri. Purposive sampling type of non-random sampling was used in the study. Purposive sampling is the selection of information-rich situations suitable for the study. It is preferred when it is desired to work with individuals with certain characteristics (Fraenkel & Wallen, 1996). In this study, high school students were preferred because the relationship between high school students' perception of role modeling and their self-regulation for science will be investigated. The study sample consisted of 362 high school students studying at a public high school and agreeing to participate in the research (Table 1). Turkish national education is based on the 4+4+4 system. At the end of their high school education, students prepare for university exams that greatly affect their future careers. Thus, they enter an intensive course of study at the 11th and 12th-grade levels. Therefore, the number of 11th and 12th-grade students participating in the study was less than that of 9th and 10th-grade students. There were 26 and 30 items in the scales used in the study. To contribute to the validity of the study results, it was tried to reach 10 times the number of items in the scale. CN (critical sample size) was also used to evaluate the adequacy of the sample size beyond model fit (Jöroskog & Sörbom, 1993). As a result of the SEM analyses developed in this study, the critical sample number for the study was CN = 154.31. Due to the pandemic, the application was made on the online platform, and sufficient sample numbers were reached.

Grade Level		Total	Gender
9	123	362	**Girl:** 216 **Boy:** 146
10	115		
11	81		
12	43		

Table 1: Sample of the study, 2017-2018, (source: own calculation)

Data Collection Tools

In order to check the validity and reliability of the data collection tools, a sample different from the research sample was used. In addition, data from different samples were used for EFA and CFA. The data relating to the scales used in the study were collected with the "Perceived Role Models Scale" and "Self-Regulation Scale for Science", which were developed during the first author's doctoral dissertation study, and validity and reliability checks were performed (Karaca & Bektas, 2022). To determine the scale items, the authors created a pool of questions based on the literature (Eker & İnce, 2018; Ilgaz & Gül, 2014; Kayan Fadlelmula, 2011; Zimmerman & Martinez-Pons,1986). Karaca and Bektas (2022) created self-regulation scale items based on Zimmerman's (1986) self-regulation model and strategies. Authors created role model scale items by taking into account the modeling processes (attention, retention, reproduction, motivation) and definitions (Malone, 2002; Rutledge, 2000) of Bandura's (1971) social learning theory, which forms the theoretical basis of the concept of role model. There were 51 items on the self-regulation draft scale for science and 49 items on the draft scale for determining the perception of role modeling. A pilot study was conducted with ten times the number of items in the draft scales (Mertens, 2019). In this context, draft scales were applied to 500 students (125 students in the fifth grade, 132 in the sixth grade, 127 in the seventh grade, and 116 in the eighth grade) studying in three secondary schools in the Melikgazi district of Kayseri in the spring term of 2017-2018 academic year. As a result of the application, the data, which were checked for completeness and objectivity, were entered into the SPSS 25 package program to conduct validity and reliability studies. After the data entry was completed, the authors assigned an average value to the items that were left blank because the missing data entry was below 5%. Reverse-coded items in the scale were recoded. The normal distribution of the scores obtained from the items prepared for analysis was checked. As a result of the reliability analysis, Cronbach's Alpha reliability coefficient of the self-regulation draft scale was .950. The Cronbach's Alpha reliability coefficient of the role modeling perception draft scale was calculated as .910. In addition, the authors decided which items should remain in the scale by checking the reliability coefficients of each item in the draft scales and the effect of the coefficient on the whole scale if they were removed from the scale. After examining the content validity of the draft scales, they calculated the item difficulty and discrimination indexes. After the content validity analysis, the authors checked the construct validity for both draft scales. Construct validity is the theoretical basis of the measurement tool (Cronbach & Meehl, 1955). Construct validity can be checked with factor analysis. After the explanatory factor analysis for both draft scales using the SPSS.25 program, the authors applied the relevant draft scales to a different sample from the sample in which the pilot study was conducted. Confirmatory factor analysis was performed using the LISREL 8.80 program with the obtained data. Within the scope of construct validity, the KMO value for the "Science-Oriented Self-regulation

Scale" was .952 and .922 for the "Perceived Role Models Scale". Since this value meant that factor analysis could be carried out, that the data were normally distributed, and that there was a sufficient sample, the authors switched to exploratory factor analysis (Pallant, 2020). As a result of repetitive factor analysis, items with overlapping extraction values below .30 or excluded in any factor were removed from the scales. Afterward, the authors continued the factor analysis with the remaining items in the "Self-Regulation Scale for Science". They determined that the remaining 26 items on the scale were grouped under three significant factors. When the distribution of the items to the factors using the Direct Oblimin vertical rotation technique was considered, the authors observed that the Eigenvalue was gathered in three factors greater than 1, and all the items had acceptable loading values in the factor they entered (the lowest item load value was .339; the highest item load value was .820). Similarly, they determined that the remaining 30 items in the "Perceived Role Models Scale" were grouped under three significant factors.

As a result, the "Self-regulation Scale for Science" was obtained, which explained 48% of the variance, had a reliability coefficient of .940, and consisted of 26 questions and three factors. The factors, names, and sample items of the scale are given in Table 2. Similarly, the "Perceived Role Models Scale" consisting of 30 questions and three factors, with a reliability coefficient of .911, explaining 41.96% of the variance was obtained. The factors were renamed, taking into account the item contents under the factors and the literature-supported factors initially predicted theoretically (Table 3).

Factors	Number of Items	Reliability Coefficient	Sample Items
Learning Strategies	2, 9, 11, 19, 25, 26, 32, 33, 34, 39, 40, 41, 44, 51	.905	I list and memorize important information related to the science course.
Critical Thinking	8, 17, 18, 20, 28, 35, 36	.808	I try to develop my ideas about what I learned in science class.
Regulating time and effort	14, 15, 31, 37, 43	.780	I use my study time efficiently for the science course.

Table 2: "Self-regulation Scale for Science" reliability values and sample items, 2017-2018, (source: own calculation)

Factors	Number of Items	Reliability Coefficient	Sample items
Reproduction	30, 13, 10, 5, 20, 16, 45, 24, 1, 42, 33,47	.876	I like it when my parents learn from the mistakes they've made in the past.
Retention	25, 9, 12, 15, 49, 36, 11, 8, 4, 19	.838	My friends' studies summarizing in science class attract my attention.
Motivation	41, 29, 35, 18, 26, 43, 23, 27	.752	I find it unnecessary to reward celebrities for their achievements.

Table 3: "Perceived Role Models Scale" reliability values and sample items, 2017-2018, (source: own calculation)

The accuracy of the factor structure revealed by the EFA was checked by confirmatory factor analysis (CFA) (Mertens, 2019) using LISREL 8.80 software. To obtain CFA data, again following the rule of 10, a total of 308 students studying at each level of a different secondary school from the pilot study in the seventh region of Melikgazi District of Kayseri in the 2018-2019 academic year were selected, a 26-item "Self-regulation Scale for Science" and the 30-item "Perceived Role Models Scale" were applied simultaneously. After the research data were complete and objective, they were transferred to the computer environment. Considering the factor structure determined by EFA, syntax commands were written, and CFA was performed. The authors examined the t values obtained as a result of the CFA, the factor loading value of each item, and the model fit indices (Jöroskog & Sörbom, 1993). The authors confirmed the factor structures obtained as a result of EFA with CFA.

Data Collection Process

Both scales used in the study were adapted to the Google Form format by the authors. The form consisted of three parts: demographic information, "Perceived Role Models Scale" and "Self-regulation Scale for Science". For the participants to answer each question and not leave it blank, the "required" tab was marked. After reaching a sufficient number of samples, the data obtained were transferred to the Excel program. After the demographic information was converted into numerical codes, participant answers were prepared for analysis. The normality of the data was checked with the SPSS 25 program. LISREL 8.80 program was used for SEM analysis.

Data Analysis

In the study, SEM's model development strategy was used to determine the relationship between high school students' role-modeling perceptions and their self-regulation levels toward science. SEM is a combination of factor analysis and regression analysis techniques applied to test causal and reciprocal relationships between variables (Raykov & Marcoulides, 2006). In this study, the SEM model development strategy was preferred since the possible relationship between the erception of role modeling and self-regulation variables and their sub-dimensions was investigated.

RESULTS

SEM Assumptions

In this section, it was primarily examined whether the data obtained from the self-regulation and role-modeling perception scales applied to the sample met the assumptions of SEM. To analyze the data of this study with SEM, independence of observations, random sampling of participants, linearity of the relationship between variables, univariate and multivariate normality, appropriate measurement level, and sufficient sample size were assumed (Reisinger & Turner, 2003; Tabachnick & Fidell, 1989). The independence of observations, which means that each observation or measurement is independent of the other, emerges as a basic requirement for almost every hypothesis testing (Gravetter & Wallnau, 2007). This study assumed that each participant answered both scales used in the research independently of each other.

The random sampling assumption, in which the participants are determined without a specific selection criterion, is important in terms of representativeness and generalizability (Gravetter & Wallnau, 2007). This study randomly selected participants studying at a public high school based on voluntariness.

In SEM analysis, there should be an assumption of linearity between both latent and observed variables. Violation of this assumption may indicate that the model fit and standard estimates are biased. One of the ways to check assumptions such as outlier, linearity, covariance, and independence of residuals is to check normality values (Pallant, 2020: 176). In SEM analysis, it is very important to control univariate and multivariate normality assumptions to determine the estimation method to be used during hypothesis testing. In the LISREL program, the generally preferred method is the Maximum Likelihood (EO) estimation (Jöroskog & Sörbom, 1993). To use this method, the variables must have a normal distribution. The multivariate normality assumption states that all univariate distributions are normal, each variable is normally distributed with the other, all bivariate graphs are linear, and residuals are covariate (Kline, 2011: 60). The LISREL program provides univariate and multivariate normality, which shows the skewness and kurtosis values for all variables measured in the model, with the chi-square test (Kunnan, 1998). The univariate normality test result is shown in Table 4.

Variables	Mean	Median	Minimum	Maximum	Skewness	Kurtosis
Learning strategies	56.60	56.00	28.00	70.00	-.41	-.30
Critical thinking	27.19	27.00	14.00	35.00	-.12	-.46
Regulation time and effort	18.36	18.00	7.00	25.00	-.06	-.54
Reproduction	47.87	48.00	20.00	60.00	-.71	.37
Retention	38.09	39.00	15.00	50.00	-.39	-.29
Motivation	29.83	31.00	12.00	40.00	-.88	1.13
Role model	115.78	117.50	61.00	150.00	-.40	-.02
Self-regulation	102.15	102.00	55.00	130.00	-.11	-.51

Table 4: The univariate normality test result, 2017-2018, (source: own calculation)

Findings Related to the Research Model

After the SEM assumptions were checked, model development studies were started. According to the results of exploratory and confirmatory factor analysis, three factors of role modeling perception (reproduction, retention, motivation) and three factors of self-regulation for science (learning strategies, critical thinking, time and effort regulation) were defined as latent variables, a structure consisting of a total of six latent variables, 56 indicator variables and 362 observations was studied. While developing the model, first of all, the χ^2/df ratio was checked, and then other fit indices were examined. In the first experiment, a model was studied in which three sub-dimensions of the role model predicted the three sub-dimensions of self-regulation. Although all of the t values obtained in this model were shown with black arrows, it was determined that the standardized coefficients were not in the desired range. The possible reason for this situation was multicollinearity. When the correlation coefficients of both scales within their dimensions were examined, it was determined that the correlation between the sub-dimensions of the self-regulation scale was high. To solve the multicollinearity problem, self-regulation was made into a single variable. The model in which the three dimensions of the perception of role modeling predicted the self-regulation variable was tested, and it was found that there was no red arrow in the t values (Figure 1), and the standardized coefficients were in the range of -1/+1 (Figure 2).

We reported the fit of the model in Table 5 by utilizing the goodness of fit indices commonly used in the literature (Jöroskog & Sörbom, 1993). According to Kline (2011), it is good if the ratio of the Chi-square value to degrees of freedom is less than five; if it is below three, it means perfect harmony. In this case, we can assert that the developed model has an acceptable fit.

In the SEM obtained as a result of the study, it was found that each of the three sub-dimensions of role modeling perception predicted high school students' self-regulation skills (25% reproduction, 44% retention, and 11% motivation) and explained 46% of the variance.

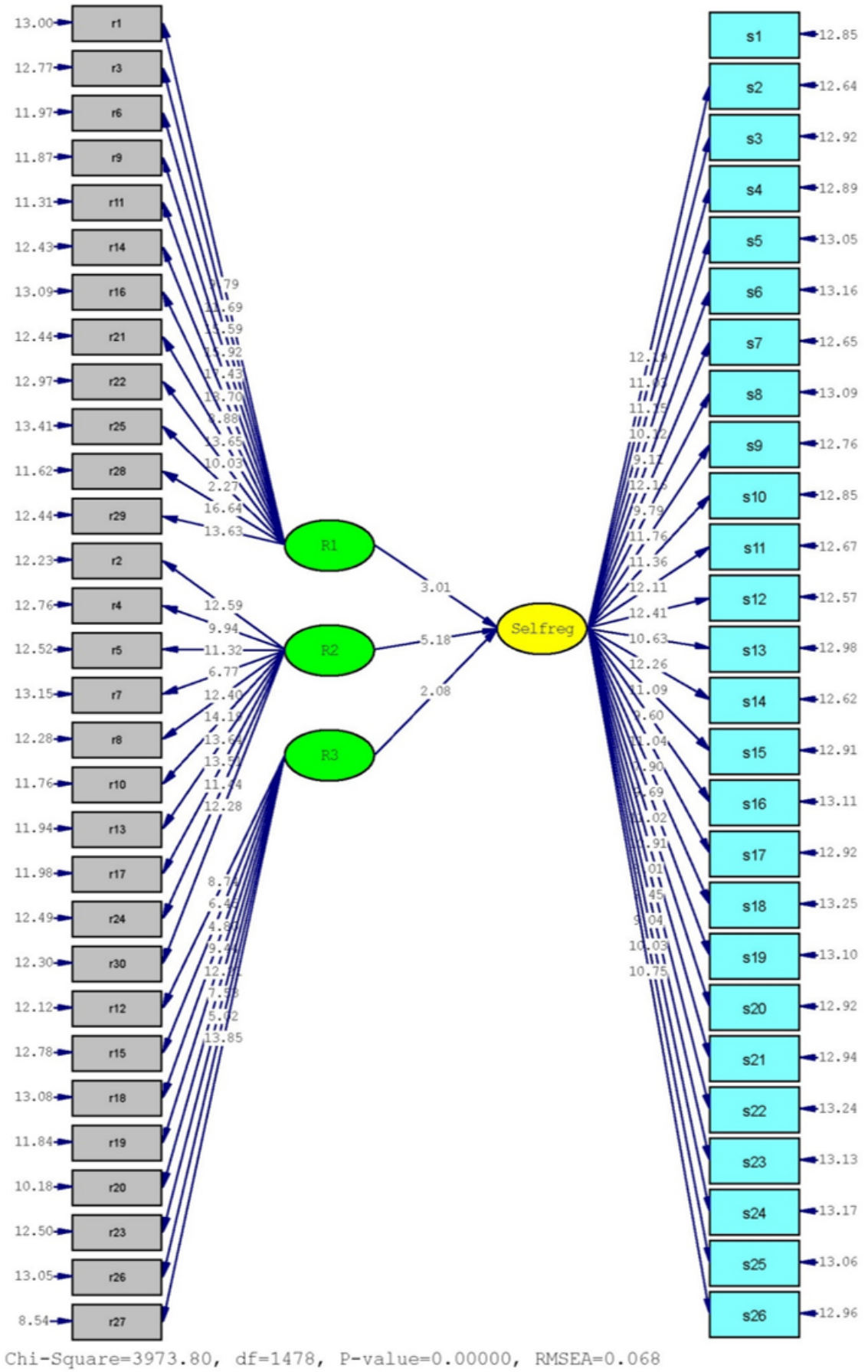

Figure 1: SEM t values, 2017-2018, (source: own calculation)

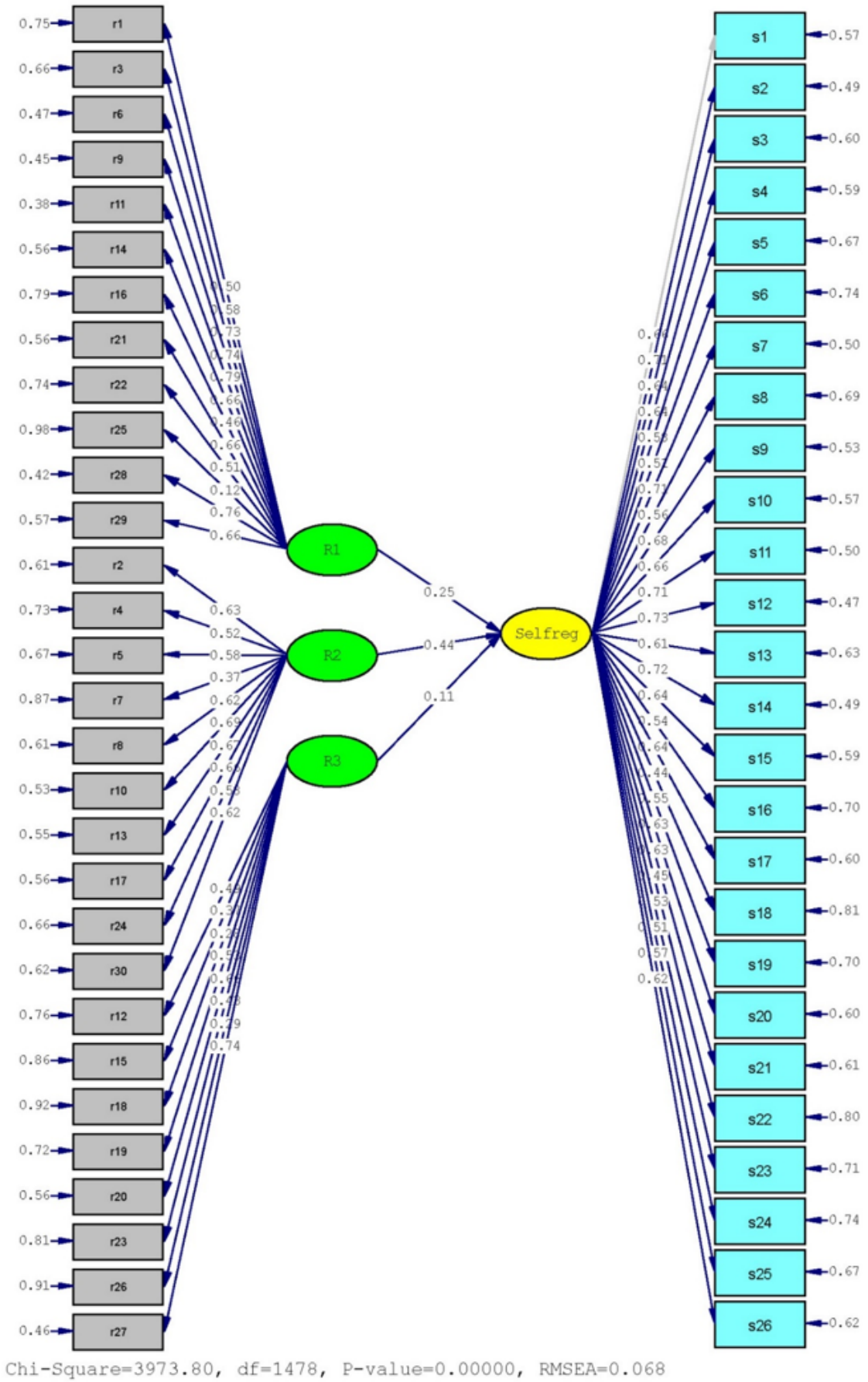

Figure 2: SEM standardized coefficient values, 2017-2018, (source: own calculation)

Fit indices	Acceptable	Excellent	Value on the scale	The fit of the scale
	χ^2/df		2.68	Excellent
NFI	.90 and above	.95 and above	.91	Acceptable
NNFI	.90 and above	.95 and above	.94	Acceptable
IFI	.90 and above	.95 and above	.94	Acceptable
RFI	.90 and above	.95 and above	.91	Acceptable
CFI	.95 and above	.97 and above	.94	Reject
GFI	.85 and above	.90 and above	.72	Reject
AGFI	.85 and above	.90 and above	.70	Reject
SRMR	Between = .05 and = .08	Between = .00 and < .05	.080	Acceptable
RMSEA	Between = .05 and = .08	Between = .00 and < .05	.068	Acceptable

Table 5: Structural equation model goodness of fit indices, 2017-2018, (source: own calculation)

DISCUSSION

Motivation and education programs are very effective in the development of children's self-regulation skills (Montroy et al., 2016). As mentioned before, the perceived social environment has an important place for the development of self-regulation skills (Bronson, 2000). As competence levels of learners increase, the role model builds a kind of scaffolding (Wood et al., 1976) by transferring the responsibility to them and slowly withdrawing their support. These scaffolds are more robust in situations such as environments where the active participation of the learner is ensured (Kangas, 2016) and seen as a stakeholder (Flekkøy & Kaufman, 1997), which is important for the development of self-regulation skills, including the learner in assessment processes (Dinsmore & Wilson, 2016). On the other hand, it is known that children start and continue their learning processes as they reach higher levels, and adult support decreases (Lundy, 2007; Smith et al., 2002). It is also known that the development of self-regulation in individuals is not directly proportional to their biological age. Therefore, the reason why the sub-dimensions of the role model could not explain the sub-dimensions of self-regulation in the first model tried in this study may be that high school teachers do not use the necessary strategies for the development of students' affective characteristics and do not show sufficient support. The reason why high school teachers have such an attitude may be that they believe students' affective characteristics are like adults because they are at the age to go to university after a few years, they often ignore that they are still children, and they are not aware that self-regulation skill development will continue throughout life even if the pace slows down. Teachers should be role models for their students to gain self-regulation skills (Siddiqui & Habib, 2021). Teachers should realize that they are role models in raising students with skills such as questioning, debating, logical inference, and problem-solving (Ashton, 1988; Wilks, 2018).

The combined use of motivation and self-regulation strategies helps to create self-regulated learning (Pintrich, 2000). One of the reasons for the differences in the results of these two studies, in which the same scales were used, maybe the age of the students who made up the sample. Self-regulation includes reflexive and learned responses, whether desired or not (Strauman, 2017). It is known that younger ages are more important for the development of these skills (Ringoot et al., 2021; Thomas et al., 2022). On the other hand, studies report that older children exhibit higher self-regulation skills than younger children (Vitiello et al., 2022; von Suchodoletz et al., 2013). In addition, experience is also an important variable for the development of self-regulation skills (Alexander et al., 1998). Moreover, children often find it difficult to generalize the use of experienced strategies to new contexts (Alexander et al., 1998). Therefore, another reason for the difference in the results of these two studies conducted in the same socioeconomic and sociocultural region may be students' different individual experiences regarding self-regulation development. Therefore, looking at these two models, it could be argued that self-regulation skills, which include internal and external processes, continue to develop throughout life, even if the pace changes depending on different variables.

Although self-regulation skills are defined as an individual process, research has documented that the social dimension is undeniably effective (Ijaz et al., 2022). By its nature, the relationship between the perception of role modeling and self-regulation is intertwined with the environment. This relationship has an important role in lifelong learning (Broda et al., 2020; Lenes et al., 2020; Lerner et al., 2021). Lerner et al. (2021) stated that role models can help learners develop their character traits and increase their awareness, especially during adolescence. Fung and Chung (2021) stated that the role model parent response is important in the self-regulation development of preschool children. There are many studies in the literature proving that role models are especially effective in the development of children's self-regulation (Acar et al., 2022; Augustine et al., 2022; Broda et al., 2020; Cai et al., 2022; Fung & Chung, 2021; Karaca & Bektas, 2021; Kurtovic et al., 2021; Lenes et al., 2020; Lerner et al., 2021; Leslie, 2021; Ringoot et al., 2021; Vitiello et al., 2022; Xie & Li, 2022; Zieli'nska et al., 2022).

CONCLUSION

In this study, reproduction, retention, and motivation dimensions of high school students' role modeling perceptions did not predict the sub-dimensions of self-regulation learning strategies, critical thinking, and time-effort management for science.

In light of the analyses, another model in which the dimensions of the perception of role modeling predicted the whole self-regulation was tried and statistically verified. Karaca and Bektas (2021) developed a model by applying the scales used in this study to secondary school students. As a result of their study, the authors confirmed the model showing that

secondary school students' reproduction and retention skills predicted self-regulation skills for science. This model proved that as the number of positive role models in a student's social environment increases, it becomes easier for them to develop self-regulation skills. On the other hand, the same model shows that motivation does not predict self-regulation. In the model, the authors developed in light of the data obtained from high school students that all three dimensions of the perception of role modeling predicted self-regulation separately.

The model developed as a result of this research enriched the results obtained from field studies by revealing that role models are important for the development of self-regulation skills of adolescent high school students. Therefore, being aware of the fact that they are role models, high school teachers need to use strategies to develop students' self-regulation skills, ensure active participation, and support students effectively until they reach a sufficient affective level. Enrichment of education and training programs with self-regulation will facilitate the work of program practitioners. The model obtained in this study also shows that individuals in the social environment should be aware that they are candidate role models for the learners around them. As this awareness matures, the social learning network that Bandura foresees will become widespread, and it will become easier and faster for children to gain self-regulation skills.

The limitations of the study and the suggestions put forward in the light of the results are listed below.

1. The sample of this study is limited to students studying at a public high school in Kayseri, Türkiye. Since self-regulation is affected by environmental factors, this research can be conducted with high school students studying in different regions. The results can be expanded by conducting this study, which was carried out in an Anatolian high school, with high school students at different academic levels studying in private high schools, vocational high schools, and science high schools.
2. This study was conducted based on the quantitative research method. Especially with the extreme values in the research, it is possible to interview by using the qualitative research method. Thus, the opportunity to examine the possible causes of the results in more detail is obtained.
3. The model developed in this study can be extended by including demographics (gender, class level, education levels of mother and father), socioeconomic, and sociocultural characteristics.
4. In the study, the sub-dimensions of the perception of role modeling predicted the whole self-regulation but did not predict the sub-dimensions of self-regulation. The possible reasons for this result can be investigated in detail using the qualitative research method.

REFERENCES

Acar, I. H., Veziroglu Celik, M., Garcia, A., Colgrove, A., Raikes, H., Gonen, M., and Encinger, A. (2019) 'The Qualities of Teacher-Child Relationships and Self-Regulation of Children at Risk in The United States and Turkey: The Moderating Role of Gender', *Early Childhood Education Journal,* Vol. 47, No. 1, pp. 75-84. https://doi.org/10.1007/s10643-018-0893-y

Adesola, S., den Outer, B., and Mueller, S. (2019) 'New Entrepreneurial Worlds: Can The Use of Role Models in Higher Education Inspire Students? The Case of Nigeria', *Journal of Entrepreneurship in Emerging Economies,* Vol. 11, No. 4, pp. 465-491. http://dx.doi.org/10.1108/JEEE-08-2018-0076

Alexander, P. A., Graham, S., and Harris, K. R. (1998) 'A Perspective on Strategy Research: Progress and Prospects' *Educational psychology review,* Vol. 10, No. 2, pp. 129-154. https://doi.org/10.1023/A:1022185502996

Ashton, P. (1988) *Teaching Higher-Order Thinking and Content: An Essential Ingredient in Teacher Preparation*, Gainesville, FL: University of Florida.

Augustine, D. A., Koss, K. J., Smith, E. P., and Kogan, S. M. (2022) 'The Influence of Family Cohesion on Self-Regulation and Anxiety Problems Among African American Emerging Adults', *PLOS ONE,* Vol. 17, No. 1, e0261687. https://doi.org/10.1371/journal.pone.0261687

Bandura, A. (1971) *Social Learning Theory,* Morristown.: General Learning Press.

Bandura, A. (1977) 'Self-Efficacy: Toward a Unifying Theory of Behavioral Change', *Psychological Review,* Vol. 84, No. 2, pp. 191-215. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1986) 'The Explanatory and Predictive Scope of Self-Efficacy Theory', *Journal of Social and Clinical Psychology,* Vol. 4, No. 3, pp. 359-373. https://doi.org/10.1521/jscp.1986.4.3.359

Berry, J. M., and West, R. L. (1993) 'Cognitive Self-Efficacy in Relation to Personal Mastery and Goal Setting Across the Life Span', *International Journal of Behavioral Development*, Vol. 16, No. 2, pp. 351-379. https://doi.org/10.1177/016502549301600213

Broda, M., Ekholm, E., and Zumbrunn, S. (2020) 'Assessing the Predictive Nature of Teacher and Student Writing Self-Regulation Discrepancy', *Frontline Learning Research,* Vol. 8, No. 4, pp. 52-73. https://doi.org/10.14786/flr.v8i4.505

Bronson, M. B. (2000) 'Recognizing and Supporting the Development of Self-Regulalion in Young Children', *Young Children*, pp. 32-37.

Cai, J., Wen, Q., Jaime, I., Cai, L., and Lombaerts, K. (2022) 'From Classroom Learning Environments to Self-Regulation: The Mediating Role of Task Value', *Studies in Educational Evaluation,* Vol. 72, 101119. https://doi.org/10.1016/j.stueduc.2021.101119

Cronbach, L. J., and Meehl, P. E. (1955) 'Construct validity in psychological tests', *Psychological Bulletin*, Vol. *52*, No. 4, pp. 281-302. https://doi.org/10.1037/h0040957

Dignath, C., and Büttner, G. (2008) 'Components of Fostering Self-Regulated Learning Among Students. A Meta-Analysis on Intervention Studies at Primary and Secondary School Level', *Metacognition and Learning,* Vol. 3, No. 3, pp. 231-264. https://doi.org/10.1007/s11409-008-9029-x

Dinsmore, D. N. L., and Wilson, H. E. (2016) 'Student Participation in Assessment: Does it Influence Self-Regulation?', in Brown, G. and Harris, L. (ed.) *Handbook of Human and Social Conditions in Assessment*, Oxfordshire: Routledge.

Dix, S., Phau, I., and Pougnet, S. (2010) '"Bend it like Beckham": The Influence of Sports Celebrities on Young Adult Consumers', *Young Consumers*, Vol. 11, No. 1, pp. 36-46. https://doi.org/10.1108/17473611011025993

Eker, C., and İnce, M. (2018) 'Teacher Opinions on Use of Self-Regulation Strategies at Science and Art Centers: Turkey Sample', *Universal Journal of Educational Research,* Vol. 6, No. 7, pp. 1486-1496. https://doi.org/10.13189/ujer.2018.060709

Elkjaer, B. (2004) 'Organizational learning: the 'third way', *Management Learning*, Vol. *35*, No. 4, pp. 419-434. https://doi.org/10.1177/1350507604048271

Flekkøy, M. G., and Kaufman, N. H. (1997) *The Participation Rights of The Child: Rights and Responsibilities in Family and Society*, London: Jessica Kingsley Publishers.

Fraenkel, J., and Wallen, N. (1996) 'Valididity and Reliability', in Fraenkel, J., Wallen, N. and Hyu, H. (ed.) *How to Design and Research in Education*, New York: McGraw-Hill.

Fung, W. K., and Chung, K. K. H. (2021) 'The Roles of Social Mastery Motivation and Parental Response in Preschoolers' Vocabulary Knowledge and Self-Regulation', *Early Child Development and Care,* Vol. 191, No. 1, pp. 21-35. https://doi.org/10.1080/03004430.2019.1596902

Gibson, D. E. (2003) 'Developing the Professional Self-Concept: Role Model Construals in Early, Middle, and Late Career Stages', *Organization science,* Vol. 14, No. 5, pp. 591-610. https://doi.org/10.1287/orsc.14.5.591.16767

Gravetter, F., and Wallnau, L. (2007) *Statistics for The Behavioral Sciences*, Belmont: Thomson Learning.

Hackett, G., Esposito, D., and O'Halloran, M. S. (1989) 'The Relationship of Role Model Influences to The Career Salience and Educational and Career Plans of College Women', *Journal of Vocational Behavior,* Vol. 35, No. 2, pp. 164-180. https://doi.org/10.1016/0001-8791(89)90038-9

Hurd, N. M., Wittrup, A., and Zimmerman, M. A. (2011) 'Role models in adolescent development', in Levesque, R. J. R. (ed.) *Encyclopedia of Adolescence,* New York: Springer.

Ijaz, I., Zubair, A., and Saeed, Q. (2022) 'Relationship Self-Regulation and Dyadic Coping in Married Women: Role of Demographic Factors', *Pakistan Journal of Psychological Research,* Vol. 37, No. 1, pp. 119-133. https://doi.org/10.33824/pjpr.2022.37.1.08

Ilgaz, G., and Gül, A. (2014) 'İlköğretim İkinci Kademe Öğrencilerinin Fen ve Teknoloji Dersi Öz-Düzenlemeli Öğrenme Stratejilerinin İncelenmesi', *Ondokuz Mayıs University Journal of the Faculty of Education,* Vol. 33, No. 1, pp. 465-465. https://doi.org/10.7822/egt207

Jöroskog, K., and Sörbom, D. (1993) *Lisrel 8: Structural Equation Modeling with The Simplis Command Languag*', Lincolnwood: Scientific Software International, Inc.

Kangas, J. (2016) *Enhancing Children's Participation in Early Childhood Education Through Participatory Pedagogy*, [Ph.D thesis], Helisinky: University of Helsinky.

Karaca, M., and Bektas, O. (2021) 'The Relationship Between Perceived Role Models and Self-Regulation in Science', *International Journal of Educational Research,* Vol. 110, p. 101884. https://doi.org/10.1016/j.ijer.2021.101884

Karaca, M., and Bektas, O. (2022) 'Self-Regulation Scale for Science: A Validity and Reliability Study', *International Journal on Social & Education Sciences,* Vol. 4, No. 2, pp. 236-256. https://doi.org/10.46328/ijonses.302

Kayan Fadlelmula, F. (2011) 'The Effect of Gender on Elementary Students' Goal Orientations, Use of Learning Strategies, and Mathematics Achievement', *International Journal of Learning,* Vol. 18, No. 1, pp. 337-346.

Kline, R. B. (2011) *Principles and Practice of Structural Equation Modeling*, 3rd Edition, New York: Guilford.

Kuhn, D. (1999) 'A Developmental Model of Critical Thinking', *Educational Researcher,* Vol. 28, No. 2, pp. 16-46. https://doi.org/10.3102/0013189X028002016

Kunnan, A. J. (1998) 'An Introduction to Structural Equation Modelling for Language Assessment Research', *Language Testing,* Vol. 15, No. 3, pp. 295-332. https://doi.org/10.1177/026553229801500302

Kurtović, A., Vrdoljak, G., and Hirnstein, M. (2021) 'Contribution to Family, Friends, School, and Community is Associated with Fewer Depression Symptoms in Adolescents Mediated by Self-Regulation and Academic Performance', *Frontiers in Psychology,* Vol. 11, 615249. https://doi.org/10.3389/fpsyg.2020.615249

Lenes, R., Gonzales, C. R., Storksen, I., and McClelland, M. M. (2020) 'Children's Self-Regulation in Norway and the United States: The Role Of Mother's Education and Child Gender Across Cultural Contexts', *Frontiers in Psychology,* Vol. 11, pp. 1-17. https://doi.org/10.3389/fpsyg.2020.566208

Lerner, J. V., Wong, C. A., Weiner, M. B., and Johnson, S. K. (2020) 'Profiles of Adolescent Character Attribute Associations with Intentional Self-Regulation and Character Role Model Relationships', *Journal of Moral Education,* Vol. 50, No. 3, pp. 293-316. https://doi.org/10.1080/03057240.2020.1755242

Leslie, A. M. (2021) 'Rendering Latinas invisible: The Underrepresentation of Latina Role Models in K-12 History Textbooks', *International Journal of Multicultural Education,* Vol. 23, No. 1, pp. 87-109. https://doi.org/10.18251/ijme.v23i1.2103

Locke, E. A., and Latham, G. P. (2006) 'New Directions in Goal-Setting Theory', *Current directions in psychological science,* Vol. 15, No. 5, pp. 265-268. https://doi.org/10.1111/j.1467-8721.2006.00449.x

Lundy, L. (2007) 'Voice'is Not Enough: Conceptualising Article 12 Of The United Nations Convention on The Rights of The Child', *British Educational Research Journal,* Vol. 33, No. 6, pp. 927-942. https://doi.org/10.1080/01411920701657033

Malone, Y. (2002) 'Social Cognitive Theory and Choice Theory: A Compatibility Analysis', *International Journal of Reality Therapy,* Vol. 22, No. 1, pp. 10-13.

Mertens, D. M. (2019) *Research and Evaluation in Education and Psychology: Integrating Diversity with Quantitative, Qualitative, and Mixed Methods*, Thousand Oaks: Sage Publications.

Montroy, J. J., Bowles, R. P., Skibbe, L. E., McClelland, M. M., and Morrison, F. J. (2016) 'The Development of Self-Regulation Across Early Childhood', *Developmental Psychology,* Vol. 52, No. 11, pp. 1744-1762. https://doi.org/10.1037/dev0000159

Mueller, M. K., Phelps, E., Bowers, E. P., Agans, J. P., Urban, J. B., and Lerner, R. M. (2011) 'Youth Development Program Participation and Intentional Self-Regulation Skills: Contextual and Individual Bases of Pathways to Positive Youth Development', *Journal of Adolescence,* Vol. 34, No. 6, pp. 1115-1125. https://doi.org/10.1016/j.adolescence.2011.07.010

Pajares, F. (1996) 'Self-efficacy Beliefs in Academic Settings', *Review of Educational Research,* Vol. 66, No. 4, pp. 543-578.

Pallant, J. (2020) *SPSS User Guide Step-By-Step Data Analysis with SPSS*, Ankara: Anı Publishing.

Pell, B., Hawkins, J., Cannings-John, R., Charles, J. M., Hallingberg, B., Moore, G., and Morgan, K. (2022) 'Choosing active role models to Inspire Girls (CHARMING): protocol for a cluster randomized feasibility trial of a school-based, community-linked programme to increase physical activity levels in 9-10-year-old girls', *Pilot and Feasibility Studies*, Vol. 8, No.1, pp. 1-14. https://doi.org/10.1186/s40814-021-00961-6

Pintrich, P. R. (2000) 'The Role of Goal Orientation in Self-Regulated Learning', In Boekaerts, M., Pintrich, P. R. and Zeidner, M. (ed.) *Handbook of Self-Regulation*, Amsterdam: Elsevier.

Raykov, T. and Marcoulides, G. A. (2006) 'On Multilevel Model Reliability Estimation from the Perspective Of Structural Equation Modeling', *Structural Equation Modeling,* Vol. 13, No. 1, pp. 130-141. https://doi.org/10.1207/s15328007sem1301_7

Reisinger, Y., and Turner, L. W. (2003) *Cross-cultural Behaviour in Tourism: Concepts and Analysis*, Abingdon: Routledge.

Ringoot, A. P., Jansen, P. W., Kok, R., van IJzendoorn, M. H., Verlinden, M., Verhulst, F. C., Bakermans-Kranenburg, M., and Tiemeier, H. (2021) 'Parenting, Young Children's Behavioral Self-Regulation, and The Quality of Their Peer Relationships', *Social Development*, Vol.31, No. 3, pp. 715-732. https://doi.org/10.1111/sode.12573

Rutledge, K. (2000) '*Social Learning Theory - Notes on Ormond's Psychology of Learning*'

Schermelleh-Engel, K., Moosbrugger, H., and Müller, H. (2003) 'Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures', *Methods of psychological research online,* Vol. 8, No. 2, pp. 23-74.

Siddiqui, R., and Habib, Z. (2021) 'Moral Education at Primary Level in Selected Private Schools of Karachi: Role of Teachers and Parents', *Pakistan Journal of Humanities and Social Sciences*, Vol. *9*, No. 2, pp. 59-73. https://doi.org/10.52131/pjhss.2021.0901.0113

Smith, S. G., English, R., and Vasek, D. (2002) 'Student and Parent Involvement in The Transition Process for College Freshmen with Learning Disabilities', *College Student Journal,* Vol. 36, No. 4, pp. 491-504.

Strauman, T. J. (2017) 'Self-Regulation and Psychopathology: Toward an Integrative Translational Research Paradigm', *Annual Review of Clinical Psychology,* Vol. 13, pp. 497-523.

Tabachnick, B. G., and Fidell, L. (1989) *Using Multivariate Statistics,* New York: Harper & Row.

Thomas, V., Peeters, J., De Backer, F., and Lombaerts, K. (2022) 'Determinants of Self-Regulated Learning Practices in Elementary Education: A Multilevel Approach', *Educational Studies,* Vol. 48, No. 1, pp. 126-148. https://doi.org/10.1080/03055698.2020.1745624

Urban, J. B., Lewin-Bizan, S., and Lerner, R. M. (2010) 'The Role of Intentional Self-Regulation, Lower Neighborhood Ecological Assets, and Activity Involvement in Youth Developmental Outcomes', *Journal of Youth and Adolescence,* Vol. 39, No. 7, pp. 783-800. https://doi.org/10.1007/s10964-010-9549-y

Vitiello, V. E., Nguyen, T., Ruzek, E., Pianta, R. C., and Whittaker, J. V. (2022) 'Differences Between Pre-K and Kindergarten Classroom Experiences: Do They Predict Children's Social-Emotional Skills and Self-Regulation Across The Transition to Kindergarten?', *Early Childhood Research Quarterly,* Vol. 59, pp. 287-299. https://doi.org/10.1016/j.ecresq.2021.11.009

von Suchodoletz, A., Gestsdottir, S., Wanless, S. B., McClelland, M. M., Birgisdottir, F., Gunzenhauser, C., and Ragnarsdottir, H. (2013) 'Behavioral Self-Regulation and Relations to Emergent Academic Skills Among Children in Germany and Iceland. *Early Childhood Research Quarterly,* Vol. 28, No. 1, pp. 62-73. https://doi.org/10.1016/j.ecresq.2012.05.003

Wilks, S. (2018) 'Resourceful teachers and teacher resources', in Burgh, G. and Thornton, S. (ed.) *Philosophical Inquiry with Children: The development of an inquiring society in Australia*, Oxforshire: Routledge.

Wirthwein, L., Sparfeldt, J. R., Heyder, A., Buch, S. R., Rost, D. H., and Steinmayr, R. (2020) 'Sex Differences in Achievement Goals: Do School Subjects Matter?', *European Journal of Psychology of Education,* Vol. 35, No. 2, pp. 403-427. https://doi.org/10.1007/s10212-019-00427-7

Wood, D., Bruner, J. S., and Ross, G. (1976) 'The Role of Tutoring in Problem Solving', *Child Psychology & Psychiatry & Allied Disciplines,* Vol. 17, No. 2, pp. 89-100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Xie, S., and Li, H. (2022) 'Self-Regulation Mediates The Relations Between Family Factors and Preschool Readiness', *Early Childhood Research Quarterly,* Vol. 59, pp. 32-43. https://doi.org/10.1016/j.ecresq.2021.10.005

Zielińska, A., Lebuda, I., Ivcevic, Z., and Karwowski, M. (2022) 'How Adolescents Develop and Implement Their Ideas? On Self-Regulation of Creative Action', *Thinking Skills and Creativity*, Vol. 43, 100998. https://doi.org/10.1016/j.tsc.2022.100998

Zimmerman, B. J. (1986) 'Becoming A Self-Regulated Learner: Which Are The Key Subprocesses?', *Contemporary Educational Psychology,* Vol. 11, No. 4, pp. 307-313. https://doi.org/10.1016/0361-476X(86)90027-5

Zimmerman, B. J. (2002) 'Becoming A Self-Regulated Learner: An Overview', *Theory into Practice,* Vol. 41, No. 2, pp. 64-70. https://doi.org/10.1207/s15430421tip4102_2

Zimmerman, B. J., and Pons, M. M. (1986) 'Development of A Structured Interview for Assessing Student Use of Self-Regulated Learning Strategies', *American Educational Research Journal,* Vol. 23, No. 4, pp. 614-628. https://doi.org/10.3102/00028312023004614

Full research paper

INTERNAL IMAGE OF CZECH TERTIARY BUSINESS SCHOOLS AND ITS INFLUENCE ON THE INTEREST OF NEW APPLICANTS FOR STUDY

Jana Pavelková[1]✉
Jana Turčínková[2]
Jakub Šácha[3]

[1]Mendel University in Brno, Czech Republic

[2]Mendel University in Brno, Faculty of Business and Economics, Department of Marketing and Trade, Czech Republic

[3]Mendel University in Brno, Faculty of Business and Economics, Czech Republic

✉ xpavelko@mendelu.cz

ABSTRACT

Attracting prospective students could be difficult and expensive. Candidates, when choosing a future school for their studies, consider a great number of factors. Information from official university websites does not suffice to persuade. Word of mouth plays a significant role, among others. The aim of this paper is to present findings about important factors influencing the overall satisfaction of current students with university life as well as their willingness to share positive references. The study took place in the Czech Republic with students attending business schools at universities. Primary data was collected via an online questionnaire with students with bachelor's, master's, and doctorate degrees (n = 274) and in-depth interviews (n = 10). Data was processed with regression analysis and Spearman's correlation coefficient. The main findings suggest that there is a close positive correlation between students' satisfaction and the willingness to recommend the university. The quality of student life and the reputation of the school were identified as the most important factors influencing this satisfaction and willingness to recommend.

KEYWORDS

References, satisfaction, university students, WOM

HOW TO CITE

Pavelková J., Turčínková J., Šácha J. (2024) 'Internal Image of Czech Tertiary Business Schools and Its Influence on The Interest of New Applicants for Study', *Journal on Efficiency and Responsibility in Education and Science*, vol. 17, no. 1, pp. 12-22. http://dx.doi.org/10.7160/eriesj.2024.170102

Article history
Received
July 7, 2022
Received in revised form
March 22, 2023
Accepted
September 12, 2023
Available on-line
March 31, 2024

Highlights

- *Influence of students' satisfaction on willingness to recommend the university to potential students.*
- *The key factors:*
  - *The comparison of the university with its competition and the university's reputation (from the area of the institution's reputation and image).*
  - *The attractiveness of courses and the teachers' attitude towards students (from the area of study factors).*

INTRODUCTION

Schools invest considerable amounts of money in their promotion to attract more students. No more than three schools or universities typically make it to a student's final shortlist, thanks to marketing activities (Caffee, 2017). Yet not all institutions are fully aware of the power of word of mouth (WOM) from their current students. Referrals are important not just as feedback for the institution to improve its services and lead it towards further innovations, but they can also become an important tool of promotion (Jalkala and Terho, 2014; Siering et al., 2018). Often, referrals also play an important role in decision-making because consumers are more likely to choose a product or service when that particular product is recommended to them by a trusted friend (He et al., 2016).

When choosing a university, students are influenced by a great number of factors. In general, tuition, school fees, and location rank amongst the most influential factors (Kinzie et al., 2004; Drewes, Michael, 2006). Nowadays, these aspects still play a role in decision-making. However, other factors, such as the learning environment, future job prospects (Agrey and Lampadan, 2014), and the university's reputation (Gamoga and Ambang, 2020), are all considered. Research by Schlesinger et al. (2021) suggests that alumni satisfaction and identification with the university and the university's brand image are also key factors for recommendations.

In the Czech Republic, where most students attend public schools, and, thanks to the country's relatively small size, location is not as important as in countries like the U.S., the promotion, reputation, and internal image of universities may have a significant impact on students' decision making.

Referrals from current students could be a source of important information for prospective students when choosing a university, as this information is actually personal and could be considered more reliable than the official information presented by the university

itself. Shields and Peruta (2019) found in their research in the USA that 55% of students find speaking to current students as one of the primary sources helping them choose a school. Referrals from students could be reliant on lots of partial factors connected with student life at the university. Examples include the appearance and amenities of the university campus, the teachers and their approach to students, the applicability of knowledge earned at the university to practical life, the university's culture (including communication between management and students), etc.

The objective of the paper is to determine which factors could have a major influence on the overall impression and satisfaction of current students and, therefore, could also influence their recommendation to prospective students. The connection of the identified factors and the willingness to spread positive WOM will be assessed.

LITERATURE REVIEW

Importance of brand image and internal image

The image of any organization contains several elements (Avenarius, 1993). The most important of these elements are reputation, the degree of being known, and the specific profile. Thanks to global integration and business competition, companies are encouraged to pay more attention to their brand image and its potential (Alhaddad, 2015). According to Wood (2004), a strong brand image enables the creation of a strong relationship between customers and companies.

There are two important relationships connected with brand image: one with brand trust, which leads to advocacy intention, and the second with repurchase intention (Huang et al., 2020). A successful brand can be created through brand experience because a positive experience with the brand leads to brand satisfaction (Brakus et al., 2009). According to Dass et al. (2021), intellectual brand experience is the most important for achieving brand love and trust in higher education. Also, sensory and affective dimensions in brand experience and behavioral dimensions have a high impact.

The brand image also plays an important role in differentiating brands from one another (Anwar and Jalees, 2020). It is formed through various instances of communication that create associations. All this leads to the creation of a certain perception in the minds of consumers (Dülek and Saydan, 2019). Companies should pay attention to their brand image, which is constructed from their visual, product, and service images (Huang et al., 2020). A well-chosen logo could also help a company distinguish itself from competitors (Erjansola et al., 2021) and with self-expression (Park et al., 2013). In the creation of a strong logo and brand, Sadeghvaziri et al. (2022) suggest universities focus on functional, visual, and self-expressive aspects and, thus, stimulate students to create a strong emotional bond. This premise is based on the tendency to connect high-quality products with the attractive visual effect of the logo. In higher education, branding strategies mostly concentrate on advertising activities, aiming to gain new students (Sujchaphong et al., 2020). Leonnard (2019) mentioned the importance of improving and increasing the online presentation for universities (such as websites, social media, etc.). His findings confirmed the positive relationship between loyalty and e-trust.

Internal branding is also of great importance for a company, as employees' brand-supportive behavior leads to strengthening the corporate brand (Hoppe, 2018). Internal image is linked with employee satisfaction and loyalty (Hejlová, 2015). With more and more organizations on the market, it is not easy to retain employees. Employees' motivation plays a key role here (Mehta et al., 2010), and it is precisely motivation and organizational commitment that lead to employee loyalty (Khuong et al., 2020). The construct of satisfaction and loyalty leading to long-term relationships is also valid for the school environment (Borraz-Mora et al., 2020). The identification of students with their schools takes place thanks to the sharing of values and attributes, due to which the students can psychologically attach themselves to their schools (Bhattacharya, Sen, 2003). According to Nguyen et al. (2016), a successful school brand signifies the ability to fulfil students' needs, leads to trust in inadequate services, and helps in making school- and course-related decisions.

Today's society cannot escape the impact of the social factors that have been promoted through globalization. In this way, people are exposed to global brands (Wu et al., 2019). Social media and networking also play an important role in everyday life. People get used to using social media not just for communication but also become more and more interested in receiving gratification through it rather than in person (Phua et al., 2017). In 2019, in the European Union, almost 88% of young people aged 16-24 used social networks (CZSO, 2020b), and social networks are an important channel for reaching new potential students. For students, it is now easier to communicate and share their opinions about their university via social media than to do so face-to-face.

Recommendation and WOM

Gaining trust is a key element in customer relationship management. Customers who believe in the company and its products are more willing to share a positive experience with friends and relatives (Sernovitz, 2009; He et al., 2016; Eldegwy et al., 2018). Another factor that plays an important role in sharing positive recommendations is consumer satisfaction. Kotler (2007) states that customers whose expectations of a company are met, or the company even exceeds their expectations, are more loyal and speak positively about the company. In their research on hospital employees, the relationship between satisfaction and positive recommendations can also be found in employee satisfaction and willingness to give recommendations, as shown by Grass et al. (2021). Each consumer is affected by certain reference groups, whether or not they are members and their family. Close surroundings present the greatest influence (Novotny and Duspiva, 2014). Consumers' opinions change through the transmission of information, beliefs, attitudes, and behaviors (Wu et al., 2019). According to Shen and Sengupta (2018), consumers present their personalities through consumption and discussing brands.

Consumers often rely on WOM during purchase decisions to help reduce uncertainty and the level of perceived risk (Murray, 1991). Positive recommendations spread spontaneously by WOM then influence customers' purchase intentions. If their experience matches the positive review, they will then spontaneously generate further positive recommendations (Barreda et al., 2015). Sharing recommendations amongst consumers is way more effective than traditional advertising techniques (Haikel-Elsabeh et al., 2019), and

they also help define the problems and functionality of institutional services, even potentially serving as an improvement of such services (Jalkala and Terho, 2014).

In their research, Ghosh et al. (2001) established that the students' trust in their alma mater leads to the willingness to advocate for their school in front of other people and to customer advocacy. According to Sarkar and Sarkar (2016), a consumer who has a strong connection with the brand and trusts the brand deeply is not only spreading positive WOM but also advocating this brand to attract new consumers to consume the brand as well. Schlesinger et al. (2021) showed in their research that the key to alumni recommending their alma mater through WOM is identification, satisfaction, and the university's brand image.

School choice

In the U.S., in the past, some of the most important factors impacting the choice were the distance from home, tuition, and compulsory fees (Drewes and Michael, 2006). Also, Judson et al. (2006) considered these factors to be the most influential while also adding the factors of image and reputation. In their research in Thailand, Agrey and Lampadan (2014) determined the learning environment and potential future job prospects as the most important factors influencing school choice as factors applicable in an international institution. Another factor they mentioned was the institution's reputation. Dirin et al. (2021) point out the importance of using relevant digital channels, which are used by students to search for information.

According to Safari et al. (2020), teachers also play a key role in the education system. Berková et al. (2020) point out the increasing importance of the implementation of entrepreneurship education in higher education programs. This could be accomplished by the employment of entrepreneurs to lessons.

In their study, Misran et al. (2012) recommend that schools promote intensively and spread more information about the school via mass media and by appointing an ambassador who would share what campus lifestyle is like and students' overall experience in high school.

Universities around the world are assessed according to various factors and lined up to rankings. Examples of international rankings are The Times Higher Education World University Rankings (THE) or the U.S. News & World Report (USNWR) (Dearden et al., 2019). Each raking uses its own criteria. For example, THE ranking assesses the quality of teaching, science, and research, citations, public opinion, companies' opinions, etc. (THE, 2021). Apart from other factors, the USNWR (2021) evaluates universities according to regional reputation indicators, citations, and research. University rankings create an opportunity to attract prospective students as they create prestige and provide information for students about each universities' attributes (Dearden et al., 2019).

Joseph et al. (2012) confirmed that branding efforts are important during the student's search process, wherein making the final choice, experiential factors such as personal visits to the campus, interactions with students and university representatives, and WOM from family members and friends play the key role. In the research from Shields and Peruta (2019), institutional websites, campus visits, and speaking with current students are the primary sources of information received by students about a school in the USA.

MATERIAL AND METHODS

For the purposes of this research, in-depth interviews ($n = 10$) with business school students from Czech universities were conducted, followed by a questionnaire survey ($n = 271$). The interviews served as preliminary research and to gain a deeper understanding of the topic. Students from public universities in the Czech Republic, which have a dominant position in the market, students from business schools of the Czech Republic, as well as students from private schools, were targeted. In total, there were respondents from 19 public and 2 private business schools from the Czech Republic. The aim was to address respondents in quota according to the level of their current study (see Table 1).

	Bachelor's	Master's	Doctor's
Statistics from the Czech Statistical Office	63%	32%	5%
Respondents	59%	37%	4%

Table 1: Compliance with the quota (source: CZSO, 2020a and questionnaire sample demographics)

The respondents evaluated 50 factors related to students' satisfaction with the university they attended (see Table 2). These factors represent the following areas selected based on the literature survey:

- learning environment (Agrey and Lampadan, 2014);
- institution reputation and image (Judson et al., 2006);
- factors related to students' studies (Safari et al., 2020; Berková et al., 2020).

Multivariate regression analysis was used to assess the importance of these factors. The regression analysis estimates the relationship between two variables - the response variable (explained) and the explanatory variables (Evangelos, 2010). The variables explained are:

- students' satisfaction with their institution (the university they currently attended);
- willingness to recommend their institution to secondary school students (potential applicants).

Some of the questions were intentionally asked twice with minor modifications, first at the beginning of the questionnaire before evaluating the set of satisfaction factors and for a second time at the end of the questionnaire. The aim was to compare how the top-of-the-mind (quick-thinking) assessment of their satisfaction and willingness to recommend would differ from their assessment after a more thorough reflection on these factors. To assess the dependence between the individual variables explained, the Spearman correlation coefficient was used.

All variables in the models were rated either on a Likert scale of 1 to 5 (1 = highest satisfaction, 5 = lowest) or are binary variables (e.g., involvement of experts from practice in courses 1 = yes, 0 = no). For binary variables, a higher rating (1) means higher satisfaction, so it can be assumed that these variables will be negatively correlated with the explained variables. Statistically insignificant

variables at the 10% significance level were removed from the models by gradual sequential elimination. The models were further subjected to econometric verification and adjusted based on the results.

Variable	Average	Median	Mode	Mode frequency	Standard deviation
Comparison of the institution with the competition	2.21	2	2	105	0.968
University reputation	2.20	2	2	106	1.160
Availability of study rooms	1.25	1	1	213	0.582
Availability of IT equipped study rooms	1.33	1	1	229	0.955
Availability of library	1.69	2	2	166	0.543
Availability of dormitories	3.08	3	3	178	0.702
Availability of school canteen	1.98	2	2	166	0.911
Availability of cafe	1.25	1	1	231	0.742
Availability of sports grounds	3.45	4	4	111	0.953
Availability of gym	3.35	4	4	107	1.094
Availability of relaxation area	1.93	2	1	119	1.125
Indoor school environment	.1.87	2	2	111	0.887
Outdoor campus environment	1.99	2	1	108	1.029
Quality of school canteen	2.67	3	2	89	1.109
Organization of workshops	0.80	1	1	217	0.400
Organization of concerts	0.18	0	0	222	0.386
Bonus lectures outside regular courses	0.83	1	1	224	0.379
Organization of job fairs	0.72	1	1	196	0.448
Organization of parties	0.30	0	0	191	0.457
University logo	2.27	2	2	111	1.042
Promotional items	2.37	2	3	98	0.914
Feeling proud of the university	2.14	2	2	104	1.023
Feeling proud of the school	2.26	2	2	100	1.064
University relationship towards students	2.37	2	2	114	0.983
Assessment of graduates by employers	2.14	2	2	88	1,000
PR and media	2.37	2	2	100	0.964
Usefulness of subjects	2.48	2	2	124	0.938
Study plan	2.61	3	2	100	1.001
Level of exam difficulty	2.88	3	3	98	0.896
Mediation of internships	2.87	3	3	85	1.190
Individual approach to students	2.77	3	3	94	1.174
Attitude of study counselors	2.25	3	2	108	1.056
Troubleshooting while studying	2.23	2	2	97	0.996
Student organizations	0.31	0	0	188	0.462
Quality of foreign language teaching	2.66	3	2	97	1.051
Opportunity to communicate with international students	2.56	3	3	76	1.212
Study abroad options	1.75	1	1	140	0.916
Level of fun during lessons	2.81	3	3	111	0.938
Attractiveness of courses	2.45	2	2	113	0.889
Expert knowledge of teachers	1.89	2	2	121	0.821
Willingness of teachers to consult	1.99	2	2	112	0.894
Teachers' approach towards students	2.40	2	2	117	0.921
Presentation skills of teachers	2.41	2	2	122	0.792
Use of practical examples	0.70	1	1	190	0.459
Involvement of experts from practice in courses	0.70	1	1	190	0.459
Projects in cooperation with companies	0.24	0	0	206	0.428
Possibility of work in real projects	0.17	0	0	224	0.379
Competitions	0.13	0	0	235	0.340
Simulation games	0.23	0	0	209	0.421
Availability of e-learning	0.53	1	1	143	0.500

Table 2: Variables used in models (source: Questionnaire 2020, *n* = 271)

RESULTS

Satisfaction is one of the crucial factors influencing one's willingness to recommend (Schlesinger et al., 2021). Thus, the first model (see Table 3) explains the students' satisfaction with the institution (university) they attend depending on the observed factors. After sequential elimination, the resulting model contains 5 variables and explains 35% of the variability. The variables were sorted according to the size of the regression coefficient, the size of which characterizes the degree of influence of the given variable on students' satisfaction. For example, for the variable "comparison with the competition", the value of the regression coefficient of 0.29 can be interpreted that an increase of such a factor by one point on the Likert scale translates to an increase in the student's satisfaction with their school by an average of 0.29.

This model, as with all presented models, was tested for the assumptions of the linear model, all assumptions were fulfilled (at a significance level of 0.01), and the multicollinearity among the explanatory variables was not demonstrated (see Table 8 in Appendix 1).

Model 1: OLS using observations 1-271 Dependent variable: Satisfaction					
	coefficient	std. error	*t*-ratio	*p*-value	
Const	0.332	0.181	1.831	0.0683	*
Comparison of the institution with its competition	0.291	0.065	4.478	< 0.001	***
Attractiveness of courses	0.221	0.066	3.350	0.0009	***
University reputation	0.151	0.047	3.230	0.0014	***
University relationship towards students	0.127	0.055	2.239	0.0260	**
Study plan	0.114	0.055	2.073	0.0391	**
Mean dependent var.		2.446	S. D. dependent var.		0.968
Sum squared resid.		160.500	S. E. of regression		0.778
R-squared		0.656	Adjusted R-squared		0.354

*Note: *, **, *** indicates the level of significance (0.1; 0.05; 0.01)*

Table 3: Model 1 - students' satisfaction (source: Questionnaire 2020, *n* = 271)

Based on the results of the regression model, only two of the three originally defined areas are significantly important for students' satisfaction with their university, namely, the institution's reputation and image and factors related to the provided education.

The area of an institution's reputation and image is represented by three factors. First, "Comparison of the institution with its competition" is also the factor with the highest impact on students' satisfaction. Comparing a school with competing institutions can be important given the number of universities providing education in business and economics on the Czech market and the availability of information about the school, its awards, placement in university rankings, etc. The second factor in this area was the "university reputation." The last factor in this area is the university's relationship with the students, the way they are treated, and how they are perceived.

The importance of the university's position in comparison to competing institutions and the reputation of the university were also confirmed by the results of in-depth interviews, where two factors were highlighted in particular: 1. whether the content of courses allows easy use of attained know-how in real life and 2. free time activities organized by the university and student organizations operating at the university.

The university's reputation and its position within the market are quite often important for future good job prospects, as Agrey and Lampadan (2014) also pointed out as one of the most important factors influencing school choice.

As expected, another area represented by statistically significant factors is factors related to the students' studies. The most important factor in this area is "the provision of attractive courses". This is understandable, as finding something attractive could mean a positive experience with this subject, which leads to customer satisfaction (Brakus et al., 2009).

It is somewhat surprising that this model lacks factors related to relationships and communication between teachers and students that were also assessed in detail by students in the questionnaire survey.

The importance of the second factor of this group, which is the "study plan" (representing the study plan students must pass in their selected study programs), points to the students' awareness of the study content and its importance. They perceive the need for a comprehensive and practical approach to the study plan, allowing the transfer of acquired knowledge in real life, not just a formal requirement to earn their degree.

Using the same methodology, a regression analysis, the dependence of the willingness to provide a recommendation of the student's home institution to secondary school students was examined. In this model, significant factors are reflected in all three areas (see Table 4).

In the area of institution reputation and image, the factors "comparison of the institution with its competition" and "university reputation" are repeated again. The significantly higher value of the regression coefficient for the factor of "comparison of the institution with its competition" compared to the previous model is worth noting. This difference is not surprising, as we can be more inclined to make more objective evaluations of universities when giving recommendations than when evaluating one's own satisfaction with the university, meaning a better position of the university compared to the competition will play a more important role in one's willingness to give recommendations.

Again, there were other factors from the area of "factors related to the students' studies" in the model. Compared to the

Model 2: OLS using observations 1-271 Dependent variable: Recommendation1					
	coefficient	std. error	*t*-ratio	*p*-value	
const	-0.341	0.274	-1.248	0.2132	*
Comparison of the institution with its competition	0.655	0.067	9.794	< 0.0001	***
Involvement of experts from practice in courses	-0.248	0.110	-2.251	0.0252	**
Teachers approach towards students	0.194	0.0561	3.467	0.0006	***
University reputation	0.133	0.050	2.640	0.0088	***
Availability of sports grounds	0.103	0.053	1.934	0.0542	*
Mean dependent var.		2.044	S. D. dependent var.		1.101
Sum squared resid.		182.053	S. E. of regression		0.829
R-squared		0.444	Adjusted R-squared		0.434

*Note: *, **, *** indicates the level of significance (0.1; 0.05; 0.01)*

Table 4: Model 2 - the students' willingness to recommend at the beginning of the questionnaire (source: Questionnaire 2020, *n* = 271)

previous model of students' satisfaction, this time, the factor is related to the "teachers' approach towards students". The inclusion of this factor may mean that students take the teachers' dignity and communication more into account when providing recommendations about an institution. The factor "the involvement of experts from practice in courses" is a binary variable, thanks to which the negative sign of the regression coefficient expresses a positive dependence. The participation of practitioners in teaching was also mentioned and positively perceived by respondents of in-depth interviews.

Compared to the previous model, this model also includes a factor from the area of "learning environment", specifically the availability of sports activities at the university campus. At first glance, this is more of an additional factor in this area, but for some students, the availability of nearby sports facilities can be very important.

At the end of the questionnaire, respondents answered a question about their overall impression of their school. This question was a modification of the evaluation of the students' satisfaction with their university. Again, there was also a question about one's willingness to recommend their own institution to secondary school students; however, this time, with an emphasis on taking all the assessed areas into consideration.

Model 3: OLS using observations 1-271 Dependent variable: Overall impression					
	coefficient	std. error	t-ratio	p-value	
const	-0.079	0.137	-0.574	0.5662	
Attractiveness of courses	0.207	0.046	4.540	< 0.0001	***
Student organizations	-0.202	0.071	-2.848	0.0047	***
Study plan	0.148	0.039	3.834	0.0002	***
Promotional items	0.135	0.038	3.587	0.0004	***
Troubleshooting while studying	0.113	0.037	3.130	0.0019	***
University relationship towards students	0.104	0.042	2.497	0.0131	**
Teachers' approach towards students	0.093	0.047	1.992	0.0474	**
Outdoor campus environment	0.091	0.035	2.627	0.0091	***
Assessment of graduates by employers	0.081	0.035	2.307	0.0218	**
Mean dependent var.		2.151	S. D. dependent war.		0.791
Sum squared resid.		69.426	S. E. of regression		0.516
R-squared		0.589	Adjusted R-squared		0.575

*Note: *, **, *** indicates the level of significance (0.1; 0.05; 0.01)*

Table 5: Model 3 - the students' overall impression (source: Questionnaire 2020, *n* = 271)

The evaluation of the overall impression of the university is influenced by several factors. In contrast to students' satisfaction, all three areas are represented in the model (see Table 5).

In the area of "institution reputation and image," this time, "comparison of the institution with its competition" and the "university reputation" are missing among the significant factors; only the "university relationship towards students" remained statistically significant. New significant factors from this area are the "assessment of graduates by employers" and satisfaction with the school's "promotional items". The in-depth respondent emphasized the role of promotional items in building a sense of belonging to the university.

No factor was eliminated in the area of "factors related to the students' studies"; on the contrary, several factors were added. This may indicate that respondents, after deeper consideration, took the practical aspects of the provided education more into account. The "student organizations" factor was again a binary variable, so a negative sign of the regression coefficient is expected in the model.

From the third area, "learning environment", not represented in the model explaining students' satisfaction, the factor of the external environment of the university (outdoor campus environment) appears among the significant variables.

Model 4: OLS using observations 1-271 Dependent variable: Recommendation2					
	coefficient	std. error	*t*-ratio	*p*-value	
const	-0.460	0.266	-1.728	0.0852	*
Projects in cooperation with companies	-0.249	0.098	-2.535	0.0119	**
Willingness of teachers to consult	0.198	0.062	3.169	0.0017	***
Teachers' approach towards students	0.185	0.062	2.981	0.0031	***
Bonus lectures outside regular courses	-0.179	0.108	-1.664	0.0973	*
Mediation of internships	0.159	0.040	3.992	< 0.0001	***
Assessment of graduates by employers	0.144	0.047	3.062	0.0024	***
Study abroad options	0.126	0.048	2.610	0.0096	***
Comparison of school with competition	0.126	0.050	2.528	0.0121	**
Availability of dormitories	0.114	0.059	1.946	0.0527	*
Outdoor campus environment	0.112	0.044	2.534	0.0119	**
Mean dependent var.		2.011	S. D. dependent var.		0.944
Sum squared resid.		105.972	S. E. of regression		0.641
R-squared		0.557	Adjusted R-squared		0.539

*Note: *, **, *** indicates the level of significance (0.1; 0.05; 0.01)*

Table 6: Model 4 - the students' willingness to recommend at the end of the questionnaire (source: Questionnaire 2020, *n* = 271)

The last model resulting from the regression analysis explained which factors influence one's "willingness to give a recommendation", this time after considering all the partial factors (see Table 6).

Amongst the significant variables, factors from all three assessed areas are repeated. As expected, when compared to the previous model (one's willingness to give recommendation 1), there is a higher number of significant factors. For example, from the area "learning environment", the factor "dormitories" has been added, as their availability in the vicinity of the university campus is assessed as positive and plays an important role in giving a recommendation.

We can see that when students were evaluated at the end of the questionnaire, they took into account the particular factors they had to evaluate during the survey. There are more statistically important factors as well as some more practically focused ones.

We assumed minor differences between the explanatory variables in all the models above at the beginning of the research. These differences were demonstrated in individual models, where the explanatory variables differ slightly. Spearman's correlation coefficient was used to determine individual dependencies (see Table 7).

Variable	Spearman correlations			
	Satisfaction	Recommendation1	Overall impression	Recommendation2
Satisfaction	1.000	0.595	0.441	0.368
Recommendation1	0.595	1.000	0.364	0.521
Overall impression	0.441	0.364	1.000	0.641
Recommendation2	0.368	0.521	0.641	1.000

Table 7: Spearman correlations (source: Questionnaire 2020, *n* = 271)

The values of the Spearman coefficients show a higher correlation between the variables that were evaluated in the same part of the questionnaire survey (satisfaction × recommendation 1; overall impression × recommendation 2) than between the individual modifications of the variables (satisfaction × overall impression; recommendation 1 × recommendation 2).

DISCUSSION

The aim of this study was to examine the factors influencing student satisfaction in relation to willingness to give recommendations. The study confirmed the importance of image and reputation (Judson et al., 2006; Agrey and Lampadan, 2014) as one of the key factors in student satisfaction, which, in consequence, affects recommendations and willingness to give those recommendations (Kotler, 2007, Schlesinger et al., 2021). On the other hand, the relationships with the learning environment, as the literature suggests (Judson et al., 2006), were not found. This could be thanks to focusing on current student, who are probably already used to the university campus and, thus, considering it as one of the less important factors for their lives at university. As the campus is often visited during open days, it is still important to keep it up for a positive first impression (Shields and Peruta, 2019).

As Schlesinger et al. (2021) discovered, alumni tend to recommend their alma mater through WOM if they feel identification and satisfaction. We can conclude it applies also to business school students who have not graduated yet. Our results support the findings of Ghosh et al. (2001), who linked students' trust in their alma mater and their positive WOM, as also published by Sarkar and Sarkar (2016).

The importance of teachers' willingness to consult and teachers' approach towards students in students' satisfaction leading to higher willingness to share good references are in line with the results of García and García (2021), who highlight that good teachers are one of the top factors influencing students' academic success. Teaching quality as crucial for competitiveness is also stressed by Fajčíková and Fejfarová (2019). Orientation on practice in provided education as an important factor for students' satisfaction was also proven by Depoo et al. (2022). Practically

oriented education was also proven earlier by Berková et al. (2018) as an important motivator.

This research also has its limitations. It is focused only on students from universities in the Czech Republic. Data collection was disrupted by the situation caused by precautions due to the COVID-19 pandemic, which complicated the involvement of a larger number of universities, especially private universities. In one way, it could be convenient thanks to the popularity of public universities, thus avoiding the influence of factors such as school tuition and other fees. On the other hand, the burden of tuition could affect the motivation of students and, thus, make some other factors reconsidered. Future studies in this area should consider repeating the research with a higher representation of private schools and in other countries and comparing the results. Another aspect that could have influenced the obtained data is the time-consuming nature of the questionnaire, which had 41 questions with a total of 211 possible answers, and it took respondents an average of 10 minutes to complete. It is, therefore, possible that they could already have been losing attention at the end of the questionnaire.

As research on high school students has shown, the referral factor of current students plays a significant role in choosing a college. Universities should, therefore, strive to make the references of their current students as positive as possible.

Based on our findings, it follows that the school should focus on the following:

1. Attention to feedback from students and staff, suggestions for improvement, and motivation of teachers, as relationships between students and staff (variables: the willingness of teachers to consult, teachers' approach towards students) resulted in significant results in the formation of positive WOM.
2. Development of cooperation with companies, i.e., creation of a functional database and communication with companies from various industries, informing educators about new collaborations, creating an internship portal for students, and promoting it not only among students but also among companies (based on the significance of variables: projects in cooperation with companies, mediation of internships, involvement of experts from practice in courses).
3. Building the school's reputation and sense of belonging, i.e., high-quality and visually appealing university-wide promotional items, efforts, and activities that will help achieve the highest possible placement in world and national university rankings (based on the significance of a variable university reputation).
4. Development of relationships with students, especially better communication, i.e., via social networks, a well-designed and clear website, and understandable presentations of schools' achievements (based on the significance of variable university relationships towards students).
5. Building a friendly and comfortable campus and student residencies, i.e., a pleasant atmosphere while visiting the campus, providing the needs of students outside of class time - relaxation places and the capacity of study rooms (based on the significance of variable outdoor campus environment, availability of dormitories).
6. Support of extracurricular activities at school, i.e., regular communication with associations and student bodies, support of their activities, and use of them as a communication channel with students (based on the significance of variable student organizations).
7. Appropriate communication with potential applicants, i.e., providing teaching/learning experience to high school students, promoting cooperation with the private sector, building alumni clubs (based on the significance of variables: assessment of graduates by employers, involvement of experts from practice in the course and comparison of the institution with its competition).
8. The motivation of current students to provide references, i.e., supportive attitude of teachers and university staff in teaching and communication with students and others (based on the significance of variables: university relationship towards students, the attractiveness of courses).

CONCLUSION

This study was conducted to evaluate the importance of different areas of student life at universities and the importance of these areas for students' satisfaction and willingness to give recommendations to prospective students.

The analysis confirmed the significant importance of factors in the areas of the institution's reputation and image and factors related to the student's studies. Factors from these two areas were included in all four models. The most significant factors were the comparison of the university with its competition and the university's reputation from the area of the institution's reputation and image, the attractiveness of courses, and the teachers' attitude towards students from the area of study factors.

The comparison of factors influencing satisfaction and one's willingness to recommend at the beginning of the survey and the end of the survey brought interesting results. It is obvious that after deeper consideration, the students involved more aspects in their assessment of their university, especially in terms of their personal experience with their education and the school environment. In the quick assessment, the students mostly evaluated their university according to external factors (such as the university's reputation and the comparison of the university with its competition), even though internal factors still played an important role.

The motivation for this paper was to explore what drives students' recommendations besides the varying levels of tuition and other costs related to higher education, especially since there is such a large number of students studying at public universities in the Czech Republic. The tuition and costs at Czech public universities are still, to this day, quite low, and therefore, not the issue influencing the choice. Our research has proven that the role of WOM is more significant, and universities should focus on the analyzed areas impacting the rate of positive recommendations.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the financial support of the Internal Grant Agency project of the Faculty of Business and Economics of Mendel University in Brno (PEF_DP_2021018).

REFERENCES

Agrey, L. and Lampadan, N. (2014) 'Determinant factors contributing to student choice in selecting a university', *Journal of Education and Human Development*, Vol. 3, No. 2, pp. 391- 404.

Alexopoulos, E., C. (2010) 'Introduction to Multivariate Regression Analysis', *Hippokratia*, Vol. 14, pp. 23-8.

Alhaddad, A. (2015) 'A structural model of the relationships between brand image, brand trust and brand loyalty', *International Journal of Management Research & Review*, Vol. 5, pp. 137-144.

Anwar, A. and Jalees T. (2020) 'Brand Orientation and WOM: Mediating Roles of Brand Love', *Journal of Management Sciences,* Vol 7, No. 1, pp. 14-30. https://doi.org/10.20547/jms.2014.2007102

Avenarius, H. (1993) 'Introduction: Image and Public Relations Practice', *Journal of Public Relations Research*, Vol. 5, No. 2, pp. 65-70. https://doi.org/10.1207/s1532754xjprr0502_01

Barreda, A. A., Bilgihan, A., and Kageyama, Y. (2015) 'The Role of Trust in Creating Positive Word of Mouth and Behavioral Intentions: The Case of Online Social Networks.' *Journal of Relationship Marketing*, Vol. 14, No. 1, pp. 16-35. https://doi.org/10.1080/15332667.2015.1006002

Berková K., Frendlovská D., Pospíšil Závodný J., Vojáčková H., Kolářová D. (2020) 'Education towards Entrepreneurial Careers in a Czech College: An Empirical Study', *Journal on Efficiency and Responsibility in Education and Science*, Vol. 13, No. 1, pp. 1-9. https://doi.org/10.7160/eriesj.2020.130101

Bhattacharya, C., and Sen, S. (2003) 'Consumer-company identification: A framework for Understanding consumers' relationships with companies', *Journal of Marketing*, Vol 67, No. 2, pp. 76-88. https://doi.org/10.1509/jmkg.67.2.76.18609

Borraz-Mora, J., Hernandez-Ortega, B. and Melguizo-Garde M. (2020) 'The Influence of Generic-Academic Competences on Satisfaction and Loyalty: The View of Two Key Actors in Higher Education', *Journal of Higher Education Policy*, Vol. 42, No. 5, pp. 563-578. https://doi.org/10.1080/1360080X.2019.1689802

Brakus, J. J., Schmitt, B. H., and Zarantonello, L. (2009) 'Brand experience: What is it? How is it measured? Does it Affect Loyalty?', *Journal of Marketing*, Vol. 73, No. 3, pp. 52-68. https://doi.org/10.1509/jmkg.73.3.052

Caffee, A. (2017) 'College admissions report'. https://articles.niche.com/the-classof-2017-goes-to-college/

CZSO (2020a) *Czech Statistical Office | Students and graduates of universities and colleges in the Czech Republic: An analysis*, [Online], Available: https://www.czso.cz/csu/czso/vzdelavani-analyzy-komentare [4 November 2020].

CZSO (2020b) *Czech Statistical Office | Use of information and communication technologies in households and among individuals*, [Online], Available: https://www.czso.cz/csu/czso/5-pouzivani-mobilniho-telefonu-a-internetu-na-mobilnim-telefonu [8 April 2021].

Dass, S., Popli, S., Sarkar, A., Sarkar, J. G. and Vinay M. (2021) 'Empirically examining the psychological mechanism of a loved and trusted business school brand', J*ournal of marketing for higher education*, Vol. 31, No. 1, pp. 23-40. https://doi.org/10.1080/08841241.2020.1742846

Dearden, J. A., Grewal, R. and Lilien, G. L. (2019) 'Strategic Manipulation of University Rankings, the Prestige Effect, and Student University Choice', *Journal of Marketing Research*, Vol. 56, No. 4, pp. 691-707. https://doi.org/10.1177/0022243719831258

Depoo, L., Urbancová, H. and Smolová, H. (2022) 'Factors of Quality Assessment in Higher Education and Its Impact on Business Students' Development and Interest in University Education', *Journal on Efficiency and Responsibility in Education and Science*, vol. 15, no. 2, pp. 63-71. https://doi.org/10.7160/eriesj.2022.150201

Dirin, A., Niemunen, M. and Alamäki, A. (2021) 'Social Media and Social Bonding in Students' Decision-Making Regarding Their Study Path', *International journal of information and communication technology education,* Vol. 17, No. 1, pp 88-104. https://doi.org/10.4018/IJICTE.2021010106

Drewes, T. and Michael, Ch. (2006) 'How do students choose a university?: an analysis of applications to universities in Ontario, Canada', *Research in Higher Education*, Vol. 47, No. 7, pp. 781-800. https://doi.org/10.1007/s11162-006-9015-6

Dülek, B. and Saydan, R. (2019) 'The Impact Of Social Media Advertisement Awareness On Brand Awareness, Brand Image, Brand Attitude And Brand Loyalty: A Research On University Students', *International Journal of Contemporary Economics,* Vol. 9, No. 2, pp. 470-494. https://doi.org/10.5281/zenodo.3596116

Eldegwy, A., Elsharnouby, T. H., and Kortam, W. (2018) 'How sociable is your university brand? An empirical investigation of university social augmenters' brand equity', *International Journal of Educational Management*, Vol. 32, No. 5, pp. 912-930. https://doi.org/10.1108/IJEM-12-2017-0346

Erjansola, A. M., Lipponen, J., Vehkalahti, K., Aula, H. M., and Pirttilä-Backman, A. M. (2021) 'From the brand logo to brand associations and the corporate identity: Visual and identity-based logo associations in a university merger', *Journal of Brand Management*, Vol. 28, No. 3, pp. 241-253. https://doi.org/10.1057/s41262-020-00223-5

Fajčíková A., 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. http://dx.doi.org/10.7160/eriesj.2019.120201

Gamoga, K. and Ambang, T. (2020) 'Factors influencing decision making on the choice of higher education institutions by prospective students: The experiences of students at Divine Word University in Papua New Guinea', *Contemporary PNG Studies*, Vol. 33, pp. 16-29.

García y García B. E. (2021) 'To What Factors do University Students Attribute Their Academic Success?', *Journal on Efficiency and Responsibility in Education and Science*, Vol.14, No. 1, pp. 1-8. http://dx.doi.org/10.7160/eriesj.2021.140101

Ghosh, A. K., Whipple, T. W. and Bryan, G. A. (2001) 'Student trust and its antecedents in higher education', *The Journal of Higher Education*, Vol. 72, No. 3, pp. 322-340. https://doi.org/10.2307/2649334

Gross, H. P., Ingerfurth, S., Willems, J. (2021) 'Employees as reputation advocates: Dimensions of employee job satisfaction explaining employees' recommendation intention', *Journal of Business Research*, Vol. 134, pp. 405-413. https://doi.org/10.1016/j.jbusres.2021.05.021

Haikel-Elsabeh, M., Zhao, Z., Ivens, B. and Brem, A. (2019) 'When is brand content shared on Facebook? a field study on online Word-of-Mouth', *International Journal of Market Research*, Vol. 61, No. 3, pp. 287-301. https://doi.org/10.1177/1470785318805301

He, J., Liu, H. and Xiong, H. (2016) 'SocoTraveler: Travel-package recommendations leveraging social influence of different relationship types', *Information and Management*, Vol. 53, No. 8, pp. 934-950. https://doi.org/10.1016/j.im.2016.04.003

Hejlová, D. (2015) *Public relations*, Praha: Grada Publishing.

Hoppe, D. (2018) 'Linking employer branding and internal branding: establishing perceived employer brand image as an antecedent of favourable employee brand attitudes and behaviours', *Journal of product*, Vol. 27, No. 4, pp. 452-467. https://doi.org/10.1108/JPBM-12-2016-1374

Huang, L.,Wang, M., Chen, Z., Deng, B. and Huang, W. (2020) 'Brand image and customer loyalty: Transmitting roles of cognitive and affective brand trust', *Social Behavior*, Vol. 48, No. 5, pp. 1-12. https://doi.org/10.2224/sbp.9069

Jalkala, A. and Terho, H. (2014) 'Customer Reference Marketing: Conceptualization and Link to Performance', *AMA Winter Educators' Conference Proceedings*, Nashville.

Joseph, M., Mullen, E. W. and Spake, D. (2012) 'University branding: Understanding student's choice of an educational institution', *Journal of Brand Management*, Vol. 20, No. 1, pp. 1-12. https://doi.org/10.1057/bm.2012.13

Judson, K. M., Gorchels, L. and Aurand, T. W. (2006) 'Building a university brand from within: A comparison of coaches' perspectives of internal branding', *Journal of Marketing for Higher Education*, Vol. 16. No. 1, pp. 97-114. https://doi.org/10.1300/j050v16n01_05

Khuong, M. and Linh, U. (2020) 'Influence of work-related stress on employee motivation, job satisfaction and employee loyalty in hospitality industry', *Management Science Letters*, Vol. 10, No. 14, pp. 3279-3290. https://doi.org/10.5267/j.msl.2020.6.010

Kinzie, J., Palmer, M., Hayek, J., Hossler, D., Jacob, S. A. and Cummings, H. (2004) 'Fifty years of college choice: Social, political and institutional influences on the decision-making process', *Lumina Foundation for Education*.Vol. 5, No. 3.

Kotler, P. (2007) *Moderní marketing, 4th edition,* Praha: Grada.

Leonnard (2019) 'Exploring the Relationship among E-service Quality, E-trust, E-satisfaction and Loyalty at Higher Education Institutions', *Journal on Efficiency and Responsibility in Education and Science*, Vol. 12, No. 4, pp. 103-110. https://doi.org/10.7160/eriesj.2019.120401

Mehta, S., Singh, T., Bhakar, S. and Sinha, B., (2010) 'Employee loyalty towards organization—a study of academician', *International Journal of Business Management and Economic Research*, Vol. 1. No. 1, pp. 98-108.

Misran, N., Aziz, N.A., Arsad, N., Hussain, H., Zaki, W.M.D.W. and Sahuri, S.N.S. (2012) 'Influencing factors for matriculation students in selecting university and program of study', *Procedia-Social and Behavioral Sciences*, Vol. 60, pp. 567-574.

Murray, Keith B. (1991) 'A test of services marketing theory: consumer information acquisition activities', *Journal of marketing*, Vol. 55, No. 1, pp. 10-25. https://doi.org/10.1177/002224299105500102

Nguyen, B., Yu, X., Melewar, T. and Hemsley-Brown, J. (2016) 'Brand ambidexterity and commitment in higher education: An exploratory study', *Journal of Business Research*, Vol. 69, No. 8, pp. 3105-3112. https://doi.org/10.1016/j.jbusres.2016.01.026

Novotny, J. and Duspiva, P. (2014) 'Factors Influencing Consumers' Buying Behavior and Their Importance for Enterprises. With English summary', *E +M Ekonomie a Management,* Vol. 17, No. 1, pp. 152-166.

Park, C. W., Eisingerich, A. B., Pol, G. and Park, J. W. (2013) 'The role of brand logos in firm performance', *Journal of Business Research*, Vol. 66, No. 2, pp. 180-187. https://doi.org/10.1016/j.jbusres.2012.07.011

Phua, J., Jin, S. V. and Kim, J. (2017) 'Uses and gratifications of social networking sites for bridging and bonding social capital: A comparison of Facebook, Twitter, Instagram, and Snapchat', *Computers in Human Behavior*, Vol. 72, pp. 115-122. https://doi.org/10.1016/j.chb.2017.02.041

Sadeghvaziri, F., Gomar, O., Azimi, M., Shoja, A. and Mozafari, A. (2022) 'Brand Logo Benefit: Is There any Evidence of its Impact in the Higher Education Sector?', *Iranian Journal of Management Studies,* Vol. 15, No. 2, pp. 305-320. https://doi.org/10.22059/ijms.2021.308166.674175

Safari I., Davaribina M., Khoshnevis I. (2020) 'The Influence of EFL Teachers' Self-Efficacy, Job Satisfaction and Reflective Thinking on their Professional Development: A Structural Equation Modeling', *Journal on Efficiency and Responsibility in Education and Science*, Vol. 13, No. 1, pp. 27-40. https://doi.org/10.7160/eriesj.2020.130103

Sarkar, A., & Sarkar, J. G. (2016) 'Devoted to you my love: Brand devotion amongst young consumers in emerging Indian market', *Asia Pacific Journal of Marketing and Logistics*, Vol. 28, No. 2, pp. 180-197. https://doi.org/10.1108/APJML-06-2015-0095

Sernovitz, A. (2009) *Word of mouth marketing,* New York: Kaplan Publishing.

Shen, H. and Sengupta, J. (2018) 'Word of Mouth versus Word of Mouse: Speaking about a Brand Connects You to It More Than Writing Does', *Journal of Consumer Research*, Vol. 45, No. 3, pp. 595-614. https://doi.org/10.1093/jcr/ucy011

Shields, A. B. and Peruta, A. (2019) 'Social media and the university decision. Do prospective students really care?', *Journal of marketing for higher education,* Vol. 29, No. 1, pp. 67-83. https://doi.org/10.1080/08841241.2018.1557778

Schlesinger, W., Cerveta-Taulet, A. and Wymer, W. (2021) 'The influence of university brand image, satisfaction, and university identification on alumni WOM intentions', *Journal of Marketing for Higher Education*, Vol. 33, No. 1, pp. 1-19. https://doi.org/10.1080/08841241.2021.1874588

Siering, M., Deokar, A. V. and Janze, Ch. (2018) 'Disentangling consumer recommendations: Explaining and predicting airline recommendations based on online reviews', *Decision Support Systems*, Vol. 107, pp. 52-63. https://doi.org/10.1016/j.dss.2018.01.002

Sujchaphong, N., Nguyen, B., Melewar, T. C., Sujchaphong, P. and Chen, J. (2020) 'A framework of brand-centred training and development activities, transformational leadership and employee brand support in higher education', *Journal of Brand Management*, Vol. 27, No. 2, pp. 143-159. https://doi.org/10.1057/s41262-019-00171-9

THE (2021) *The Times Higher Education World University Rankings | World University Rankings 2021* [Online], Available: https://www.timeshighereducation.com/world-university-rankings/2021/world-ranking [17 July 2021].

USNWR (2021) *U.S. News & World Report* [Online], Available: https://www.usnews.com/ [17 July 2021].

Wood, L. (2004) 'Dimensions of brand purchasing behaviour: Consumers in the 18-24 age group', *Journal of Consumer Behaviour: An International Research Review*, Vol. 4, No. 1, pp. 9-24. https://doi.org/10.1002/cb.154

Wu, X., Chen, H. X. and Li, X. (2019) 'The Moderation Effect of Social Factors on Marketing Factors in Consumer Research', *Transnational Corporations Review,* Vol. 11, No. 2, pp. 157-165. http://dx.doi.org/10.1080/19186444.2019.1603778

APPENDIX

Assumption	Tests	Satisfaction	Recommendation1	Overall impression	Recommendation2
Homoskedasticity	Breusch-Pagan test (*p*-value)	0.19	0.04	0.02	0.05
Normality	Normality of residues (*p*-value)	0.03	0.06	0.09	0.13
Specification	LM test (*p*-value)	0.14	0.02	0.02	0.36
	Reset test (*p*-value)	0.22	0.04	0.05	0.24
Multicollinearity	VIF <10	fulfilled	fulfilled	fulfilled	fulfilled

Table 8: The classical assumption of models (source: Questionnaire 2020, *n* = 271)

GRADE POINT AVERAGE: THE RELATIONSHIP WITH RESULTS OF ENTRANCE ASSESSMENT, LEARNING MOTIVATION, ACHIEVEMENT MOTIVATION, AND PERCEPTION OF TEACHER LEADERSHIP

Nataliia Sereda[1]✉
Svitlana Reznik[1]
Tetiana Solodovnyk[1]
Zhanna Bogdan[2]
Oleksandr Romanovskyi[1]

[1]National Technical University «Kharkiv Polytechnic Institute», Ukraine

[2]Simon Kuznets Kharkiv National University of Economics, Ukraine

✉ nataliia.sereda@khpi.edu.ua

ABSTRACT

The purpose of this study is to determine the relationship between the GPA of graduates of social majors at National Technical University «Kharkiv Polytechnic Institute» (Ukraine) with the results of the entrance assessment, learning motivation, achievement motivation, and perception of teacher leadership, based on the correlation and regression analysis of the study with a total of 502 respondents.

Results of the presented study demonstrate a significant correlation between the level of academic success and the level of educational and cognitive motivation of graduates (r = 0.644, p < 0.010). A less strong connection was found between graduates' GPA and the entrance examination results (r = 0.502, p < 0.010). The weakest, albeit statistically significant, relationship is between GPA variables and students' perception of teacher leadership (r = 0.160, p < 0.010), as well as between GPA and motives for creative self-realization (r = 0.139, p < 0.010). The article also carried out a correlation analysis for groups of students by level of education, form of study, majors, and gender of respondents. Obtained results are discussed with the purpose of improving the procedure for selecting applicants for admission to universities and improving the educational process.

KEYWORDS

Achievement motivation, entrance assessment, Grade Point Average, learning motivation, teacher leadership

HOW TO CITE

Sereda N., Reznik S., Solodovnyk T, Bogdan Z., Romanovskyi O. (2024) 'Grade Point Average: The Relationship with Results of Entrance Assessment, Learning Motivation, Achievement Motivation, and Perception of Teacher Leadership', *Journal on Efficiency and Responsibility in Education and Science*, vol. 17, no. 1, pp. 23-34. http://dx.doi.org/10.7160/eriesj.2024.170103

Article history
Received
August 23, 2022
Received in revised form
June 22, 2023
Accepted
September 18, 2023
Available on-line
March 31, 2024

Highlights

- *Compared to other variables researched, students' educational and cognitive motives have the strongest connection with GPA.*
- *The focus on reforming the system of admission in Ukrainian HEI should be based on a statistical assessment of its effectiveness.*
- *The identified relationships between GPA and student motivation can be utilized to improve students' learning outcomes.*
- *The students' perception of their teachers, particularly their leadership skills, is also important for the educational process results.*

INTRODUCTION

The results of teaching and learning in universities are characterized by a number of indicators, including the academic performance of students. During the study process at Higher Education Institutions (HEI), the most complete academic performance is represented by the average score of the graduate's diploma (GPA), as the entire period of a student's education and the success of their studies in all disciplines of the curriculum are taken into account. GPA can be related to many different factors. In our research, we focused attention, firstly, on the results of the entrance assessment (competition score) as a reflection of the level of initial training of students,

secondly, on the motivation of students as a reflection of their interests, desires, and aspirations, and, thirdly, on their perception of teacher leadership.

Students' perception of teachers is important for learning outcomes (Nabaho et al., 2017). Havik and Westergård (2020) determined the relationship between students' perception of classroom interactions, teachers' emotional support, and the involvement of students in the learning process. Marksteiner et al. (2021) determined that the peculiarities of students' perception of their teachers may be associated with the frequency of their unethical behavior (writing off homework and exam tests). Students' perception of teachers is a component of the educational environment that affects students' learning and academic achievements (Shah et al., 2019) and may be related to student motivation (Noels et al., 1999).

Our study includes students' perception of teacher leadership due to the fact that such leadership aims to improve students' learning and success level. However, the problem of the relationship between teacher leadership and students' achievements has not been studied enough by empirical methods of research (Wenner and Campbell, 2017).

Student's learning outcomes can also be influenced by the particular characteristics of admission to HEI. In Ukraine, after graduation from high school, applicants for HEI are based on the calculation of the competitive score. In 2008, a radical reform took place in the system of admission for the bachelor's level of education, which introduced External Independent Testing (EIT) instead of entrance exams, which were organized and conducted by each HEI independently. Bekeshkina et al. (2015: 5) note that the EIT system «made it possible to improve the efficiency and fairness of the HEI admission system». Since 2010, the competitive score for admission has been calculated based not only on the EIT results but also on the average score of the secondary education document (average score of the school leaving certificate).

At the same time, reforms in the admission system to HEI at both the bachelor's and master's levels of education in Ukraine still go on, and certain changes occur almost every year. These changes need to be evaluated in terms of their effectiveness, feasibility, and impact on students' further education at HEI. However, a generalized statistical assessment of the reform at the all-Ukrainian level was carried out based on the results of admission to the bachelor's programs in 2008-2011 and also in 2015 (Bekeshkina et al., 2015). For the years 2008-2011, the predictive validity of the criteria for the competitive selection of students was checked, i.e., to what extent the criteria for the competitive selection of students (the EIT average score and the average score of the school leaving certificate) allow predicting the academic success of students in the first year of HEI. In 2015, only a statistical assessment of the correlation between the average score of the EIT and the average score of the school leaving certificate was conducted, and the predictive validity of the entrance assessment results was not evaluated (Bekeshkina et al., 2015).

None of such all-Ukrainian statistical assessments were conducted for the master's programmes. The State Institution «Ukrainian Center for Educational Quality Assessment» (UCEQA) carries out EIT for admission to the bachelor's degree, as well as the foreign language Unified Entrance Exam (UEE «Foreign (language)») and the Unified Professional Entrance Test (UPET) for admission to the master's program of some majors. One of the activities of UCEQA is also monitoring studies of the quality of education. However, they are carried out for primary education only (Ukrainian Center for Educational Quality Assessment, 2007). It can be noted that the reforms of the Ukrainian system of admission to HEI continue. Still, there is a lack of statistical assessment of the relationship between the results of the entrance assessment and the subsequent success of students in HEI.

It should also be noted that one of the most influential factors of successful teaching and learning is motivation (Filgona et al., 2020). Motivation is important for a student's academic progress (Hamdan et al., 2010). That is why, in our work, we decided to compare the strength of the connection between motivation and other variables and students' academic results.

Therefore, the purpose of this work is to determine the relationship between the GPA of graduates and the corresponding variables - the results of the entrance exam, students' motivation, and their perception of teacher leadership.

The structure of the article consists of an introduction with a justification of the objectives of the study, a theoretical framework with an analysis of the features of joining the Ukrainian HEI, an analysis of educational motivation and achievement motivation, the definition of teacher leadership and the justification of its importance for the educational process; the section «research methodology» provides data on the methods used and the research procedure; the obtained quantitative results are presented in the section «results of the study», and their interpretation and comparison with the results of other studies are given in the section «discussion», the article ends with conclusions drawn from the results of the study.

THEORETICAL FRAMEWORK

Learning outcomes can be influenced by various factors. The 3P learning model (Biggs, 2003) considers factors that relate to three points in time: presage, process, and product, respectively. Presage includes two major aspects: student characteristics and teaching context. Pascarella's learning model (1985) also pays attention to the structural/organizational characteristics of institutions. Trigwell and Prosser (1997) added to the Biggs model (2003) such an important factor as students' perception of the learning context. In our study, we considered the factors that characterize students (prior knowledge, motivation), as well as their perception of their teachers.

Learning motivation

Learning motivation has a significant impact on the effectiveness of student learning (Gao, 2019). Filgona et al. (2020) declare learning motivation the central element of good teaching and the single most important element of learning. Hu et al. (2016) indicate academic motivation as a key determinant of academic performance.

The results of a number of studies confirm the relationship between learning motivation and students' academic achievement (Chang and Tsai, 2022; Duchatelet and Donche, 2019; Farnam and Anjomshoaa, 2020; Titrek et al., 2018).

Teachers pay considerable attention to student motivation to ensure effective teaching and learning (Hamdan et al., 2010; Law and Chuah, 2009; Sedden and Clark, 2016; Tabatabaei et al., 2017).

According to the model by Rean and Yakunin in the modification of Badmaeva (Kominko and Kucher, 2005), educational motivation may include external (communicative, prestige, social, avoidance of failure) and internal (educational and cognitive, professional and creative self-realization) motives of the student. Since research results indicate that internal (intrinsic) motivation is more effective for the results of students' academic activities (Ghanizadeh et al., 2017; Taylor et al., 2014), the relationship between these motives and GPA is studied in our work.

Achievement motivation

Achievement motivation means striving for high results in various activities. According to the theory of Atkinson (1957, 1964) individual differences in achievements are determined by the predominance of the level of motivation to achieve success over the motivation to avoid failure. People with high achievement motivation tend to be more persistent in achieving success (Cooper, 1983).

McEwan and Goldenberg (1999) conclude that Atkinson's theory of achievement motivation is confirmed based on the fact that the participants in their experiment had high achievement motivation, and academic success was determined by their first-semester grade point average (GPA). The results of a significant positive relationship between achievement motivation and students' academic results have been obtained in a number of studies (Busato et al., 2000; Ergin and Karataş, 2018; Mahdavi et al., 2021; Richardson and Abraham, 2009).

The results of a meta-analysis by Robbins et al. (2004) found that the best predictors for the GPA of college students were academic self-efficacy and achievement motivation.

Thus, one of the tasks of our work was to compare the strength of the relationship between students' achievement motivation and GPA.

Teacher leadership

Teacher leadership is determined by the high quality of pedagogical work in the classrooms, but at the same time, «going beyond the boundaries of the classroom», which involves active cooperation with other teachers, participation in professional communities, conducting professional master classes, developing programs, etc. (Wenner and Campbell, 2017). Sometimes, teacher leadership involves a formal position, but more often, it is an informal role (Wenner and Campbell, 2017; Katzenmeyer and Moller, 2011). Teacher-leaders influence students, other teachers, the professional-pedagogical community, the educational institution, but the ultimate goal of their leadership is the improvement of the educational process and the students' success (Wenner and Campbell, 2017).

Sometimes, the leadership of scientific and pedagogical workers of universities is considered exclusively in the context of their scientific activity. Liu et al. (2016) developed a mechanism for calculating the bibliometric indicator of Academic Leadership, which makes it possible to evaluate the effectiveness of the teacher's research activities as the ratio of the results obtained and the resources used. At the same time, a number of researchers do not consider it right that in the evaluation of the leadership of teachers at universities, more attention is paid to the research activities of teachers than to their success in teaching (Hofmeyer et al., 2015; Nunn and Pillay, 2014).

Draper et al. (2015) especially emphasize the interconnectedness of leadership and innovation in teaching. Leadership is important for innovation in education through the following: leading by example, creation of professional teams and communities for interaction aimed at improving the education process, persuasion, and influence. Wenner and Campbell (2017) in their review of theoretical and empirical research on teacher's leadership in K-12 schools conducted between 2004 and 2013, they report that the teacher's leadership is the second most important factor affecting the students' learning.

At the same time, scientific and teaching activities can be interconnected. Swihart et al. (2016) conducted a study at 33 research universities in the United States, showing that teachers with a bigger number of scientific results (number of publications and citation indices) were more effective. Still, this dependence was the lowest for teachers of social and management disciplines.

Taking into account the importance of teacher leadership in improving the educational process, in our work, we studied the relationship between students' perception of their teachers as leaders and their GPA.

Peculiarities of admission to higher education institutions in Ukraine

After receiving a general secondary education, a person has the opportunity to enter Ukrainian Higher Education Institutions. The terms of admission to higher education (both the first bachelor's and second master's level of education) are subject to approval by the Ministry of Education and Science of Ukraine every year. Every year, in these conditions, there are certain changes related to the search for optimal ways to admit applicants to study at HEI. Changes could relate to the number of applications that applicants could submit, the rules for providing places for study at the expense of the state budget, the rules for passing competitive selection, etc. (Ministry of Education and Science of Ukraine, n.d.). For example, the innovations of 2016 included the acceptance of documents for admission exclusively in electronic form through the electronic cabinets of the applicant, the ability to submit 15 applications for five different specialties under the state order for full-time education, and the introduction of a new principle for the distribution of places of study at the expense of the state budget for those students who scored the highest scores in the competition for a particular specialty in Ukraine (Ministry of Education and Science of Ukraine, n.d.).

There are different rules for admission of applicants to HEI in the world practice. High-quality selection in HEI should be based on the readiness of applicants for higher education as well as take into account the accuracy of predicting the learning success of students in HEI (Maruyama, 2012). In the USA, it is a common practice to use tests of educational competence - Scholastic Assessment Test (SAT) и American Test College Programme Assessment Test (ACT). The SAT contains a verbal and math subtest. The ACT contains four subtests: English, math, reading, and scientific reasoning. SAT and ACT are among the best predictors of academic performance in college (Coyle, 2006).

In Ukraine, when entering a bachelor's degree, applicants take subject tests that determine the results of learning in various subjects in the scope of the secondary school program. In 2010-2012, in Ukraine, an experiment was carried out on the use of a test of general educational competence, but it did not gain wide use. For admission to HEI, an applicant should choose the specialty in which he wants to study, and entrance tests depend on this specialty (only the Ukrainian language test is mandatory for everyone). The testing procedure involves passing External Independent Testing (EIT), which is conducted in specially organized testing centers.

For example, the students of the specialty «Psychology» participating in our study for admission to the bachelor's degree compiled EIT in three subjects: Ukrainian language and literature (as mandatory), biology (as profile), and a test on choice of the following subjects - History of Ukraine, Mathematics and/or Foreign Language. In 2021, there were changes in the rules and mandatory subjects for the specialty «Psychology» became Ukrainian Language and Literature and Mathematics, and it was suggested to choose one subject from six: History of Ukraine, Biology, Foreign Language, Physics, Chemistry, and Geography (Ministry of Education and Science of Ukraine, n.d.). For each test, a threshold score is determined, the only one for the whole of Ukraine, and defining the border «passed / failed the test». In the future, participation in the competition depends on the total of points scored and assumes that applicants with the highest scores enter the HEI.

Admission to master's degrees in Ukraine is also being reformed. The students who participated in our study for admission to the magistracy passed a test in Foreign Language (mandatory for all specialties) and a professional entrance test (depending on the specialty for which the applicant is applying). The third variable in determining the competitive score is the average score of the bachelor's degree.

It is common for undergraduate or graduate admissions to use a 600-point grading scale to determine the overall competitive score. If necessary to transfer one scale to another, generally recognized scales are used (for example, in cases where the applicant's scores on a document on previously received education are set on a different scale) (Ministry of Education and Science of Ukraine, n.d.; National Technical University «Kharkiv Polytechnic Institute», n.d.).

Master's Admission Reforms Concern Testing Procedure and Test Content. The focus on subject testing remains, but starting from 2017, gradually, for some specialties, not only the definition of the subject but also educational competencies become a component of professional testing. However, in 2021, this rule has not yet been extended to all master's degrees within Ukraine.

The change in the procedure for admission to the magistracy consists of the gradual, starting from 2016, the conduct of entrance examinations not in the educational institutions where applicants enter but in independent centers, similar to the EIT carrying out. The testing for admission to the master's degree in these centers was called the Unified Entrance Exam in a foreign language (UEE (FL)) and the Unified Professional Entrance Test (UPET).

Due to the insidious military aggression of Russia against Ukraine, the rules for entry in 2022 were simplified to preserve the intellectual potential of the nation and ensure the safety of test participants.

The post-war revival of Ukraine will also require a definition of what higher education should be, what the entrance should be for applicants to higher education, and what reformation paths will need to be chosen in the future in order to make education effective and meet the social demands of Ukrainian society.

MATERIALS AND METHODS

Participants

The study took into account the results of graduates of three years (2018 - June 2020) of the Department of Pedagogy and Psychology of Social System Management named after academician I.A. Ziaziun of the Faculty of Social and Humanitarian Technologies of National Technical University «Kharkiv Polytechnic Institute» (NTU «KhPI», Ukraine). The work was carried out within the framework of departmental studies of the problems of increasing the efficiency of the educational process of students of the department and studies of leadership in education. The total number of graduates of the mentioned Department in the specialties «Psychology», «Educational and Pedagogical Sciences» and «Public Management and Administration» was 562 people during this period. However, some of them did not answer the questions of the test and questionnaire in full. Also, a small number of students transferred to study at NTU «KhPI» from other educational institutions. Therefore, their competitive scores on the entrance assessment were not taken into account. So, statistical calculations were carried out for 502 respondents. 271 of them were full-time graduates, 231 were part-time graduates, 168 were bachelor's degree graduates, 334 were master's degree graduates, 124 were male, and 378 were female.

Instruments

The results of the entrance assessment of applicants, which are taken into account for participation in the competition and enrollment for training, are published every year in the public domain on the website (Unified State Electronic Database for Education, n.d.), which data were used by us in our work. As it has already been noted, the competitive score of applicants for both bachelor's and master's programs was calculated on a 600-point scale, taking into account the results of the relevant subject tests and the average score of the diploma of preliminary education, which allows for statistical calculations and comparisons of results.

The academic success of students in the Ukrainian HEI is assessed by teachers simultaneously on three scales: traditional for Ukraine (which existed even before joining the Bologna Process) four-point scale («excellent» - five points, «good» - four points, «satisfactory» - three points, «unsatisfactory» - two points), one hundred points (0-100) and the scale of the European Credit Transfer and Accumulation System (ECTS) (A, B, C, D, E, F, FX). We used a more accurate and flexible scale, namely a 100-point one in our study. The GPA was calculated according to

the practice accepted in Ukraine as the sum of the points received by the student for the entire period of study, divided by the number of grades (Bekeshkina et al., 2015).

The study also used the results of assessing the positive motivation of students' learning, which were determined by three scales of the methodology for diagnosing the learning motivation of students, developed by Rean and Yakunin in the modification of Badmaeva (Kominko and Kucher, 2005): educational and cognitive motives, professional motives and motives of creative self-realization. The methodology provides for students' answers to a set of questions on a 5-point Likert scale: 1 point corresponded to the minimum significance of the motive, 5 - to the maximum. The scale of educational and cognitive motives involves diagnosing students' interest in new knowledge and the desire to learn successfully; the scale of professional motives determines the propensity for a particular profession, education for implementation in future professional activities; the scale of creative self-realization is aimed at determining interest in creative activity. In total, there were 15 questions of the methodology on these scales, and the arithmetic mean of the students' answers was determined for each scale (Kominko and Kucher, 2005).

Diagnostics of students' achievement motivation was determined according to the Ehlers method. Students gave «yes» or «no» answers to 41 questions; the number of points according to the method was calculated using a corresponding key, according to which one point was awarded for answers «yes» or «no» to certain questions (Danylchenko and Vertel, 2012).

Students' perception of teacher leadership was determined by a questionnaire. Graduates were asked to answer the question, «During your studies at the University, how often did your lecturers show teacher leadership?» in accordance with a 5-point Likert scale: 1 - very rarely; 2 - rarely; 3 - mediocre; 4 - often; 5 - very often. It should be noted that leadership is one of the scientific directions of the graduation department of these students, and some of them participated in research in this field. This was manifested through participation in student conferences and seminars, writing theses of reports and articles, as well as in the topics of coursework and final qualification papers. Almost every student studied one or more leadership courses, including the «Pedagogical leadership» discipline. In addition, the students' conscious answers to the questions of the questionnaire were facilitated by the instruction and discussion with them of the essence of the concept of «teacher leadership».

Data analysis

Correlation-regression analysis was used to assess the relationship between variables. Correlation analysis allows you to assess the strength of the relationship between variables. This type of statistical analysis is used to assess the predictive validity of input assessment methods (Bekeshkina et al., 2015; Westrick et al., 2019). The study used Pearson's correlation coefficient for statistical analysis. It is widely used in science to measure the degree of linear dependence between two variables, which takes values between or equal to −1 to +1 (Stigler, 1989). Regression analysis provides an opportunity to obtain more information and determine the degree to which changes in one variable can lead to changes in another. However, the mutual correlation of variables among themselves (Table 1) in our study does not allow applying regression analysis in these cases. Therefore, the construction of the regression model was carried out only for several independent variables. The quality characteristic of the regression model is determined by the coefficient of determination (R squared), which shows what percentage of the variation of the dependent variable is explained by the variation of the independent variable. Another important indicator in calculations is the F-test, which is used to assess the significance of the coefficient of determination (Rouaud, 2013).

Null hypotheses (H_0) were formulated that no statistically significant relationship exists between the average score of the graduates' diploma (independent variable) and the corresponding dependent variables (results of entrance assessment, professional motives, motives of creative self-realization, educational and cognitive motives, achievement motivation, students' perception of teacher leadership). For example, H_{01}: there is no statistically significant relationship between the average score of a graduate's diploma and their entrance exam result. When the calculated values were below the significance level of 0.05, the null hypothesis was rejected. Descriptive statistics are presented in Table 1. The average GPA was 85.020 ± 8.200, the average result of the entrance assessment was 460.570 ± 67.020, the average values of professional motives, motives for creative self-realization, educational and cognitive motives were, respectively, 3.770 ± 0.88.000, 3.710 ± 0.970, 3.440 ± 0.810, achievement motives were 17.300 ± 4.070, perception of teaching leadership was 3.020 ± 1.120.

Variables	Min	Max	*M*	*SD*
Grade Point Average	64.520	98.760	85.020	8.200
Result of entrance assessment	189.000	600.00	460.570	67.020
Professional motives	0.800	5.800	3.770	0.880
Motives of creative self-realization	0.500	5.000	3.710	0.970
Educational and cognitive motives	1.000	5.000	3.440	0.810
Achievement motivation	3.000	28.000	17.300	4.070
Perception of teacher leadership	1.000	5.000	3.020	1.120

Table 1: Descriptive statistics for all study variables

RESULTS

The results of the correlation analysis for the entire sample of graduates (502 people) are presented in Table 2. As can be seen from Table 2, GPA and the competitive score of the entrance assessment are significantly correlated with each other ($r = 0.502$, $p < 0.010$); however, the relationship between

GPA and the level of educational and cognitive motivation of graduates is stronger ($r = 0.644$, $p < 0.010$). The weakest, albeit statistically significant, relationship is between variables representing GPA and students' perception of teacher leadership ($r = 0.160$, $p < 0.010$), as well as between GPA and motives for creative self-realization ($r = 0.139$, $p < 0.010$).

Variables	Grade Point Average	Result of entrance assessment	Professional motives	Motives of creative self-realization	Educational and cognitive motives	Achievement motivation	Perception of teacher leadership
Grade Point Average	1.000						
Result of entrance assessment	0.502**	1.000					
Professional motives	0.308**	0.070	1.000				
Motives of creative self-realization	0.139**	0.022	0.287**	1.000			
Educational and cognitive motives	0.644**	0.315**	0.502**	0.419**	1.000		
Achievement motivation	0.449**	0.267**	0.348**	0.255**	0.574**	1.000	
Perception of teacher leadership	0.160**	0.024	0.067	0.042	0.176**	0.182**	1.000

** - *correlation is significant at the 0.010 level (two-tailed).*

Table 2: Correlation between GPA, competitive entrance assessment score, motivation, and students' perception of teacher leadership, 2018-2020 (source: own calculation)

All variables on the scales of positive learning motivation and achievement motivation of graduates correlate with each other. Students' perception of teacher leadership is also related to students' educational and cognitive motives ($r = 0.176$, $p < 0.010$) and achievement motivation ($r = 0.182$, $p < 0.010$). To obtain more information regarding the relationship between certain variables, a regression model was built (Tables 3, 4, 5) for the GPA dependent variable and independent variables that do not correlate with each other, namely: competitive score of entrance assessment, motivation of professional activity, perception of teacher leadership. The choice of these independent variables is due to the fact that the model with other variants of independent variables without cross-correlation results in a lower coefficient of determination.

Model	*R*	R-squared	Adjusted R-squared	Standard error of estimate
	0.586[a]	0.344	0.340	6.661

[a] - *predictors: (const) competitive entrance assessment score, professional motivation, perception of teaching leadership.*

Table 3: The value of the coefficient of determination, 2018-2020 (source: own calculation)

As can be seen from the table, 34% of the variance of the GPA of graduates is determined by the variation of the set of variables «competitive score of entrance assessment», «motivation of professional activity», and «perception of teacher leadership».

Model*	Sum of squares	Degrees of freedom	Mean square	*F*	Significance
Regression	11583.290	3	3861.097	87.013	0.000[b]
Residual	22098.132	498	44.374		
Total	33681.422	501			

* - *analysis of variance, where a is the dependent variable: GPA.*

[b] - *predictors: (const) competitive entrance assessment score, professional motivation, perception of teaching leadership.*

Table 4: Results of model testing using Fisher's test, 2018-2020 (source: own calculation)

The F-test (Table 4) shows that the coefficient of determination is statistically significant. Regression analysis also shows the significance of all independent variables and that the most significant factor is the result of the entrance assessment.

Model	Unstandardized coefficients		Standardized coefficients	*t*	Significance
	B	Standardized error	Beta		
(Constant)	45.704	2.455		18.613	0.000
Result of entrance assessment	0.059	0.004	0.480	13.197	0.000
Professional motives	2.485	0.341	0.266	7.284	0.000
Perception of teacher leadership	0.958	0.266	0.131	3.603	0.000

Table 5: The value of the regression coefficient of the model, 2018-2020 (source: own calculation)

It should also be noted that the construction of a single-factor regression model, in which educational and cognitive motivation is an independent variable, allows us to obtain a determination coefficient of 0.414, which shows greater adequacy of the model. The results of correlation analysis for groups of students by level of education, form of study, specialties, and gender of respondents are given further.

As noted above, out of 502 respondents, 168 people completed a bachelor's degree, 334 a master's degree. Table 6 shows the Pearson correlation coefficients depending on the level of education.

Variables	Result of entrance assessment	Professional motives	Motives of creative self-realization	Educational and cognitive motives	Achievement motivation	Perception of teacher leadership
Bachelor's graduates (n = 168)						
Grade Point Average	0.453**	0.412**	0.148	0.588**	0.432**	0.191*
Master's graduates (n = 334)						
Grade Point Average	0.451**	0.241**	0.141*	0.663**	0.499**	0.156**

** - correlation is significant at the 0.010 level (two-tailed).

* - correlation is significant at the 0.050 level (two-tailed).

Table 6: Correlation between GPA, competitive entrance assessment score, students' motivation and perception of teacher leadership for the groups of bachelors (n = 168) and masters (n = 334), 2018-2020 (source: own calculation)

As can be seen from the tables, for bachelors, motives of creative self-realization do not have a significant relationship with GPA; for masters, such a relationship is significant at the 0.05 level. The relationship between GPA and graduate perceptions of teacher leadership is significant at the 0.05 level for undergraduates and 0.01 for masters.

Table 7 provides data for samples by form of education.

Variables	Result of entrance assessment	Professional motives	Motives of creative self-realization	Educational and cognitive motives	Achievement motivation	Perception of teacher leadership
Graduates of full-time education (n = 271)						
Grade Point Average	0.559**	0.332**	0.154*	0.652**	0.473**	0.212**
Graduates of part-time education (n = 231)						
Grade Point Average	0.313**	0.300**	0.136*	0.632**	0.440**	0.102

** - correlation is significant at the 0.010 level (two-tailed).

* - correlation is significant at the 0.050 level (two-tailed).

Table 7: Correlation between GPA, competitive entrance assessment score, motivation, and students' perception of teacher leadership for full-time and part-time graduates, 2018-2020 (source: own calculation)

For part-time graduates, not only GPA and educational and cognitive motivations (r = 0.632, p < 0.010) but also GPA and achievement motivation (r = 0.440, p < 0.010) have a stronger relationship compared with the indicators of the correlation between GPA and the results of the entrance assessment (r = 0.313, p < 0.010). The correlation coefficient between GPA and professional motives turned out to be close in strength (r = 0.300, p < 0.010). The correlation coefficient between GPA and graduates' perception of teacher leadership is not statistically significant, in contrast to full-time graduates (r = 0.212, p < 0.010). For both full-time and part-time graduates, the relationship between GPA and motives for creative self-realization is the weakest among other variables and is significant at the level of 0.050 (r = 0.154 and r = 0.136, p < 0,050, respectively).

Table 8 provides data for samples by specialty.

Variables	Result of entrance assessment	Professional motives	Motives of creative self-realization	Educational and cognitive motives	Achievement motivation	Perception of teacher leadership
Graduates of the specialty «Psychology» (n = 233)						
Grade Point Average	0.548**	0.133	0.050	0.714**	0.554**	0.144
Graduates of the specialty «Educational, Pedagogical Sciences» (n = 138)						
Grade Point Average	0.463**	0.349**	0.059	0.590**	0.457**	0.283**
Graduates of the specialty «Public Management and Administration» (n = 131)						
Grade Point Average	0.398**	0.171	0.195*	0.639**	0.457**	0.094

** - correlation is significant at the 0.010 level (two-tailed).

* - correlation is significant at the 0.050 level (two-tailed).

Table 8: Correlation Between GPA, Competitive Entrance Assessment Score, Motivation, and Student Perceptions of Teacher leadership for Social Science Graduates, 2018-2020 (source: own calculation)

For graduates of all three majors, the correlation between GPA and learning-cognitive motives is stronger compared to other variables. For graduates of the specialties «Psychology» and «Educational, Pedagogical Sciences,» the correlation coefficient between GPA and professional motives is significant at the level of 0.01 (r = 0.364, r = 0.349, respectively), for graduates of the specialty «Public Management and Administration» the correlation between these variables are insignificant. For graduates majoring in «Psychology» and «Public Management and Administration», the relationship between GPA and motives for creative self-realization is significant at the level of 0.050 (r = 0.140, r = 0.195 respectively), but for graduates majoring in «Educational, Pedagogical Sciences», the relationship between these variables is insignificant. For graduates majoring in «Psychology» and «Educational, Pedagogical Sciences», the correlation coefficient between GPA and perception of teacher leadership is significant at the level of 0.010 (r = 0.178, r = 0.283 respectively), for graduates majoring in «Public Management and Administration», the relationship between these variables is insignificant.

Table 9 represents data for samples by male and female gender.

Variables	Result of entrance assessment	Professional motives	Motives of creative self-realization	Educational and cognitive motives	Achievement motivation	Perception of teacher leadership
Male graduates (n = 123)						
Grade Point Average	0.548**	0.351**	0.155	0.687**	0.509**	0.167
Female graduates (n = 379)						
Grade Point Average	0.488**	0.286**	0.115*	0.622**	0.414**	0.165**

** - *correlation is significant at the 0.010 level (two-tailed).*

* - *correlation is significant at the 0.050 level (two-tailed).*

Table 9: Correlations between GPA, Competitive Entrance Assessment Score, Motivation, and Student Perceptions of Teacher leadership for Male and Female Graduates), 2018-2020 (source: own calculation)

For male graduates, the correlation between GPA and motives for creative self-realization, as well as between GPA and perception of teacher leadership, is insignificant. For female graduates, the relationship between these variables is significant: between GPA and motives for creative self-actualization at the level of 0.050 (r = 0.115), between GPA and perceptions of teacher leadership at the level of 0.010 (r = 0.165).

DISCUSSION

In all comparisons made in our study, a statistically significant relationship was found between the results of the entrance exam and the GPA, which, to a certain extent, corresponds to other similar studies (Bekeshkina et al., 2015; Čechová et al., 2019; Vulperhorst et al., 2018; Westrick et al., 2019). At the same time, the strength of the relationship varies across studies.

In the all-Ukrainian study of the predictive validity of the results of the entrance assessment for admission to the bachelor's degree, the following data were obtained: the correlation of the sum of the average EIT score and the average score of the document on secondary education with grades for the first year of study in HEI for 2010 was 0.549 (number of study participants was 27817), for 2011 it was 0.526 (number of study participants was 21227) (Bekeshkina et al., 2015). In our study, for undergraduate students, the correlation coefficient was 0.453; however, as it was noted, we considered students' grades not only for their first year only but also for their entire HEI period of study.

The correlation coefficient of the cumulative result of the introductory assessment, which included the average EIT score and the average score of the document of secondary education, and the first-year grades for male students was 0.470 (for 2011, the number of participants is 8832), for female students it was 0.504 (for 2011, number of participants 9533) (Bekeshkina et al., 2015). In contrast to these results, in our study, the correlation coefficient for male students was higher than for female students: it made 0.548 and 0.488, respectively. In the same study, the correlation data for different fields of knowledge were given: «Education» 0.538, «Social and Behavioral Sciences» 0.581, «Management and Administration» - 0.540 (Bekeshkina et al., 2015). In our work, the following results were obtained for specialties belonging to the relevant fields of knowledge: «Educational, Pedagogical Sciences» 0.463, «Psychology» 0.506, «Public Management and Administration» 0.398.

In contrast to the all-Ukrainian monitoring (Bekeshkina et al., 2015), research at the University College of the Netherlands under the Bachelor of Arts program (Vulperhorst et al., 2018) in order to find the most effective way to select high school students for participation in University programs studied the academic performance of students both in the short term (after completion of the first year of study - FYGPA (First Year GPA)) and in the long-term perspective. Selection to the college is based on HSGPA (High School GPA)). The predictive validity of HSGPA was higher for FYGPA than for final GPA, but in the context of our study, we considered the latter calculations. For two different high school programs, the correlation between HSGPA and final GPA was 0.580 (sample size was 315 individuals) and 0.550 (sample size was 113 individuals), respectively. For the entire population of study participants, the correlation was 0.620 (Vulperhorst et al., 2018). In our study, the correlation for bachelors is 0.453.

A number of studies have been aimed at determining the correlation between the results of various types of entrance exams and the subsequent academic performance of students during their training. Admission to European universities is mostly based on grades obtained at the previous stage of study (grade-based admissions).

Makransky et al. (2017) conducted a study at the University of Southern Denmark that demonstrated that admission based on a cognitive ability test followed by multiple mini-interviews (MMIs)) was more productive in terms of students' subsequent academic performance compared to students who were admitted based on grades. Thus, among students who were admitted to the University through the assessment of cognitive (test) and non-cognitive (MMI) skills, there was a significantly lower percentage of those who left training after the 1st year of study, the risk of failing the final exam after the 1st and 2nd years of study was lower. In addition, such students demonstrated a higher level of self-efficacy in academic learning and critical thinking.

Mengash's (2020) study, which lasted from 2016-2019 in Saudi Arabia with the participation of 2039 students, revealed that among the criteria for selecting applicants at the stage of preliminary admission, namely the High School Grade Average (HSGA), the Scholastic Achievement Admission Test (SAAT) and the General Aptitude Test (GAT); the SAAT is the most indicative. It seems to be the best way of predicting the Cumulative Grade Point Average (CGPA) of students while studying at the University the most accurately. This even led to changes in the College of Computer and Information Sciences (CCIS) admission system: it was decided to increase the weight of the SAAT criterion and change the weight of the three admission criteria (HSGA, SAAT, and GAT) from 60%, 20%, 20% to 30%, 40% and 30% respectively. As a result, the level of students with excellent or very good first-year CGPAs increased by 31% in 1 year of study.

Bestetti et al. (2020) considered the relationship between the university admissions test (UAT) and academic achievements at the end of the study, i.e., with the grades obtained in the Organized Structured Clinical Examination (OSCE), in the progress testing (PT), and in the final marks of the clerkship (FMC) in the conditions of the Department of Medicine, University of Ribeirão Preto, Ribeirão Preto City, Brazil. The UAT includes multiple-choice questions (MCQ) (the items include Biology, Portuguese Language, Mathematics, Geography, General and Brazilian History, Chemistry, and Physics) and an essay, which is aimed at revealing the applicant's ability to organize thoughts, present ideas in a logical sequence, and the ability to think analytically. Admission to the University is based on a formula that combines the UAT score with the scores obtained in the National High School Exam. The results of the study showed a moderate correlation between the UAT indicators and OSCE, PT, and FMC marks ($r = 0.460$; $p = 0.010$). No correlation was found between essay grades and OSCE, PT, and FMC marks.

As has already been mentioned, a feature of admission to Ukrainian HEI (for both bachelor's and master's programs) is subject-oriented assessment. In contrast to this admission criterion, in some countries of the world, tests of general educational competence, i.e., general readiness for successful study in HEI, are used. The American SAT is one of these tests. A study by Westrick et al. (2019) revealed that the correlation of SAT and HSGPA with first-year academic performance (FYGPA) is 0.61. An experimental pilot study was conducted on the introduction of a test of general educational competencies as an entrance test for HEI in Ukraine in 2010-2011. The test included two components: verbal and communicative (analytical reading, critical reading, essay writing) and logical and mathematical (Liashenko and Rakov, 2012). However, this practice was not implemented in the system of evaluating applicants for admission to HEI in Ukraine. A greater focus on tests of general academic competence, rather than on subject tests, when entering HEI in Ukraine can be the subject of debate.

The fact that the work (Bekeshkina et al., 2015) has determined a tendency to decrease the prognostic validity of all indicators of the competitive selection of HEI students during 2008-2011 deserves special attention. As we mentioned above, not all Ukrainian studies were conducted after 2011.

In general, we believe that the focus on reforming the Ukrainian HEI admission system should be based on a statistical assessment of its effectiveness. In order to improve the entrance assessment, it is necessary to take into account the connection between the points of the entrance assessment and the further academic performance of students at the University. In particular, the choice of whether admission should be based on the results of a subject test or a test of general academic aptitude; which subjects for admission to a particular major should be mandatory for testing; a list of which subjects can be offered to applicants to choose from all these questions require statistical evaluation. Our results demonstrate that the level of some motives has a higher correlation coefficient with GPA than the result of the entrance assessment, and these may be implemented as additional admission requirements - letter of motivation or admission interview for applicants. In such cases, the assessment criteria should be the applicant's strong interest in education, gaining new knowledge, acquiring a profession, as well as the need for high achievements, etc.

Our research work showed that, for all calculations, in comparison with the correlation between GPA and the result of the entrance assessment, the correlation between GPA and educational-cognitive motives of graduates had a higher coefficient. In addition, for Master's degree graduates, part-time graduates and graduates of the specialty «Public Management and Administration», a higher correlation coefficient was also noted between GPA variables and achievement motivation. The lowest GPA correlation coefficient from the results of the entrance evaluation was obtained from graduates in absentia and graduates in the specialty «Public Management and administration», the highest - among the graduates of full-time education, male graduates and graduates of the specialty «Psychology». As for the variables of study motivations, they had a statistically significant positive relationship with GPA. The obtained result corresponds to the results of similar studies (Chang and Tsai, 2022). The weakest relationship was observed between the motives of creative self-realization and GPA, and the strongest one was between educational and cognitive motives and GPA. It is possible to recommend that teachers use various means of motivation in the educational process, given that different motives have different connections with student learning outcomes.

At the same time, achievement motivation had a statistically

significant relationship with GPA for all groups of students, which teachers can use to improve teaching and learning outcomes.

Students' perception of their teachers is also important for the results of the educational process. We observe a positive statistically significant relationship between GPA and perceptions of teacher leadership for the entire collection of observations, as well as for bachelor's and master's. The significant relationship between GPA and the perception of teacher leadership among full-time students compared to part-time students can be explained by the greater number of interactions with teachers of the first group of students compared to those of the second group. The significance of the relationship also turned out to be dependent on the gender and major (specialty) of the students, which requires further research. In general, the results of our work correspond to the model of learning by Trigwell and Prosser (1997), who considered students' perception of the learning context as an important factor in learning outcomes. Recent research by García y García (2021) also proved that good teachers are one of the most important factors determining the academic success of university students. On the basis of this, teachers can be recommended to pay attention to how students perceive the presence or absence of teacher leadership.

The variable that had the strongest relationship with GPA among other variables, both for the entire collection of observations and for all separate groups of students, is students' educational and cognitive motives. These results to some extent correspond to the results of studies in which the relationship between the need for cognition and learning outcomes has been determined (Grass et al., 2017). The need for cognition is defined as an individual tendency to engage in and enjoy cognitive endeavors (Grass et al., 2017; Grass et al., 2018; Strobel et al., 2019). We agree with the thesis of Grass et al. (2017) that it is necessary to intensify the study of the need for knowledge in the context of higher education.

The potentially debatable question is whether it is possible to use the knowledge about the significant connection between the educational and cognitive motives of students and GPA when determining the admission procedure to HEI. At the same time, teachers should take into account the strength of the connection between these motives and the academic performance of students. Encouraging the desire to learn and maintaining students' cognitive interests can be an important strategy for improving learning outcomes and may help to reduce or eliminate negative factors of students' learning, such as gaps in basic (initial) knowledge, etc. Further research can be carried out in this direction.

The conduct of this study was limited to the students of three majors at one university. Further research with students of other majors may show different results regarding a greater or lesser relationship between GPA and entrance assessment results, which is confirmed by a 2015 Ukrainian study (Bekeshkina et al., 2015). In addition, other results may be obtained by comparing the strength of the relationship between GPA, motives, and perceptions of teacher leadership. Nevertheless, the obtained data is significant and can be used for further discussion regarding effective conditions for university admissions. We also consider it promising to continue research to compare the influence of the relationship between various factors on the academic success of students in order to determine ways to improve the educational process.

CONCLUSION

For the entire collection of observations, the results showed a statistically significant relationship between GPA and all variables - the results of the entrance assessment (competition score), learning motivation, achievement motivation, and students' perception of teacher leadership. Accordingly, the null hypotheses (H_0) were rejected. A comparison of groups of students was also carried out according to the level of education, form of study, specialties, and gender of the respondents.

The variable that had the strongest relationship with GPA among other variables, both for the entire collection of observations and for all separate groups of students, is students' educational and cognitive motives. The correlation between GPA and entrance test scores was weaker (less strong). For master's degree graduates, part-time graduates, and «Public Management and Administration» graduates, a higher correlation coefficient was also noted between the variables GPA and achievement motivation.

The results of the regression analysis showed that from the set of independent variables represented by «competitive score of entrance assessment,» «motivation of professional activity,» and «perception of teaching leadership,» the result of entrance assessment can be considered the most significant factor. However, the construction of a one-factor regression model, in which educational and cognitive motivation is an independent variable, demonstrates a greater adequacy of the model.

In order to improve the entrance assessment, it is necessary to take into account the connection between the points of the entrance assessment and the further academic performance of students at the University. In particular, the results of our study may be implemented as additional admission requirements - a letter of motivation or admission interview for applicants. In such cases, the assessment criteria should include the applicant's strong interest in education, gaining new knowledge, acquiring a profession, as well as the need for high achievements, etc. The identified relationships between GPA and student motivation, as well as between GPA and student perceptions of teacher leadership, are noteworthy for improving students' learning outcomes.

REFERENCES

Atkinson, J. W. (1957) 'Motivational determinants of risk-taking behavior', Psychological Review, Vol. 64, No. 6, pp. 359-372. https://doi.org/10.1037/h0043445

Atkinson, J. W. (1964) *An introduction to motivation*. Princeton: Van Nostrand.

Bekeshkina, I. E. ., Boliubash, Ya. Ya., Buhrov, V. A., Horokh, V. P., Hudzynskyi, V. Ye., Kovalchuk, Yu. O., Kovtunets, V. V., Kovtunets, O. V., Lisova, T. I., Mazorchuk, M. S., Rakov, S. A. and Sereda, L. I. (2015) *Study on the quality of the selection of students entering higher education institutions based on the results of external independent testing: analytical materials*. Kyiv: Nora-Druk.

Bestetti, R. B., Couto, L. B., Roncato-Paiva, P., Romão, G. S., Faria-Jr, M., Furlan-Daniel, R. A., Geleilete, T. J. M., Jorge-Neto, S. D., Mendonça, F. P., Garcia, M. E. and Durand, M. T. (2020) 'University Admission Test Associates with Academic Performance at the End of Medical Course in a PBL Medical Hybrid Curriculum', *Adv Med Educ Pract*, Vol. 11, pp. 579-585. https://doi.org/10.2147/AMEP.S255732

Biggs, J. B. (2003) *Teaching for quality learning at university: What the student does*, 2nd edition. London: Buckingham Open University Press/Society for Research into Higher Education.

Busato V. V., Prins F. J., Elshout J. J. and Hamaker C. (2000). 'Intellectual ability, learning style, personality, achievement motivation and academic success of psychology students in higher education', *Personality and Individual Differences,* Vol. 29, No. 6, pp. 1057-1068. https://doi.org/10.1016/s0191-8869(99)00253-6

Čechová, I., Neubauer, J. and Sedlačík, M. (2019) 'Tracking the University Student Success: Statistical Quality Assessment', *Journal on Efficiency and Responsibility in Education and Science*, Vol. 12, No. 1, pp. 12-25. http://dx.doi.org/10.7160/eriesj.2019.120102

Chang, Y-C. and Tsai, Y-T. (2022) 'The Effect of University Students' Emotional Intelligence, Learning Motivation and Self-Efficacy on Their Academic Achievement-Online English Courses', *Frontiers in Psychology*, Vol. 13, p. 818929. https://doi.org/10.3389/fpsyg.2022.818929

Cooper, W. H. (1983) 'An achievement motivation nomological network', Journal of Personality and Social Psychology, Vol. 44, No. 4, pp. 841-861. https://doi.org/10.1037/0022-3514.44.4.841

Coyle, T. (2006) 'Test-retest changes on scholastic aptitude tests are not related to g.', *Intelligence*, Vol. 34, No. 1, pp. 15-27. https://doi.org/10.1016/j.intell.2005.04.001

Danylchenko, N. O. and Vertel, A. V. (2012) *Psychodiagnostics: practicum.* [Online], Available: https://library.sspu.edu.ua/wp-content/uploads/2018/04/41-4.pdf [14 Jun 2022].

Draper, P., Cleaver, E., Cooper, H., Heath, T., Hilton, A. and Kember, M. (2015) *The Experiences of Innovators in Learning and Teaching*. London: Leadership Foundation for Higher Education.

Duchatelet, D. and Donche, V. (2019) 'Fostering self-efficacy and self-regulation in higher education: a matter of autonomy support or academic motivation?', *Higher Education Research and Development,* Vol. 38, No. 4, pp. 733-747. https://doi.org/10.1080/07294360.2019.1581143

Ergin, A. and Karataş, H. (2018) 'Achievement-oriented motivation levels of university students', *Hacettepe University Journal of Education,* Vol. 33, No. 4, pp. 868-887. https://doi.org/10.16986/HUJE.2018036646

Farnam, A. and Anjomshoaa, H. (2020) 'The Relationship between Metacognition Skills with Academic Motivation and Academic Achievement among High School Students in Kerman, Iran', *International Journal of Pediatrics*, Vol. 8, No. 1, pp. 10759-10766. https://dx.doi.org/10.22038/ijp.2020.45608.3731

Filgona, J., Sakiyo, J., Gwany, D.M. and Okoronka, A.U. (2020) 'Motivation in Learning', *Asian Journal of Education and Social Studies*, Vol. 10, pp. 16-37. https://doi.org/10.9734/AJESS/2020/v10i430273

Gao, F. (2019) 'Teaching Research on Learning Motivation of Database Course Based on Cloud Teaching Platform', *Advances in Social Science Education and Humanities Research,* Vol. 322, pp. 350-353. https://doi.org/10.2991/iserss-19.2019.92

García y García, B. E. (2021) 'To What Factors do University Students Attribute Their Academic Success?', *Journal on Efficiency and Responsibility in Education and Science*, Vol. 14, No. 1, pp. 1-8. http://dx.doi.org/10.7160/eriesj.2021.140101

Ghanizadeh, A., Jahedizadeh, S. and Blessinger, P. (2017) 'The Nexus Between Emotional, Metacognitive, and Motivational Facets of Academic Achievement among Iranian University Students', *Journal of Applied Research in Higher Education,* Vol. 9, No. 4, pp. 598-615. https://doi.org/10.1108/JARHE-05-2017-0060

Grass, J., John, N. and Strobel, A. (2018) 'Freude am Denken als Schlüssel zum Erfolg? Die Bedeutung von Need for Cognition für subjektives Erleben und Leistung im Studium', Zeitschrift für Pädagogische Psychologie, Vol. 32, No. 3, pp. 145-154. https://doi.org/10.1024/1010-0652/a000222

Grass, J., Strobel, A. and Strobel, A. (2017) 'Cognitive Investments in Academic Success: The Role of Need for Cognition at University', *Frontiers in Psychology*, Vol. 8, p. 790. https://doi.org/10.3389/fpsyg.2017.00790

Hamdan, K., Hamdan, A. and Ali, E. A. H. (2010). 'Motivating students to learn in academic environments', *Proceedings of the 3rd International Conference of Education, Research and Innovation (ICERI 2010),* Madrid, pp. 788-799.

Havik, T. and Westergård, E. (2020) 'Do Teachers Matter? Students' Perceptions of Classroom Interactions and Student Engagement, Scandinavian', *Journal of Educational Research*, Vol. 64, No. 4, pp. 488-507. https://doi.org/10.1080/00313831.2019.1577754

Hofmeyer A., Sheingold B. H., Klopper H. C. and Warland J. (2015) 'Leadership in Learning and Teaching in Higher Education: Perspectives of Academics in Non-Formal Leadership Roles', *Contemporary Issues in Education Research,* Vol. 8, No. 3, pp. 181-192.

Hu, W., Jia, X., Plucker, J. A. and Shan, X. (2016) 'Effects of a Critical Thinking Skills Program on the Learning Motivation of Primary School Students', *Roeper Review,* Vol. 38, No. 2, pp. 70-83. https://doi.org/10.1080/02783193.2016.1150374

Katzenmeyer, M. and Moller, G. (2011) 'Understanding Teacher Leadership', in E. Blair (ed.) *Teacher Leadership: The «New» Foundations of Teacher Education,* New York: Peter Lang AG.

Kominko, S. B. and Kucher, H. V. (2005) Best Psychodiagnostic Techniques, Ternopil: Kart-blansh.

Law, K. and Chuah, K. B. (2009) 'What Motivates Engineering Students? A Study in Taiwan', *International Journal of Engineering Education*, Vol. 25, No. 5, pp. 1068-1074.

Liashenko, O. I. and Rakov S. A. (2012) '*General educational competency test: basic principles and piloting results', Pedahohika i psykholohiia*, Vol. 2, pp. 27-35.

Liu, Y., Ou, F., Deng, Y., Wu, B., Liu, R., Hua, H., Guan, Y., Chen, R., Gjesteby, L., Yang, J., Vannier, M. W. and Wang, G. (2016) 'Bibliometric Index for Academic Leadership', *ArXiv*. https://doi.org/10.48550/arXiv.1610.03706

Mahdavi, P., Valibeygi, A., Moradi, M. and Sadeghi, S. (2021) 'Relationship Between Achievement Motivation, Mental Health and Academic Success in University Students', *International Quarterly of Community Health Education*, Vol. 43, No. 3, pp. 311-317. https://doi.org/10.1177/0272684X211025932

Makransky, G., Havmose, P., Vang, M. L., Andersen, T. E. and Nielsen, T. (2017) 'The predictive validity of using admissions testing and multiple mini-interviews in undergraduate university admissions', *Higher Education Research and Development*, Vol. 36, No. 5, pp. 1003-1016. https://doi.org/10.1080/07294360.2016.1263832

Marksteiner, T., Nishen, A. K. and Dickhäuser, O. (2021) 'Students' Perception of Teachers' Reference Norm Orientation and Cheating in the Classroom', *Frontiers in Psychology*, Vol. 12, p. 614199. https://doi.org/10.3389/fpsyg.2021.614199

Maruyama, G. (2012) 'Assessing College Readiness: Should We Be Satisfied With ACT or Other Threshold Scores?', *Educational Researcher*, Vol. 41, No. 7, pp. 252-261. https://doi.org/10.3102/0013189x12455095

McEwan, L. and Goldenberg, D. (1999) 'Achievement motivation, anxiety and academic success in first year Master of Nursing students', *Nurse Educ Today*, Vol. 19, No. 5, pp. 419-30. https://doi.org/10.1054/nedt.1999.0327

Mengash, H. A. (2020). 'Using Data Mining Techniques to Predict Student Performance to Support Decision Making in University Admission Systems', *IEEE Access*, Vol. 8, pp. 55462-55470. https://doi.org/10.1109/ACCESS.2020.2981905

Ministry of Education and Science of Ukraine (2022) *Higher education and adult education*, [Online], Available: https://mon.gov.ua/ua/tag/vishcha-osvita [14 Jun 2022].

Nabaho, L., Oonyu, J. and Aguti, J. N. (2017) 'Good teaching: Aligning student and administrator perceptions and expectations', *Higher Learning Research Communications,* Vol. 7, No. 1, pp. 27-42. https://doi.org/10.18870/hlrc.v7i1.321

National Technical University «Kharkiv Polytechnic Institute» (2022) Admissions guideline, [Online], Available: http://vstup.kpi.kharkov.ua/admission/admission_rules/ [14 Jun 2022].

Noels, K. A., Clément, R. and Pelletier, L. G. (1999) 'Perceptions of Teachers' Communicative Style and Students' Intrinsic and Extrinsic Motivation', *The Modern Language Journal,* Vol. 83, No. 1, pp. 23-34. https://doi.org/10.1111/0026-7902.00003

Nunn, R. and Pillay, A. (2014) 'After invention of the h-index, is there a place for the teaching track in academic promotion?', *Higher Education Research and Development,* Vol. 33, No. 4, pp. 848-850. https://doi.org/10.1080/07294360.2014.915465

Pascarella, E. T. (1985) 'Students' affective development within the college environment', Journal of Higher Education, Vol. 56, No. 6, pp. 640-663. https://doi.org/10.2307/1981072

Richardson, M. and Abraham, C. (2009) 'Conscientiousness and Achievement Motivation Predict Performance', *European Journal of Personality,* Vol. 23, No. 7, pp. 589-605. https://doi.org/10.1002/per.732

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R. and Carlstrom, A. (2004) 'Do Psychosocial and Study Skill Factors Predict College Outcomes? A Meta-Analysis', *Psychological Bulletin,* Vol. 130, No. 2, pp. 261-288. https://doi.org/10.1037/0033-2909.130.2.261

Rouaud, M. (2013). *Probability, Statistics, and Estimation.* London: Creative commons, [Online], Available: http://www.incertitudes.fr/book.pdf [14 Jun 2022].

Sedden, M. and Clark, K. (2016). 'Motivating Students in the 21st Century', *Radiologic technology*, Vol. 87, No. 6, pp. 609-616.

Shah, D. K., Piryani, S., Piryani, R. M., Islam, M. N., Jha, R. K. and Deo, G. P. (2019) 'Medical students' perceptions of their learning environment during clinical years at Chitwan Medical College in Nepal', *Advances in Medical Education and Practice,* Vol. 10, pp. 555-562. https://doi.org/10.2147/AMEP.S203377

Stigler, S. M. (1989) 'Francis Galton's Account of the Invention of Correlation', *Statistical Science,* Vol. 4, No. 2, pp. 73-79. https://doi.org/10.1214/ss/1177012580

Strobel, A., Behnke, A., Gärtner, A. and Strobel, A. (2019) 'The interplay of intelligence and need for cognition in predicting school grades: A retrospective study', *Personality and Individual Differences*, Vol. 144, pp. 147-152. https://doi.org/10.1016/j.paid.2019.02.041

Swihart, R., Sundaram, M., Höök, T. O. and DeWoody, J.A. (2016) 'Factors affecting scholarly performance by wildlife and fisheries faculty', *Journal of Wildlife Management,* Vol. 80, No. 3, pp. 563-572. https://doi.org/10.1002/jwmg.1034

Tabatabaei, S.-S., Ahadi, H., Bani-Jamali, S.-ol-S., Bahrami, H. and Khamesan, A. (2017) 'The Effects of Motivated Strategies for Learning Questionnaire (MSLQ) on Students' Cognitive and Meta-Cognitive Skills', *NeuroQuantology,* Vol. 15, No. 2, pp. 239-245. https://doi.org/10.14704/nq.2017.15.2.1068

Taylor, G., Jungert, T., Mageau, G., Schattke, K., Dedic, H., Rosenfield, S. and Koestner, R. (2014) 'A self-determination theory approach to predicting school achievement over time: The unique role of intrinsic motivation', *Contemporary Educational Psychology*, Vol. 39, No. 4, pp. 342-358. https://doi.org/10.1016/j.cedpsych.2014.08.002

Titrek, O., Çetin, C., Kaymak, E. and Ka¸sikçi, M. M. (2018) 'Academic motivation and academic self-efficacy of prospective teachers', *Journal of Education and Training Studies,* Vol. 6, No. 11a, pp. 77-82. https://doi.org/10.11114/jets.v6i11a.3803

Trigwell, K. and Prosser, M. (1997) 'Towards an Understanding of Individual Acts of Teaching and Learning', *Higher Education Research and Development*, Vol. 16, No. 2, pp. 241-252. https://doi.org/10.1080/0729436970160210

Ukrainian Center for Educational Quality Assessment (2007) About the Ukrainian Center for Educational Quality Assessment, [Online], Available: https://testportal.gov.ua/pro-utsoyao/ [14 Jun 2022].

Unified State Electronic Database for Education (2023) [Online], Available: https://vstup.edbo.gov.ua/ [14 Jun 2022].

Vulperhorst, J., Lutz, C., de Kleijn, R. and van Tartwijk, J. (2018) 'Disentangling the predictive validity of high school grades for academic success in university', *Assessment and Evaluation in Higher Education*, Vol. 43, No. 3, pp. 399-414. https://doi.org/10.1080/02602938.2017.1353586

Wenner, J. A. and Campbell, T. (2017) 'The Theoretical and Empirical Basis of Teacher Leadership: A Review of the Literature', *Review of Educational Research,* Vol. 87, No. 1, pp. 134-171. https://doi.org/10.3102/0034654316653478

Westrick, P., Marini, J., Young, L., Ng, H., Shmueli, D. and Shaw, E. (2019) *Validity of the SAT® for Predicting First-Year Grades and Retention to the Second Year*, [Online], Available: https://collegereadiness.collegeboard.org/pdf/national-sat-validity-study.pdf [14 Jun 2022].

EFFICIENCY ASSESSMENT ON CODIFIED KNOWLEDGE PRODUCTS: AN SFA APPROACH

Gustavo Ferro[1]✉
Nicolás Gatti[2]

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

[2]Universidad del CEMA (UCEMA) and Centro de Economía y Prospectiva CIEP-INTA, Argentina

✉ gaf97@ucema.edu.ar

Article history
Received
December23, 2022
Received in revised form
July 31, 2023
Accepted
September 19, 2023
Available on-line
March 31, 2024

ABSTRACT

Knowledge applied to innovation is increasingly recognized as an explanatory factor of economic growth. Innovation derives from applying knowledge to generate new products or processes. National Innovation Systems (NIS) performs as the formal or informal network of people within institutions interacting to produce and apply knowledge to innovation. NIS can be understood as two subsystems: one based on scientific and technological work, producing codified products (publications and patents), and the other centered on practical actions to diffuse, apply, and use knowledge. Our objective is to assess cost efficiency in the production of codified knowledge outputs (CKO), being our unit of analysis NIS (countries). To attain our goal, we apply a Stochastic Frontier Analysis (SFA) to estimate a cost frontier of CKO. The panel sample includes 1189 observations for 23 years (1996-2019) and 82 countries. Our main results identify determinants and patterns of efficiency and productivity, tendencies, and specifics of countries and groups of them.

KEYWORDS

Efficiency frontiers, cost frontier, codified scientific knowledge, National Innovation Systems

HOW TO CITE

Ferro G., Gatti N. (2024) 'Efficiency Assessment on Codified Knowledge Products: An SFA Approach', *Journal on Efficiency and Responsibility in Education and Science*, vol. 17, no. 1, pp. 35-54. http://dx.doi.org/10.7160/eriesj.2024.170104

Highlights

- *Given human and non-human resources, some National Innovation Systems perform better than others in producing codified knowledge outputs.*
- *Efficiency assessment concentrates on the best administration of resource scarcity and is useful for ex-ante planning and ex-post evaluation.*
- *Our empirical assessment identifies the best performers within 82 countries in producing scientific publications and patents.*

INTRODUCTION

Knowledge production is a key explanatory factor of economic growth. Early economic growth models treated technical change as exogenous, while more recent ones incorporated its endogenous role (Barro and Sala-i-Martin, 2003). The latter is the recognition that an important part of generated knowledge is not fortunate random discoveries; instead, it is derived from the deliberate effort in human and non-human investments, which depends on cost-benefit analysis, the resource allocation of its production process, and the efficiency in the use of those resources. Understanding the drivers of knowledge can provide useful policy implications for economic development.

Evaluating the outcomes of knowledge production is challenging. You can look at output indicators or an inventory of inputs to produce knowledge, or you can focus on the relationship between outputs and inputs and calculate partial productivity indexes. However, both analyses are insufficient. Partial productivity measures often omit the effect of interactions between relevant inputs. For instance, a ratio between production and labor units would leave out the complementarities between labor and capital in the production process. When it comes to productivity, it is relevant to consider the input and output vector altogether because it will be a more complete representation of a production or cost function for efficiency estimation.

The National Innovation System (NIS) is the formal or informal network of people within institutions, interacting to apply knowledge to innovation (that is, to generate new or improved products or processes). NIS can be divided into two subsystems: one based on scientific and technological work, producing codified knowledge outputs (CKO) (e.g., scientific publications and patents of inventions), and the other centered on practical and non-codified actions to diffuse, apply, and use knowledge. CKO can be measured directly because they are countable, and thanks to the effort of scientists working on bibliometrics and of international organizations compiling statistics of costs, inputs, and outputs. Instead, because

non-codified knowledge is embodied in people's minds or embedded in organizations, its measurement is quite elusive. Our objective is to assess cost efficiency in the production of CKO by country. CKO efficiency is related to the optimum usage of its output/input ratio, while CKO productivity considers the transformation of inputs into outputs (Nasierowski and Arcelus, 2003). Efficiency estimates can be made on pure technical conditions (output-to-input relationships) or in terms of allocative conditions (cost-to-output relationships). In the measurement of the efficiency of CKO activities, observation units (whether countries, regions, research institutes, or firms) are regarded as entities operating a production process where inputs - mainly capital and manpower - are transformed to produce CKO (Carrillo, 2019). Efficiency assessment is a tool to evaluate the administration of resource scarcity, and it is useful for ex-ante planning and ex-post evaluation. Knowing which efficient systems ex-ante could guide future investments, while ex-post evaluations teach about adjustments and possible improvements.

To estimate the technical efficiency of NIS, we apply a Stochastic Frontier Analysis (SFA) to estimate a cost frontier of CKO, considering relevant inputs and "environmental" (in the sense of contextual) conditions to address country-specific conditions. Our database is built on different sources for outputs (Scimago and WIPO), costs, inputs, and input prices (UNESCO), and macroeconomic and institutional issues to characterize the environment of CKO production (The World Bank and Heritage Foundation). The sample is a panel that includes 1189 observations for 23 years (1996-2019) and 82 countries, each one being a NIS.

Our work makes three main contributions. First, we developed an extensive literature review documenting the development of the concept of NIS. Second, we put together and performed an efficiency analysis on a database encompassing developed and developing countries, contrasting with most of the literature focusing on comparisons within OECD countries. Lastly, the method of analysis employed is econometric and concentrates on technical and allocative efficiency, which contrasts with the use of mathematical programming approaches for technical efficiency estimation only.

The article is organized as follows. Section 2 reviews the literature on NIS. Section 3 introduces the material and methods used to estimate the allocative efficiency of codified knowledge. Section 4 shows the main results, while Section 5 discusses the implications of our findings. Lastly, Section 6 summarizes the main conclusions.

LITERATURE REVIEW

On the NIS concept

A National Innovation System (NIS) can be defined as a network of private and public sector institutions whose formal or informal activities and interactions start, import, modify, and diffuse new technologies, characterizing collective innovation efforts (Manzini, 2012). It is *national* because of the central role of spatial proximity and concentration in this process (Acs et al., 2016). *Innovation* means technologies or practices that are new to a given society, made by entrepreneurs, and depend on a society's adoption (The World Bank, 2010). The *system* notion emphasizes cooperation and linkages in the innovation process (Manzini, 2012). As Lundvall (2005) points out, mechanistic versions of NIS denote something that can be constructed, governed, and manipulated by policymakers. When applied to developing countries, the emphasis is on system building and promotion (Lundvall, 2007b).

Modes of innovation

Within NIS, there are two modes of innovation: the STI mode - comprehending learning from *science, technology, and invention*, and the DUI mode - encompassing learning by *doing, using, and interacting*. The STI mode produces CKO (such as scientific papers, patents, books, presentations at conferences, etc.). On the other hand, the DUI mode produces innovations through non-codified knowledge (or know-how), which is tacit, embodied in people, or embedded in organizations (Lundvall, 2005, 2007a, 2007b; Manzini, 2012; Atkinson, 2020; Acs et al., 2016; Eggink, 2013; OECD, 1997). The output of each mode of innovation is diverse, and the sensibility to measure them is disparate. The DUI mode subsystem (experience-based) is elusive to measure (Cirillo et al., 2019). Indicators capturing institutions, linkages, policies, and social capabilities, or DUI modes of learning, are less susceptible to quantitative representation. Instead, CKO from the STI mode (science-based) is relatively easy to account for, and there was progress in bibliometrics to improve measurement, both in output quantity and quality (Lundvall, 2007a; Manzini, 2012; Atkinson, 2020; Acs et al., 2016, Eggink, 2013).

Codified products of knowledge

The CKO varies in its degree of public good: something whose consumption is non-rival as well as non-excludable. A patent is a private good (the owner can exclude third parties), and the content of a scientific paper is mostly a public good. Embodied personal knowledge is mostly private. Practices and norms are normally common knowledge within the interior of firms or other institutions. However, the benefits of research generated in one place can hardly be captured in other places. Secrecy would prevent innovation. A technological advantage can thus only be private and locally captured temporarily since people move and knowledge diffuses (Etzkowitz, 2011).

Conversely, as CKO has components of public goods, the incentives of market actors are not adequate to produce the socially desired level of scientific knowledge because of the challenges of appropriating or owning it. Economic theory provides a robust rationale for the public support of only a component of innovation (discovery or invention). In contrast, public financing for applied research and commercialization is debatable because of the private appropriation of benefits through trade secrecy, intellectual property, or maintaining a competitive lead (Schot and Steinmueller, 2018). The "market failure" argument, however, does not guide how much governments should spend on science. Besides the public good argument, uncertainty (another market failure) may also prevent firms from investing in innovation. Empirically, the most used appropriation methods are lead time and secrecy, the complexity of design, and trademarks (Faberger, 2017).

Latecomers, in comparison with first movers, are challenged with many disadvantages in developing their innovation capabilities, such as technological leadership of incumbents, preemption of assets, and buyer switching costs, but they benefit from free-rider effects, information spillovers and learning from the experiences of pioneers (Fan, 2014).

Institutions within NIS

The differences in NIS quality depend on the quality of "institutions" (Bartels et al., 2014). Institutions are intended as organizations, as well as '*habits, routines, rules, norms, and laws, which regulate the relations between people, and shape social interaction*'. Some of these interactions may be cooperative, while others may be competitive. The linkages between agents can be formal or informal, intentional or incidental, conscious or not conscious, and synergetic or not (Eggink, 2013).

The historical role of universities has been to establish what is considered 'reasonably reliable knowledge.' They had enjoyed relative autonomy from the state as well as from private interests. The primary function of universities remains to train people to solve complex problems (Heller and Eisenberger, 1998). In the late 19th century, research was added to teaching as a second university mission. In the USA, at the time, funds from philanthropists were given to fund new universities and expand old ones. There were concerns among academics that the gifts would try to influence professors' hiring and firing, as well as to decide research priorities. To preserve independence for science from economic interests, a doctrine of pure research was promoted. In 1942, Merton stated the normative structure of science with an emphasis on universalism and skepticism as a response to Nazi and Soviet political control of knowledge to also protect science from politics. The third element in establishing scientific autonomy was the Bush Report of 1945. The distribution of government funds to academic research was assigned to ''peer reviewers'', a criterion adapted from foundation practices in the 1920s and 1930s. Endowed with higher education and research goals, the increased role of knowledge and research in economic development opened the third mission for universities after WWII, which is the promotion of economic development, more pronounced since the end of the Cold War (Etzkowitz and Leydesdorff, 2000).

The so-called Triple Helix of university-industry-government relations states that the university can promote innovation in knowledge-based societies. Most countries and regions are presently trying to attain some form of Triple Helix, with university spin-off firms and strategic alliances among firms, government laboratories, and academic research groups (Etzkowitz and Leydesdorff, 2000). The model is analytically different from the NSI approach, in which entrepreneurs lead innovation, and from the "Triangle" model of Sábato (1975) and Sábato and Mackenzie (1982) in which the nation-state encompasses academia and industry and directs the relations between them. Its strongest version was the Soviet-type system. The weakest versions were present in Latin America. Both experiences are deemed as failed developmental models, with little "bottom-up" initiatives, and where innovation was discouraged rather than encouraged. Higher education and training systems that assist only public administration or produce large numbers of underemployed scholars do not promote innovation (Lundvall, 2007a). Another policy model consists of separate institutional spheres with strong borders dividing them and highly circumscribed relations among the spheres, exemplified in Sweden and the US (Etzkowitz and Leydesdorff, 2000).

Nevertheless, Lundvall (2007b) argues that American tendencies in pharmaceuticals and biotechnology face the risk of being generalized to the relationships between universities and industry in general, inspiring reforms that neglect other universities' functions. The great US entrepreneurial universities rest on a national policy of funding mission-oriented research areas mainly for defense and health (Etzkowitz, 2015), largely to federal labs, and support for basic, curiosity-directed research through university funding (Atkinson, 2020; Faberger, 2017). Lundvall (2007a) adds that the long-term implications and costs of making scholars and universities profit-oriented seem to be that scholars become less engaged in sharing their knowledge otherwise salable.

Teaching guarantees universities a comparative advantage as a source of innovations over other forms of knowledge producers, such as consultants, which is student turnover. In solving clients' problems, a consulting company reunites together dispersed personnel transiently for individual projects and then disperses them again after projects are completed. They lack a cumulative research program. The university combines organizational and research memory with flows of new persons and new ideas (Etzkowitz and Leydesdorff, 2000; Etzkowitz, 2011).

Two established models co-exist in STI innovation policy discussions. The first began with a post-WWII institutionalization of government support for CKO, seeking economic growth and addressing market failure in the private generation of new knowledge. The second emerged in the 1980s and focused on building links, clusters, and networks, stimulating learning between elements in the systems, and enabling entrepreneurship (Schot and Steinmueller, 2018).

Innovation: process or system?

There seem to be two ways to conceptualize the role of knowledge in innovation activities: a process or a system. Before the early 1970s, theorists studied innovation in terms of a process composed of "sequences" and "stages" or "chains" of activities (Godin, 2017). The linear model of innovation begins with basic research, followed by applied research, development, and commercialization. In this, innovation is seen as a process made up of sequential stages that are temporally and conceptually distinct and characterized by unidirectional causality (Guan and Chen, 2012). The conception of a linear innovation model was first proposed by White House science advisor Vannevar Bush in the post-war period, and it was based on the notion that funding basic research will lead almost automatically to innovation (Fan, 2014).

On the other hand, between 1930 and 1950, official statisticians started to define, classify, and register basic research, applied research, and development data. In 1951, the National Science Foundation (NSF) was mandated by

law to measure scientific and technological activities in the USA. The organization developed surveys on R&D based on precise definitions and categories. Industrialized countries followed the NSF definitions when they adopted the OECD Frascati manual in 1963. The manual offers methodological conventions that allowed international comparisons (Godin, 2017).

Before the linear model, there were other process models. One is the invention-diffusion framework. It came from anthropologists in the 1920s and 1930s and served to analyze changes in culture among societies. Another early process model since the 1940s is the stage model from rural sociologists, who studied the diffusion of innovation as a sequential process. Criticism of the linear model gave rise to the demand-pull model (c. 1965), which places the origin of the process of innovation on social needs or market demand instead of a supply perspective. The idea became formalized into a demand-pull model in the 1970s and 1980s, which was of limited use in explaining technological innovation (Godin, 2009).

A new kind of explanation appeared in the post-WWII era: the system model. The system concept was popular in the 1950s and 1960s. The NIS approach suggests that the research system's goal is technological innovation and that it is part of a larger system composed of government, university, and industry. The approach also emphasizes the relationships between the components or sectors to explain the performance of innovation systems. The NIS approach is due to researchers such as Chris Freeman, Richard Nelson, Bengt-Ake Lundvall, and early OECD work from the 1960s. The NIS framework has been very influential as a rationale for the development of national policies to stimulate technological innovation (Godin, 2017).

The actors in the NIS innovation model have a division of labor and responsibility. Scientists are expected to pursue scientific advancement and publish their results, disclosing their methods and findings. The public sector is expected to fund scientific research. The private sector transforms scientific discoveries into innovations that support economic growth. The NIS approach is thus complementary to a competitiveness agenda (Schot and Steinmueller, 2018). Both tacit knowledge (or know-how) exchanged through informal channels and codified knowledge are inputs for innovation (OECD, 1997).

The most traditional type of knowledge flow in a NIS may be technology dissemination in the form of new equipment and machinery. However, the innovative performance of firms increasingly depends on adopting and using innovations and products developed elsewhere. The movement of people and the tacit knowledge they carry with them is key in NIS. Personal interactions are important channels of knowledge transfer (OECD, 1997).

MATERIAL AND METHODS

The following three subsections discuss the variables and data, the method we employ, and the models we estimate.

Data

Our data is a combination of country-level sources. Table 1 presents the variable definitions, classifying them according to their role in the estimates. One of the main concepts of the Frascati manual was GERD (gross expenditures on R&D), defined as the sum of the expenditures from the four main economic sectors of the economy: government, university, industry, and nonprofit (Godin, 2017). R&D expenditures are "*current and capital expenditures (both public and private) on creative work undertaken systematically to increase knowledge, including knowledge of humanity, culture, and society, and the use of knowledge for new applications*" (Godin, 2017). In a production frontier, GERD represents the non-human resources, and in a cost frontier (our concern), it is the cost of production of the R&D outputs, the dependent variable. GERD is expressed in the US dollar, at PPP constant values of 2010, attributes which allow comparisons between countries and years. On the other hand, according to the World Bank (2010), researchers in R&D are "*professionals engaged in the conception or the creation of new knowledge, products, processes, methods, or systems, and the management of the projects concerned*" (Godin, 2017). Researchers are the human resources in a production frontier, and for a cost frontier, it is an important variable to compute, along with GERD, the relative price of inputs.

Our analysis runs different specifications using alternative measures for the outputs. We report the production of published documents or of citable published documents, which are a subset of the former (correlation 0.99). In the same vein, we report patenting by patent publication or patent grants (correlation 0.90). We include an input relative price, a time trend, and some environmental variables. These include the per capita GDP. We also defined some partial productivity indicators that are useful to characterize and compare countries and to give consistency to efficiency analysis. Inputs are human and non-human, the latter measured in monetary units. All monetary issues were converted to constant 2020 dollars at PPP values since the cost of living, salaries, and cost of materials are different among countries. Concerning the environmental conditions, we try to address the differences in costs between arts and social sciences publications and natural sciences ones through a dummy, and to identify the "modernity" of the NIS, we developed a dummy to differentiate between patents that we characterize as belonging to IV Industrial Revolution.[1]

Table 2 shows the descriptive statistics of the variables included in the analysis. We use an unbalanced panel of 82 countries over 24 years, from 1996 to 2019[2].

1 The characterization of the technologies in each industrial revolution (IR) is as follows:
1. The First IR used water and steam power for mechanization.
2. The Second IR applied electricity to create mass production.
3. The Third IR employed electronics and information technology for automation.
4. The Fourth IR combined physical, digital, and biological technologies in disruptive ways (Lacy et al., 2019).

2 To get the final number of observations, we first dropped countries with incomplete information, and we removed the countries that contributed less than 0.005% of total publications.

Name	Type	Definition
gerd	Cost	Dollar 000, PPP constant values of 2010, according to UNESCO.
docs	Output	Published documents, according to the SCIMAGO database
citabledocs	Output	Citable published documents, according to the SCIMAGO database
patpublications	Output	Patent publications, according to the WIPO database
patgrants	Output	Patent grants, according to the WIPO database
w	Relative Price of Human and non-Human Inputs	Dollar 000, PPP constant values of 2010, according to UNESCO on Number of researchers full-time equivalent, according to UNESCO
gdppc	Environmental	Per capita GDP (PPP values) in constant dollars of 2010, according to World Bank
heritageeconomicfreedom	Environmental	Global Heritage Economic Freedom Index, according to Heritage Foundation
gerdpc	Environmental	Gerd/Inhabitants
socialdocsshare	Environmental	Share of social sciences and art disciplines on total published documents
socialcitabledocsshare	Environmental	Share of social sciences and art disciplines on total citable published documents
ivirpatpublicationsshare	Environmental	Share of IV Industrial Revolution Technologies on Total Patents Publications
ivirpatgrantsshare	Environmental	Share of IV Industrial Revolution Technologies on Total Patent Grants
trend	Time trend	1 for 1996 to 23 for 2019
sqtrend	Time trend squared	Trend squared
	Partial productivity	
doc_on_res		Docs/researchers
citabledocs_on_res		Citabledocs/researchers
patpublications_on_res		Patpublications/researchers
grants_on_res		Patgrants/researchers
	Average costs	
GERD_on _docs		Gerd/docs
GERD_on_citabledocs		Gerd/citabledocs
GERD_on_patpublications		Gerd/ Patpublications
GERD_on_patgrants		Gerd /Patgrants

note: Researchers are counted as Full-Time Equivalents.

Sources: Table 1: Variable definitions (source: Authors' elaboration on Scimago Journal & Country Rank, https://www.scimagojr.com/countryrank.php, UNESCO Institute for Statistics (UIS), http://data.uis.unesco.org/, WIPO Information Resources on Patents, https://www.wipo.int/patents/en/, World Bank Open Data, https://data.worldbank.org/, Heritage Foundation Index of Economic Freedom, https://www.heritage.org/index/download)

	Observations	Mean	Sd	Min	Max
Gerd	1189	19728.50	57358.19	20.60	444589.66
docs	1189	66948.47	148742.54	142.00	1213339.00
patpublications	1189	55205.47	194938.96	2.00	2922482.00
citabledocs	1189	75329.43	175254.25	136.00	1337148.00
patgrants	1189	15339.31	47559.53	1.00	361771.00
W	1189	143.69	96.42	10.57	978.02
GDP per capita	1189	24876.13	21319.87	234.00	111968.00
Overall Score Heritage Economic Freedom	1189	66.41	9.35	41.80	90.20
GERD per capita	1189	358.33	387.26	1.00	1691.00
socialdocsshare	1189	0.09	0.05	0.01	0.31
ivirpatpublicationsshare	1189	0.40	0.14	0.00	0.83
ivirpatgrantsshare	1189	0.65	0.15	0.00	1.00
socialcitabledocsshare	1189	0.08	0.05	0.01	0.32
Researchers (FTE)	1189	110763.19	248438.50	142.00	1866109.00
docs_on_res	1189	0.90	0.72	0.03	5.75
citabledocs_on_res	1189	0.95	0.74	0.03	5.66
patpublications_on_res	1189	0.12	0.19	0.00	1.42
grants_on_res	1189	0.08	0.10	0.00	0.78
GERD_on _docs	1189	201.00	152.73	17.95	2152.50
GERD_on_citabledocs	1189	188.51	147.20	14.05	2218.73
GERD_on_patpublications	1189	21946.19	119677.48	110.18	2585170.00
GERD_on_patgrants	1189	22250.35	131032.18	146.67	3834244.00

Table 2: Descriptive statistics (sources: Authors' elaboration on Scimago Journal & Country Rank, https://www.scimagojr.com/countryrank.php, UNESCO Institute for Statistics (UIS), http://data.uis.unesco.org/, WIPO Information Resources on Patents, https://www.wipo.int/patents/en/, World Bank Open Data, https://data.worldbank.org/, Heritage Foundation Index of Economic Freedom, https://www.heritage.org/index/download)

Method

Efficiency in the production of codified outputs of knowledge in the STI mode of NIS is the focus of this assessment. The simplest possible approach consists of computing simple measures of partial productivity (i.e., output/input ratios) or average costs (i.e., costs/output ratios). These approaches neglect relations of complementarity and substitution between inputs and synergies of joint production in outputs. Most sophisticated techniques use frontier approaches, such as mathematical programming methods and econometric estimates. Inputs are usually represented by indicators such as the amount of R&D investment and the number of researchers in R&D, whereas output measures are reflected by indicators such as patents and scientific and technical journal paper publications. These data are territory-based.

The SFA approach decomposes the deviations of each observation from the frontier (residues) into two components: a stochastic error term and an inefficiency term. In a panel data context, where multiple decision-making units (DMU) and periods exist, SFA permits efficiency to vary within a DMU, over time, and among DMU. Accordingly, panel data SFA models can be classified into four groups:

1. Models with invariant inefficiency both in time and DMU (Pitt and Lee, 1981; Battese and Coelli, 1988).
2. Models with time-varying and DMU invariant inefficiency (Kumbhakar, 1990; Battese and Coelli, 1992).
3. Models with both time and DMU varying inefficiency (Battese and Coelli, 1995, Greene 2005a, and Greene 2005b).
4. 4) Models with persistent and residual inefficiency and with unobserved heterogeneity were considered across DMU (Kumbhakar and Heshmati, 1995; Kumbhakar et al., 2014).

The most used production (cost) function specifications are the Cobb-Douglas in logarithms and the Translogarithmic (Translog) defined, respectively as

$$\ln y = \beta_0 + \sum_{n=1}^{N} \beta_n \ln x_n \tag{1}$$

$$\ln y = \beta_0 + \sum_{n=1}^{N} \beta_n \ln x_n + \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} \beta_{nm} \ln x_n \ln x_m \tag{2}$$

In the former, y represents output(costs) and x inputs(outputs) in production(cost) frontiers, respectively.

The Translog is more flexible, not assuming constant elasticities over the full sample and considering quadratic effects and the possible interactions (complementarity, substitution, or no interaction) between the inputs (outputs in cost frontiers). The elasticities of the Translog frontier are:

$$\frac{\partial \ln y}{\partial \ln x_n} = \beta_n + \sum_{m=1}^{N} \beta_{nm} \ln x_m \tag{3}$$

Battese and Coelli (1995) propose a model in which $_{it}$ can be influenced by DMU-specific effects, exogenous determinants, or covariates, z_{it}, uncorrelated with the regressors of the frontier. In these time-varying SFA models, the intercept α is the same across all DMU (Belotti et al., 2013), not addressing time-invariant unobservable factors, assumed to be random on DMUs over time. Thus, their performance is underestimated. We employ the Battesse and Coelli (1995) model, where:

$$y_{it} = \alpha + x_{it}\beta + v_{it} - Su_{it} \tag{4}$$

And

$$u_{it} = z_{it}\delta + W_{it} \tag{5}$$

Where:

$S = 1$ *for production frontiers*, and $S = -1$ *for cost frontiers* where y_{it} represents the output(cost) for the i DMU in the t period; x_{it} denotes a vector of inputs(outputs) for the DMU (country in this case) i in the t period, β is a vector of parameters. The composed error term ε_{it} is the sum (or difference) of v_{it}, representing statistical noise, and a one-sided disturbance u_{it}, addressing for inefficiency. S assumes the value of 1 in production frontiers and -1 in cost frontiers.

The error term is expressed as the sum of two terms, u_{it} and v_{it}, which are assumed independent of each other, as well as independent and identically distributed.

$$e = v_{it} + u_{it} \tag{6}$$

The first part of the error term is a random error with distribution independent and identically distributed to account for possible noise, data typing, or reporting errors.

$$v_{it} \sim N\left(0, \sigma_v^2\right) \tag{7}$$

The second part of the error term is the inefficiency itself, and it accounts for unobserved factors that are in control of the decision unit.

$$u_i \sim N^{+}\left(\mu, \sigma_u^2\right) \tag{8}$$

The SFA model is usually estimated through maximum likelihood (ML) methods in two steps: firstly, the estimation of the parameters of the model and secondly, the point estimates of inefficiency through the mean of the conditional distribution:

$$E(u_{it} | v_{it} + u_{it}) \tag{9}$$

In Battese and Coelli (1995), parameters of the SFA and the model for the technical inefficiency effects are estimated simultaneously by maximum likelihood. The likelihood function is expressed in terms of the variance parameters for the compound error term σ^2, the sum of the variances $\sigma_v^2 + \sigma_u^2$ and the ratio between the variances $\gamma = \frac{\sigma^2_u}{\sigma^2}$, where $\gamma \in (0;1)$.

If $\gamma = 0$, volatility is fully explained by randomness, while if it is the unit, inefficiency explains the whole volatility.

Models

We estimate two Translog models. The dependent variable is the logarithm of GERD in constant 2010 PPP values, representing the cost of CKO of each country, regressed against the logs of its outputs (scientific publications -docs- or citable

scientific publications -citabledocs-; patent publications -patpublications- or patent grants -patgrants-), its squared and interaction (cross-) effects, and the logarithm of the relative price of human and non-human inputs (w). We added some environmental variables to capture the level of economic development of the country (logarithm of per capita GDP), the level of institutional development of the country (logarithm of Economic Freedom Index of Heritage Foundation), the importance of the activity in the country (logarithm of the per capita GERD), the share of publications which demand lower non-human resources (arts and social sciences publications or citable publications), the share of IV Industrial Revolution patents on total (patent publications or patent grants).

Variables	Model A	Model B
Costs (dependent)	lgerd	lgerd
Outputs (linear, cross-, and squared effects)	ldocs	---
	---	lcitabledocs
	lpatpublications	---
	---	lpatgrants
	lsqdocs	---
	---	lsqcitabledocs
	lsqpatpublications	---
	---	lsqpatgrants
	ldocspatpublications	---
	---	lcitabledocspatgrants
Input relative prices	lw	lw
Environmental	lgdppc	lgdppc
	lheritageeconomicfreedom	lheritageeconomicfreedom
	lgerdpc	lgerdpc
	socialdocsshare	---
	---	socialcitabledocsshare
	lvirpatpublicationsshare	---
	---	ivirpatgrantsshare

Table 3: Estimated models (source: Authors' elaboration)

RESULTS

The model used for the estimations is Battese and Coelli's (1995) time-varying model of inefficiency.

In Table 4, we present both estimates for models a and B, which are introduced in Table 3. The differences between the two models are the outputs (and their crossed and squared effects). Not all publications are cited, nor are the cited publications the same as the former. There is a lag between the paper being sent to publishing and its finally being published, and there is also a lag between the publication and the new publications citing them. We do not apply lags to publications nor the citable publications. If, say, a couple of years is needed on average to publish, and another couple of years until the former publications start to impact, we could lose four years of observations. Instead, we assume that the current costs are spent to finance the current inputs, while most probably, they are being spent on outputs that will be published in a couple of years. A similar thing happens with patents: a patent granted in the current period had a process initiated in some period in the past. The same is true for patent publications, however, the set of patent grants is different from patent publications, and they are both different from patent presentations. In the case of patents, there is no consensus on the adequate lag to apply. We perform some sensitivity tests, with two years lag, to address these complex issues, and the results are not remarkably different from the main scenario presented here.

The coefficients of outputs are positive as expected in both models, even when the linear coefficients of patpublications and patgrants are not significantly different from zero. Quadratic values are positive for both inputs, and the cross effect is negative and significant, also as expected, because patents and publications compete for the resources they employ (human and non-human inputs, researchers, and money). The log of the relative price of inputs is also significant and positive, as expected.

Concerning the environmental variables, the logarithm of the GDP per capita is negative, indicating that the costs of producing COK decline with the level of development of the country, proxied by the cited variable. Also, production in model a declines with the Heritage Foundation Index of Economic Freedom, while it is not significantly different from zero in model B. The fourth industrial revolution type of patent publication reveals no significant difference from zero in model A, while the same consideration made for patent grants is significant and negatively affects costs. This can be explained by the synergy of different types of

technologies in the Fourth Industrial Revolution's type of inventions, as stated by Lacy et al. (2019). Social sciences published documents and citable documents reveal both as significant and negatively correlated with costs. This is reasonable since the production costs of the remaining papers in natural sciences, medicine, or engineering are more expensive to produce in terms of laboratories, materials, experimentation, etc. Finally, the sign of the time trend is negative, indicating in the case of model a that costs are decreasing at a rate of -1.38 percent per year on average, and for model B, at -1.62 percent yearly.

The value of lambda is high, indicating that the standard deviation of the inefficiency component is nearly nine times the standard deviation of the pure randomness component of the composite error term ($u_{it} + v_{it}$).

It is worth mentioning that $\lambda = \frac{\sigma_u}{\sigma}$, $(0 < \lambda < 1)$, where λ is the ratio between the standard deviation of u (σ_u and σ, which is the sum of the standard deviation of v and u (σ_v and σ_u). If $\lambda = 1$, the residual variability can be totally explained by the efficiency component u. Instead, if $\lambda = 0$, all the residual variability is randomness.

	Ln(gerd)		Ln(gerd)
Model A		Model B	
Ln(docs)	0.538***	Ln(citabledocs)	0.688***
	(0.0732)		(0.0786)
Ln(patpublications)	0.0393	Ln(patgrants)	-0.0226
	(0.0443)		(0.0470)
Ln(docs)*Ln(patpublications)	-0.0493***	Ln(citabledocs)*Ln(patgrants)	-0.0325**
	(0.0168)		(0.0146)
Ln(docs)^2	0.0276***	Ln(citabledocs)^2	0.0144**
	(0.00626)		(0.00591)
Ln(patpublications)^2	0.0191***	Ln(patgrants)^2	0.0212***
	(0.00345)		(0.00287)
lnw	0.508***	lnw	0.523***
	(0.0196)		(0.0203)
lngdppc	-0.505***	lngdppc	-0.537***
	(0.0247)		(0.0259)
lnheritageeconomicfreedom	-0.167*	lnheritageeconomicfreedom	-0.0629
	(0.101)		(0.105)
lngerdpc	0.369***	lngerdpc	0.390***
	(0.0232)		(0.0232)
ivirpatpublicationsshare	0.0974	ivirpatgrantsshare	-0.285***
	(0.0728)		(0.0789)
socialdocsshare	-2.271***	socialcitabledocsshare	-1.995***
	(0.273)		(0.308)
trend	-0.0138**	trend	-0.0162**
	(0.00612)		(0.00633)
sqtrend	-0.000314	sqtrend	-0.000115
	(0.000261)		(0.000271)
Constant	8.611***	Constant	8.066***
	(0.484)		(0.504)
Mu	-15.26	Mu	-15.37
	(33.89)		(23.33)
Usigma	1.515	Usigma	1.582
	(2.105)		(1.436)
Vsigma	-2.825***	Vsigma	-2.767***
	(0.111)		(0.103)
Log-likelihood	-465.78	Log-likelihood	-515.90
Prob>chi2	0.0000	Prob>chi2	0.0000
Wald Chi2(13)	45324.04	Wald Chi2(13)	42663.74
SigmaU	2.13	SigmaU	2.20
SigmaV	0.24	SigmaV	0.25
Lambda	8.76	Lambda	8.80
Observations	1189	Observations	1189
Number of countries	82	Number of countries	82

*Note: ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels.*

Table 4: Cost SFA Estimates

In Table 5, we present the efficiency estimates and descriptive statistics of Models A and B. On average, technical efficiency is 0.7770 for Model a and 0.7660 for Model B, respectively. Even though the variables included are different and represent different timing in the publication process, we see that standard deviations and ranges in both cases are similar.

Variable	Obs	Mean	Std. Dev.	Min	Max
TE Model A	1,189	0.7760	0.1360	0.0947	0.9615
TE Model B	1,189	0.7660	0.1422	0.0842	0.9600

Table 5: Technical efficiency for Models a & B

Tables 6 and 7 show Tests for differences in characteristics by TE quantiles. Column 1 shows the average and standard deviation for each quartile of the TE distribution, going from the least to the most efficient countries. The number of countries will not be equally distributed by quartile because we use the average TE by country to split an unbalanced panel. The following columns have the t-tests for the differences by quantile, and lastly, we present a joint orthogonality test for all the distributions. Countries have significant differences in terms of inputs and partial productivity measures when looking at the joint orthogonality test for all the variables by quartiles. When looking at individual differences, the test over the 3rd and 4th quantiles shows the differences between the two most efficient groups of countries. We have positive differences in gerd, docs, citable docs, and patgrants, which means that the most efficient group has less of each of these concepts than the second efficient group. We have a negative difference in docs/res and citable docs/res, which are both partial productivity measures, meaning that the most efficient countries produce more articles and citable articles. However, they also have a higher average cost of production gerd on docs and gerd on citable docs.

Variable	Mean/SE				T-test Difference						F-test
	1st quartile	2nd quartile	3rd quartile	4th quartile	(1)-(2)	(1)-(3)	(1)-(4)	(2)-(3)	(2)-(4)	(3)-(4)	orthogonality
gerd_000	14110.849	25342.021	35838.691	3845.920	-11231.172***	-21727.842***	10264.929***	-1.05e+04*	21496.101***	31992.771***	0.000***
	[1844.626]	[3415.377]	[5258.848]	[337.119]							
docs_000	36.230	90.035	113.677	29.598	-53.806***	-77.448***	6.632*	-23.642	60.438***	84.080***	0.000***
	[3.215]	[9.549]	[13.394]	[2.330]							
patpublications_000	62.426	78.865	70.184	9.189	-16.439	-7.758	53.236***	8.681	69.676***	60.994***	0.000***
	[11.833]	[16.501]	[9.746]	[1.067]							
citabledocs_000	38.253	102.577	130.309	32.296	-64.324***	-92.056***	5.958	-27.732	70.281***	98.013***	0.000***
	[3.359]	[11.386]	[15.833]	[2.546]							
patgrants_000	19.516	18.537	20.420	2.602	0.979	-0.904	16.914***	-1.883	15.935***	17.818***	0.000***
	[3.634]	[2.733]	[2.865]	[0.289]							
W	133.437	148.458	165.559	127.706	-15.021***	-32.122***	5.730	-17.101*	20.751***	37.852***	0.000***
	[4.197]	[3.337]	[8.017]	[5.387]							
docs_on_res	0.561	0.737	1.047	1.255	-0.176***	-0.486***	-0.694***	-0.310***	-0.518***	-0.207***	0.000***
	[0.019]	[0.016]	[0.051]	[0.053]							
citabledocs_on_res	0.584	0.784	1.097	1.356	-0.200***	-0.513***	-0.772***	-0.313***	-0.572***	-0.259***	0.000***
	[0.019]	[0.016]	[0.050]	[0.056]							
patpublications_on_res	0.105	0.124	0.123	0.122	-0.019	-0.018	-0.017	0.001	0.002	0.001	0.536
	[0.013]	[0.009]	[0.009]	[0.011]							
grants_on_res	0.065	0.088	0.086	0.075	-0.023**	-0.022**	-0.010	0.002	0.013*	0.011	0.017**
	[0.008]	[0.005]	[0.005]	[0.005]							
gerd_on_docs	287.597	227.378	176.203	108.251	60.220***	111.394***	179.347***	51.175***	119.127***	67.953***	0.000***
	[11.916]	[6.871]	[6.850]	[2.837]							
gerd_on_citabledocs	276.795	209.280	163.494	99.669	67.515***	113.301***	177.126***	45.786***	109.611***	63.825***	0.000***
	[11.822]	[6.303]	[6.122]	[2.580]							
gerd_on patpublications	37965.295	16245.952	13248.288	19298.674	21719.343*	24717.008**	18666.621	2997.664	-3052.722	-6050.387	0.047**
	[10261.227]	[4456.902]	[3818.279]	[6642.684]							
gerd_on_patgrants	52631.290	9134.886	5655.729	19579.226	43496.404***	46975.562***	33052.064**	3479.157	-1.04e+04**	-1.39e+04***	0.000***
	[13577.388]	[1815.420]	[1119.183]	[4469.435]							
TE by quartile (upper bound)	0.7360	0.8160	0.8605	0.9161							
Num. of countries (Obs)	26 (312)	18 (286)	17 (298)	21 (293)							

*Note: the values displayed for t-tests are the differences in the means across the groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels.*

Table 6: T-test difference by quartile of the technical efficiency distribution Model A

Variable	Mean/SE 1st quartile	2nd quartile	3rd quartile	4th quartile	T-test Difference (1)-(2)	(1)-(3)	(1)-(4)	(2)-(3)	(2)-(4)	(3)-(4)	F-test orthogonality
gerd_000	15014.967	18212.647	40779.308	3980.447	-3197.680	-25764.341***	11034.520***	-22566.661***	14232.200***	36798.861***	0.000***
	[1899.510]	[3206.774]	[5171.379]	[346.349]							
docs_000	39.727	69.136	126.092	30.646	-29.409***	-86.366***	9.081**	-56.956***	38.490***	95.447***	0.000***
	[3.299]	[9.106]	[13.126]	[2.389]							
patpublications_000	65.482	52.403	90.744	9.487	13.079	-25.262	55.994***	-38.341**	42.916***	81.257***	0.000***
	[12.233]	[15.405]	[9.798]	[1.101]							
citabledocs_000	41.830	78.929	144.572	33.447	-37.099***	-102.742***	8.383*	-65.643***	45.482***	111.125***	0.000***
	[3.448]	[10.836]	[15.528]	[2.610]							
patgrants_000	20.400	9.353	28.227	2.680	11.047**	-7.828	17.720***	-18.874***	6.673***	25.547***	0.000***
	[3.757]	[2.295]	[2.988]	[0.298]							
W	145.677	128.047	167.556	132.991	17.630***	-21.879**	12.686*	-39.510***	-4.944	34.565***	0.000***
	[3.995]	[3.630]	[7.914]	[5.545]							
docs_on_res	0.603	0.747	0.959	1.302	-0.144***	-0.356***	-0.699***	-0.211***	-0.554***	-0.343***	0.000***
	[0.019]	[0.018]	[0.052]	[0.053]							
citabledocs_on_res	0.621	0.791	1.012	1.408	-0.170***	-0.391***	-0.787***	-0.221***	-0.617***	-0.396***	0.000***
	[0.020]	[0.017]	[0.051]	[0.056]							
patpublications_on_res	0.121	0.095	0.135	0.122	0.026	-0.015	-0.001	-0.041***	-0.027*	0.013	0.061*
	[0.014]	[0.008]	[0.010]	[0.012]							
grants_on_res	0.073	0.059	0.108	0.072	0.014*	-0.035***	0.001	-0.049***	-0.013**	0.036***	0.000***
	[0.008]	[0.003]	[0.006]	[0.006]							
gerd_on_docs	295.367	185.402	209.955	107.853	109.965***	85.412***	187.514***	-24.552**	77.549***	102.102***	0.000***
	[11.932]	[5.186]	[8.462]	[2.916]							
gerd_on_citabledocs	285.400	171.089	193.345	99.029	114.312***	92.055***	186.371***	-22.257**	72.060***	94.316***	0.000***
	[11.826]	[4.537]	[7.775]	[2.663]							
gerd_on patpublications	33549.939	23294.112	10897.203	19908.183	10255.827	22652.736**	13641.756	12396.909**	3385.929	-9010.980	0.138
	[10473.714]	[5162.270]	[2985.262]	[6874.144]							
gerd_on_patgrants	28679.260	35387.223	3916.853	20800.404	-6707.963	24762.407***	7878.856	31470.369**	14586.819	-16883.551***	0.021**
	[2681.278]	[13907.884]	[666.318]	[4648.083]							
TE by quartile (upper bound)	0.7340	0.8060	0.8500	0.9100							
Num. of countries (Obs)	24 (301)	19 (304)	18(301)	21 (283)							

*Note: the values displayed for t-tests are the differences in the means across the groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels.*

Table 7: T-test difference by quartile of the technical efficiency distribution Model B

Table 8 shows the ranking of countries by GERD participation. We add the other input (researchers), the relative price, and outputs to characterize countries. We include the cumulative summation of countries by quartile. It is worth noticing that the 20 biggest countries of the sample explain more than 92 percent of GERD, 88 percent of the researchers, 82 percent of documents and published documents, and nearly 95 percent of patent publications and grants. The big three, the USA, China, and Japan, explain the 58 percent of the GERD of the sample. Almost 20 percent of the researchers of the sample are in China, and another 20 percent are in the USA. In documents, both published and citable, the United States produces more than China, but in patent publications, China is ahead, while in grants, the USA continues to be the first. The averages mask the growth of China, a country which, at the beginning of the sample, was well behind the USA and had converged steadily. There are differences in productivity and CKO patterns among countries with similar efficiency levels. Take, for instance, South Korea and France, each one spending the same and with a similar number of researchers. France produces more publications, while South Korea produces more patents. The UK and India devote the same non-human resources. Still, the UK, on average, has four times the number of researchers than India, produces many more publications, and has overwhelmingly high patent publications or grants. A similar situation is true for Canada and Brazil. Most of the countries in the twenty biggest are developed. However, there are some big emerging, such as Brazil, India, Russia, and Turkey.

Rank	country	Gerd (%/World's)	Researchers (%/World's)	Docs (%/World's)	Patpublications (%/World's)	Citabledocs (%/World's)	Patgrants (%/World's)
1	United States	31.91000	19.13017	21.47678	18.55775	22.96982	21.56487
2	China	14.97007	20.11941	12.48207	22.32981	13.57460	12.47574
3	Japan	11.53345	11.46194	5.19384	24.94443	4.87247	27.97039
4	Germany	6.61783	5.38924	5.60740	8.25879	5.60793	7.86141
5	South Korea	3.92512	3.95123	2.29399	6.60833	2.13783	9.21395
6	France	3.82045	3.86941	3.95252	3.18725	4.10803	3.79616
7	United Kingdom	2.90494	4.05727	5.99758	2.02159	5.69084	1.96624
8	India	2.83957	1.02663	3.43079	0.23879	3.25227	0.17025
9	Canada	2.10245	2.27026	3.16762	1.03348	2.95609	0.99014
10	Brazil	1.99352	1.26631	1.72083	0.27435	1.63787	0.08036
11	Italy	1.78724	1.64756	3.21153	1.15420	3.23678	1.30946
12	Russia	1.76558	8.14913	2.16681	1.48867	2.14734	2.57173
13	Spain	1.21502	1.84335	2.57986	0.37163	2.50836	0.45294
14	Netherlands	1.05193	0.99109	1.77825	1.63181	1.65450	1.57009
15	Sweden	0.83225	0.78582	1.21145	1.03762	1.12164	1.19054
16	Austria	0.71645	0.46324	0.71074	0.46995	0.68070	0.54683
17	Belgium	0.68323	0.67439	0.99090	0.45206	0.93020	0.47227
18	Australia	0.66740	0.47310	2.47654	0.47137	2.26989	0.44821
19	Turkey	0.55956	0.87786	1.01813	0.12104	0.97128	0.04381
20	Singapore	0.51758	0.42415	0.59071	0.19378	0.54620	0.16075
	Cumulative	**92.41364**	**88.87156**	**82.05835**	**94.84671**	**82.87465**	**94.85613**
21	Mexico	0.49519	0.51102	0.61755	0.06617	0.58141	0.03081
22	Finland	0.48052	0.44653	0.63107	0.57466	0.59988	0.64885
23	Denmark	0.42993	0.51991	0.72054	0.43841	0.66115	0.41084
24	Poland	0.42435	1.18415	1.27633	0.20929	1.26919	0.22844
25	Norway	0.31024	0.36848	0.56190	0.22955	0.52023	0.24420
26	Czech Republic	0.30391	0.44213	0.62149	0.07187	0.60601	0.08136
27	Malaysia	0.28814	0.37981	0.61192	0.03722	0.57152	0.03882
28	South Africa	0.28037	0.25725	0.50467	0.10780	0.45858	0.12443
29	Argentina	0.27301	0.65299	0.38839	0.02896	0.36012	0.01094
30	Switzerland	0.26094	0.16996	1.29668	1.61823	1.20719	1.67879
31	Iran	0.25798	0.37896	1.06568	0.00429	1.09743	0.00278
32	Egypt	0.23755	0.46615	0.41427	0.00694	0.40101	0.00723

Rank	country	Gerd (%/World's)	Researchers (%/World's)	Docs (%/World's)	Patpublications (%/World's)	Citabledocs (%/World's)	Patgrants (%/World's)
33	Portugal	0.22716	0.51574	0.60382	0.03895	0.57865	0.02845
34	Ukraine	0.22542	0.53639	0.39617	0.12105	0.39374	0.22276
35	Thailand	0.20581	0.38863	0.35245	0.01382	0.31911	0.00822
36	Ireland	0.18826	0.25902	0.37480	0.18225	0.33891	0.16065
37	Hong Kong	0.17402	0.29497	0.60928	0.10559	0.57863	0.10367
38	Hungary	0.16179	0.33316	0.35325	0.07702	0.34373	0.06563
39	Greece	0.13445	0.27313	0.57695	0.04346	0.55513	0.04992
40	United Arab Emirates	0.11147	0.04815	0.11196	0.00976	0.10041	0.00638
41	Indonesia	0.09482	0.20499	0.30345	0.00376	0.27759	0.00204
	Cumulative	**97.97897**	**97.50309**	**94.45097**	**98.83578**	**94.69425**	**99.01133**
42	Romania	0.09371	0.35805	0.40516	0.05261	0.39520	0.06908
43	Pakistan	0.08221	0.19578	0.30083	0.00103	0.30542	0.00081
44	Slovenia	0.07805	0.11241	0.17983	0.03846	0.17774	0.05149
45	Colombia	0.07701	0.01312	0.20197	0.01554	0.18635	0.00336
46	Slovakia	0.06457	0.21268	0.21652	0.01942	0.20707	0.01591
47	New Zealand	0.05558	0.12773	0.42443	0.10066	0.38326	0.06501
48	Croatia	0.04457	0.10587	0.18950	0.02129	0.18051	0.01923
49	Bulgaria	0.04419	0.20686	0.14865	0.02090	0.14381	0.02043
50	Chile	0.04286	0.05639	0.28202	0.02625	0.26977	0.00890
51	Tunisia	0.03901	0.14404	0.18574	0.00491	0.17422	0.00093
52	Lithuania	0.03591	0.14125	0.10054	0.00846	0.09984	0.01284
53	Luxembourg	0.03540	0.03094	0.03961	0.11683	0.03654	0.11071
54	Vietnam	0.02872	0.15213	0.11743	0.00174	0.11863	0.00100
55	Algeria	0.02480	0.02993	0.14183	0.00082	0.13646	0.00028
56	Kuwait	0.02396	0.01204	0.04736	0.00258	0.04249	0.00369
57	Estonia	0.02352	0.06035	0.07811	0.01386	0.07788	0.01610
58	Morocco	0.01961	0.16107	0.12576	0.00934	0.12005	0.01005
59	Philippines	0.01793	0.05219	0.06724	0.00689	0.05944	0.00372
60	Costa Rica	0.01722	0.01700	0.02280	0.00247	0.02009	0.00115
61	Ecuador	0.01599	0.02337	0.04395	0.00540	0.04001	0.00052
62	Latvia	0.01444	0.06068	0.05277	0.01269	0.04940	0.01940
	Cumulative	**98.85823**	**99.77698**	**97.82301**	**99.31794**	**97.91841**	**99.44593**
63	Uruguay	0.01286	0.01984	0.03647	0.00407	0.03316	0.00151
64	Oman	0.00993	0.00513	0.03691	0.00040	0.03231	0.00021
65	Sri Lanka	0.00759	0.01943	0.03963	0.00133	0.03627	0.00052
66	Panama	0.00733	0.00431	0.01354	0.00603	0.01175	0.00549
67	Cyprus	0.00705	0.01132	0.05192	0.01694	0.05187	0.01503
68	Ethiopia	0.00592	0.01577	0.04448	0.00003	0.03947	0.00002
69	Moldova	0.00466	0.03237	0.01692	0.01594	0.01541	0.02888
70	Malta	0.00427	0.00816	0.01349	0.01298	0.01197	0.01082
71	Jordan	0.00351	0.01736	0.08142	0.00180	0.07892	0.00144
72	Georgia	0.00344	0.02126	0.03072	0.00591	0.03501	0.01034
73	Kenya	0.00338	0.00864	0.06303	0.00122	0.05485	0.00062
74	Bolivia	0.00290	0.00526	0.01384	0.00016	0.01220	0.00010
75	Paraguay	0.00276	0.00691	0.00561	0.00022	0.00411	0.00004

Rank	country	Gerd (%/World's)	Researchers (%/World's)	Docs (%/World's)	Patpublications (%/World's)	Citabledocs (%/World's)	Patgrants (%/World's)
76	Trinidad and Tobago	0.00223	0.00114	0.01257	0.00051	0.01085	0.00049
77	Madagascar	0.00221	0.00987	0.00775	0.00009	0.00678	0.00006
78	Guatemala	0.00212	0.00394	0.00567	0.00146	0.00493	0.00022
79	Senegal	0.00174	0.01797	0.01833	0.00179	0.01595	0.00327
80	Botswana	0.00132	0.00056	0.01371	0.00016	0.01203	0.00014
81	Ghana	0.00094	0.00289	0.03864	0.00023	0.03413	0.00027
82	Bahrain	0.00025	0.00037	0.01314	0.00070	0.01127	0.00048
	Total	**98.94463**	**99.98950**	**98.38079**	**99.38993**	**98.43167**	**99.52589**

Table 8: Ranking by country (sorted by average participation in World's GERD)

Table 9 shows efficiency estimates from our two estimated models ranked by GERD. Of the top 10 countries, Germany is the most efficient country (0.8429-0.8497), and Brazil is the least efficient (0.5856-0.5003). The rest of the top 10 countries have an efficiency that ranges from 0.7240 to 0.8418. There are some small countries (small should be understood as relative to the size of the country in terms of the world's figure in GERD, researchers, and their products) with good efficiency scores (they can attain relatively high-efficiency levels with low absolute levels of inputs and outputs). Nevertheless, their devoted resources and output yields are very modest in importance. Recall the averages are 0.7670 for Model a and 0.7660 for Model B, respectively.

Rank	Country	Gerd (%/World's)	TE Model A	TE model B
1	United States	31.91000	0.8174	0.8405
2	China	14.97007	0.8008	0.7791
3	Japan	11.53345	0.7240	0.7240
4	Germany	6.61783	0.8429	0.8497
5	South Korea	3.92512	0.8163	0.8177
6	France	3.82045	0.8083	0.8257
7	United Kingdom	2.90494	0.8418	0.8198
8	India	2.83957	0.8373	0.7714
9	Canada	2.10245	0.8004	0.7658
10	Brazil	1.99352	0.5856	0.5003
11	Italy	1.78724	0.8659	0.8678
12	Russia	1.76558	0.5828	0.6114
13	Spain	1.21502	0.8021	0.7920
14	Netherlands	1.05193	0.8701	0.8526
15	Sweden	0.83225	0.8461	0.8368
16	Austria	0.71645	0.7904	0.8061
17	Belgium	0.68323	0.8281	0.8023
18	Australia	0.66740	0.7575	0.7347
19	Turkey	0.55956	0.7279	0.7081
20	Singapore	0.51758	0.7894	0.7470
21	Mexico	0.49519	0.5440	0.4885
22	Finland	0.48052	0.7974	0.7920
23	Denmark	0.42993	0.8123	0.7765
24	Poland	0.42435	0.8717	0.8638
25	Norway	0.31024	0.6457	0.6322
26	Czech Republic	0.30391	0.8631	0.8623
27	Malaysia	0.28814	0.7181	0.7026

Rank	Country	Gerd (%/World's)	TE Model A	TE model B
28	South Africa	0.28037	0.7367	0.7290
29	Argentina	0.27301	0.5558	0.4975
30	Switzerland	0.26094	0.9045	0.8935
31	Iran	0.25798	0.8713	0.8720
32	Egypt	0.23755	0.7686	0.7478
33	Portugal	0.22716	0.7736	0.7618
34	Ukraine	0.22542	0.8031	0.8208
35	Thailand	0.20581	0.5352	0.5272
36	Ireland	0.18826	0.7606	0.7064
37	Hong Kong	0.17402	0.8668	0.8653
38	Hungary	0.16179	0.8479	0.8294
39	Greece	0.13445	0.8458	0.8352
40	United Arab Emirates	0.11147	0.2676	0.2586
41	Indonesia	0.09482	0.3439	0.3336
42	Romania	0.09371	0.8279	0.8267
43	Pakistan	0.08221	0.7279	0.7772
44	Slovenia	0.07805	0.8880	0.8961
45	Colombia	0.07701	0.9031	0.8910
46	Slovakia	0.06457	0.8271	0.8171
47	New Zealand	0.05558	0.8348	0.8068
48	Croatia	0.04457	0.8799	0.8755
49	Bulgaria	0.04419	0.8717	0.8643
50	Chile	0.04286	0.8731	0.8604
51	Tunisia	0.03901	0.9064	0.9087
52	Lithuania	0.03591	0.7256	0.7523
53	Luxembourg	0.03540	0.5772	0.5387
54	Vietnam	0.02872	0.3754	0.4077
55	Algeria	0.02480	0.6724	0.7217
56	Kuwait	0.02396	0.8196	0.8067
57	Estonia	0.02352	0.8680	0.8758
58	Morocco	0.01961	0.4590	0.4589
59	Philippines	0.01793	0.3121	0.2934
60	Costa Rica	0.01722	0.6473	0.6379
61	Ecuador	0.01599	0.4667	0.4419
62	Latvia	0.01444	0.8195	0.7956
63	Uruguay	0.01286	0.8010	0.7873
64	Oman	0.00993	0.8763	0.8848
65	Sri Lanka	0.00759	0.7084	0.7580
66	Panama	0.00733	0.8606	0.8298
67	Cyprus	0.00705	0.8985	0.9034
68	Ethiopia	0.00592	0.8904	0.8662
69	Moldova	0.00466	0.8723	0.8463
70	Malta	0.00427	0.8180	0.7801
71	Jordan	0.00351	0.8737	0.8893
72	Georgia	0.00344	0.8205	0.8607

Rank	Country	Gerd (%/World's)	TE Model A	TE model B
73	Kenya	0.00338	0.7795	0.8130
74	Bolivia	0.00290	0.5117	0.6190
75	Paraguay	0.00276	0.3658	0.3963
76	Trinidad and Tobago	0.00223	0.7912	0.8175
77	Madagascar	0.00221	0.8720	0.8389
78	Guatemala	0.00212	0.6129	0.5975
79	Senegal	0.00174	0.6312	0.7585
80	Botswana	0.00132	0.8192	0.8832
81	Ghana	0.00094	0.9161	0.8988
82	Bahrain	0.00025	0.7606	0.8019

Table 9: Model a and B (Efficiency sorted by average participation in the world's GERD)

DISCUSSION

Efficiency scores are similar between models, even though they represent different knowledge cost functions. Countries have significant differences in terms of inputs and partial productivity measures. The most efficient group has more of gerd, docs, citable docs and patgrants compared to lower quantiles of the TE distribution. In contrast, costs in developed countries are higher than in developing countries. We also show that the costs of producing COK decline with the level of development of the country, the costs of producing knowledge decrease with output volume, and the production costs of papers in natural sciences, medicine, or engineering are more expensive to produce than social sciences. Also, our results show that there is competition between resources for patents and publications and that the lag between the production of an article and its first citation may affect the estimations.

A concern in the literature is the separated presentation of resource and output statistics; some studies engaged with the former, some do in the latter, and hardly both are considered together. We address this issue by generating a database composed of outputs, inputs, costs, and relative prices of the inputs. Aksnes et al. (2017) raise this issue; they investigate methodological problems in measuring research productivity on the national level by comparing official R&D statistics from the OECD with data on publications from the Web of Science for 18 countries. They propose improvements to enhance the comparability of data sources. They point out that resource and output statistics are customarily presented as separated instead of combining them into productivity measurements.

In our study, the unit of analysis is more aggregated than those commonly found in the literature, and we focus on SFA. Nor is SFA superior to DEA; conversely, both methods have relative advantages and disadvantages. A comprehensive review of the application of parametric and non-parametric frontier techniques to analyze Research and Development (R&D) systems efficiency can be found in Bonaccorsi and Daraio (2004). Also, Bonaccorsi and Daraio (2003), as an example of more aggregated studies than ours, analyze data on scientific productivity at institutes of the French INSERM and biomedical research institutes of the Italian CNR for the year 1997. Available data on human capital input and geographical agglomeration allows the estimation and comparison of efficiency measures for the two institutions.

Quality of contributions is an important discussion in science and technology efficiency and productivity measurement. In our database, and because of its aggregation level, we cannot address more precisely quality aspects. Nevertheless, the generation of environmental variables points to solving this issue. We are aware that qualitative aspects have a subtle but important difference from environmental ones: quality addition implies volition and deliberate efforts, while environmental conditions can be passive from the point of view of the NIS (for example, NIS cannot influence the global quality of national institutions). For instance, using the Science and Engineering Indicators report of the US National Science Foundation, Bornmann et al. (2018) investigated 21 countries' literature cited in top-quality journals from 2004 to 2013, assuming citation as a qualitative distinction for publications. China has emerged as a major player in science. However, in the Bornmann et al. (2018) sample, China remains a low contributor in the citations of the top 1 percent of articles. That can be attributed to the recent growth of this country to the pool of contributions; on the other hand, citations are a proxy of the quality of the contribution.

Publication in scientific journals is a product of inventive effort; however, it is more an indicator of scientific exploration than of commercialization. Thus, scientific innovation can be perceived as the non-commercial final output. For us, it is a challenge to measure the DUI outputs and inputs to perform a sequel of this paper's analysis, the part we do not cover in this article on NIS efficiency. In the literature, we find that Guan and Chen (2012) propose a relational network data envelopment analysis (DEA) model for measuring the innovation efficiency of the 22 OECD countries' NIS by decomposing the innovation process into a two-stage production framework: an upstream STI knowledge production process, and a downstream DUI knowledge commercialization process. They identify in most countries a significant rank difference between STI and DUI subsystems, indicating a non-coordinated relationship between both stages. The empirical study benchmarked the relative efficiency of the two internal NIS sub-processes of 22 OECD nations. It also explored the determinants

of variations in efficiency across those nations in the two individual sub-processes.

Universities and similar institutions are evaluated either by peer review or by bibliometrics, which is cheaper and more objective than peer review, although biased to scholars and disciplines with relatively intensive publication activity. Results change according to each scientific field and technique applied. In our case, we address the differences in costs between publications in social versus natural sciences and corroborate that the latter are more expensive. Preceding us, Coccia (2008) addresses how is it possible to separate high performing from low-performing research units within each research field, recognizing the differences.

We find that the most important contributors to global R&D expenditure are not necessarily ranked as world-top performers. Several recent studies address efficiency and productivity measurement in science and technique on a national basis. Carrillo (2019) assesses the R&D efficiency of countries using DEA. Afterward, the overall performance score is obtained with the cross-efficiency method, and countries are listed according to their R&D performance. Switzerland, the United Kingdom, and the Netherlands are the three leading countries. The sample of Carrillo (2019) comprises 33 countries with significant involvement in R&D activities (above 1 percent of the World's activity), to which efficiency scores were obtained with an output-oriented VRS DEA model. Also, Ferro and Romero (2021), using a Data Envelopment Analysis (DEA) efficiency frontier approach, study which countries are more efficient at producing scientific articles and patents. They find efficient countries that are both small and not traditional knowledge producers. When scale and regional effects are controlled, the results favor developed countries and Eastern and Central Asian ones.

There are some small countries with good efficiency scores. Nevertheless, their devoted resources and output yields are very modest in importance. The small country issue is puzzling and already discussed in the literature. Kotsemir (2013) reviews the application of the DEA method for measuring the efficiency of national innovation systems (NIS), providing a comprehensive review of 11 empirical studies on a cross-country analysis. When "small" (in terms of national innovation system scope and the level of development) countries are included in the country sample, those become the efficient ones. In general, the studies use samples from less than 30 countries in the studies. The most efficient national innovation systems are OECD countries, normally overrepresented in the samples because of data availability.

Since the main drawback of the SFA approach is that it cannot include multiple outputs in its production analysis, there are two possibilities to overcome the problem: one is the cost function analysis we apply in this study, and the other is the distance function approach that is an appropriate method for the multiple input-output frameworks of SFA. Hu et al. (2014) apply the distance function approach for stochastic frontier analysis (SFA) to compare R&D efficiency across 24 nations during 1998-2005. R&D expenditure stock and R&D manpower are treated as inputs, while patents, scientific journal articles, royalties, and licensing fees are the outputs. Intellectual property rights protection, technological cooperation among business sectors, knowledge transfer between business sectors and higher education institutions, agglomeration of R&D facilities, and involvement of the government sector in R&D activities are environmental conditions that significantly improve national R&D efficiency.

The discussion on the scale is also present in the R&D efficiency debate. Some of the big ones in the top ten are developing countries, which are consistent with the trend of decreasing costs. Most of the countries in the twenty biggest are developed. However, there are some big emerging, such as Brazil, India, Russia, and Turkey. Nasierowski (2010) aims to clarify whether the so-called innovation leaders are efficient in transforming innovation inputs into outputs. Based on the European Innovation Scoreboard (EIS), the efficiency of investment in innovation is examined with the use of the DEA model. It is observed a similar phenomenon as we observed in our sample: the so-called laggards in innovation are often efficient in their use of resources, whereas leaders of innovation fall short in returns to scale and congestion.

Previous empirical results indicate that the overall technical inefficiencies of the NIS activities in European and Asian countries are primarily due to pure technical inefficiencies rather than scale inefficiencies. This is also visible in our cost-efficiency study. Pan et al. (2010) apply the traditional DEA models, bilateral models, and critical performance measures, respectively, combining multiple outputs and inputs to measure the magnitude of performance difference between NIS in 33 Asian and European countries. The bilateral comparison analysis indicates that the Asian group is a better performer than the European group in production activities.

As already mentioned, innovation leaders do not always have the most efficient innovation systems, and modest innovators are not necessarily inefficient in transforming innovation inputs into outputs of innovation. Matei and Aldea (2012) measure and compare the performance of some NIS using the IUS 2011 database to estimate efficiency.

The big three, the USA, China, and Japan, compete and alternate in productivity and efficiency rankings. The relative price between non-human and human inputs reflects the relative intensity of non-human resources technology of production in the USA compared to other countries. Nasierowski and Arcelus (2003) present a non-parametric approach to identify the extent to which a decrease in the productivity growth of many countries can be explained by differences in efficiency and differences in scale and congestion. The model recognizes two types of outputs as the result of the R&D process: patents and their spillover effect onto the economic base of the country. The database consists of the countries included in the World Competitiveness Report.

Environmental conditions are important to explain differences in the performance of NIS since "institutions" vary between national realities. Carvalho et al. (2015) examine the socio-economic factors that contribute to the EU's innovative performance, using two linear regressions, considering as dependent variables, respectively, the patents required and the percentage of innovative sales. This study concludes that the most important explanatory variables for patents are private

R&D expenditure, percentage of innovative firms, and public R&D. Similarly, addressing environmental or contextual issues, Coccia and Rolfo (2007) investigate the relationships between organizational changes and productivity in public research institutions within the Italian national system of innovation, during the period 1999-2003, which is characterized by mergers and consolidation among research units. Their sample is analyzed through DEA and applied to researchers, technicians, administrative staff, cost of personnel as inputs, and the number of domestic and international publications as outputs. They find that new policy is generating lower research productivity and scale diseconomies in larger laboratories due to the bureaucratization of these larger new bodies.

Our national focus has obvious limitations. Knowledge production is an increasingly global endeavor. Despite robust increases in scientific production by traditional leaders, their relative share has decreased in recent decades because the pace of growth in science by other nations has been even more rapid. The share of international collaborations has also increased, as has the share of citations to papers with foreign authors. However, location retains considerable importance in science (Packalen, 2019) because borders continue to influence scientist interactions and because many important science policy decisions are set at the national level.

CONCLUSIONS

Endogenous growth models emphasize the importance of knowledge to generate sustained economic growth. There are several explanations of how knowledge is produced and is conducive to innovation. An encompassing concept in this discussion is NIS, which highlights the interlinks between different kinds of actors to produce knowledge aimed at innovation. The NIS can be split into two subsets: one based on scientific and technological work, producing codified products (scientific publications and patents of inventions), and the other centered on practical and non-codified actions to diffuse, apply, and use knowledge. Our objective is to measure the cost efficiency of the codified knowledge outputs, which are produced with human and non-human resources. In the literature there are inventories of resources and outputs, often studied separately, there are also partial productivity indexes tempting to compare performance, and frontier studies are trying to capture the efficiency of the whole process. The frontier studies are developed as empirical assessments that resort to mathematical programming or econometric techniques.

We examine efficiency using an SFA model; adding to the two versions of explanatory cost frontiers, we estimate some environmental conditions to address differences between development levels of the countries and types of patented technologies to differentiate social from natural sciences in the production of publications, etc. Our database uses information from different sources on scientific publications and patents for 82 countries for 23 years, totaling 1189 observations. Patents and publications are produced by human resources (researchers) together with non-human inputs (funds).

In the sample, 20 out of 82 countries explain more than 92 percent of the financial resources devoted to research and development, 88 percent of the researchers, 82 percent of documents and published documents, and nearly 95 percent of patent publications and grants. The average efficiency of the estimates is in the order of 0.77, indicating 23 percent of cost redundancy. Of the biggest countries in the sample, the United States, spending 32 percent of the sample costs, has efficiency scores of 0.82 to 0.84, depending on the model. China, which is the second country in importance, has an efficiency score of 0.80 to 0.78, depending on the specification.

The growth of China in the last two decades is impressive. Among developed countries, the most efficient are Switzerland and the Netherlands. In Latin America, the best performers are Colombia and Chile by far, while Brazil, Argentina, and Mexico have poor efficiency scores, on the verge of 0.50. There are small countries by their participation in the sample and by all criteria (population, GDP, territory, scientific tradition) that perform well, even though the absolute levels of output and inputs are modest.

The next is to examine efficiency in the DUI subsystem of NIS and the interactions between DUI and STI subsystems, which is challenging because of the difficulty of measuring DUI outputs.

ACKNOWLEDGEMENT

We are grateful to two anonymous referees whose valuable and thorough comments improved the manuscript.

REFERENCES

Acs, Z. J., Audretsch, D. B., Lehmann, E. E., and Licht, G. (2016) 'National Systems of Innovation', *Journal of Technology Transfer. The Journal of Technology Transfer*, Vol. 42, No. 5, pp. 997-1008. https://doi.org/10.1007/s10961-016-9481-8

Aksnes D., Gunnar Sivertsen, W., 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

Atkinson, R. D. (2020) 'Understanding the U.S. National Innovation System', *Information Technology & Innovation Foundation*. https://doi.org/10.2139/ssrn.3079822

Barro, R. J. and Sala-i-Martin, X. (2003) Economic Growth. 2nd Edition. Cambridge: The MIT Press.

Bartels, F. L., Voss, H., Lederer, S., and Bachtrog, C. (2012) 'Determinants of National Innovation Systems: Policy Implications for Developing Countries', *Innovation*, Vol. 14, No. 1, pp. 2-18. https://doi.org/10.5172/impp.2012.14.1.2

Battese, G. E., and Coelli, T. J. (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

Battese, G. E., and Coelli, T. J. (1992) 'Frontier production functions, technical efficiency and panel data: With application to paddy farmers in India', *Journal of Productivity Analysis*, Vol. 3, No. 1-2, pp. 153-169. https://doi.org/10.1007/BF00158774

Battese, G. E., and Coelli, T. J. (1995) 'A model for technical inefficiency effects in a stochastic frontier production function for panel data', *Empirical Economics*, Vol. 20, No. 2, pp. 325-332. https://doi.org/10.1007/BF01205442

Belotti, F., Daidone, S., Ilardi, G., and Atella, V. (2013) Stochastic frontier analysis using Stata. *Stata Journal*. https://doi.org/10.2139/ssrn.2145803

Bonaccorsi, A., and Daraio, C. (2003) 'A Robust Nonparametric Approach to the Analysis of Scientific Productivity', *Research Evaluation*, Vol. 12, No. 1, pp. 47-69. https://doi.org/10.3152/147154403781776726

Bonaccorsi, A., and Daraio C. (2004) 'Econometric Approaches to The Analysis of Productivity of R&D Systems: Production Functions and Production Frontiers', in Moed, H. F., Glänzel, W. and Schmoch, U. (ed.) *Handbook of Quantitative Science and Technology Research*, Amsterdam: Springer.

Bornmann L., Wagner, C., and Leydesdorff, L. (2018) 'The Geography of References in Elite Articles: Which Countries Contribute to the Archives of Knowledge?', *PloS ONE*, Vol. 13, No. 3, pp. e0194805. https://doi.org/10.1371/journal.pone.0194805

Carrillo, M. (2019) 'Measuring and Ranking R&D Performance at the Country Level', *Economics and Sociology*, Vol. 12, No. 1, pp. 100-114. https://doi.org/10.14254/2071-789X.2019/12-1/5

Carvalho, N., Carvalho, L., and Nunes, S. (2015) 'A Methodology to Measure Innovation in European Union Through the National Innovation System', *International Journal of Innovation and Regional Development*, Vol. 6, No. 2, pp. 159-180. https://doi.org/10.1504/IJIRD.2015.069703

Cirillo, V., Martinelli, A., Nuvolari, A., and Tranchero, M. (2019) 'Only One Way to Skin a Cat? Heterogeneity and Equifinality in European National Innovation Systems', *Research Policy*, Vol. 48, No. 4, pp. 905-922. https://doi.org/10.1016/j.respol.2018.10.012

Coccia, M. (2008) 'Measuring Scientific Performance of Public Research Units for Strategic Change', *Journal of Informetrics*, Vol. 2, No. 3, pp. 183-194. https://doi.org/10.1016/j.joi.2008.04.001

Coccia, M., and Rolfo, S. (2007) 'How Research Policy Changes Can Affect the Organization and Productivity of Public Research Institutes: An Analysis Within the Italian National System of Innovation', *Journal of Comparative Policy Analysis: Research and Practice*, Vol. 9, No. 3, pp. 215 - 233. http://dx.doi.org/10.1080/13876980701494624

Eggink, M. (2013). The Components of an Innovation System: A Conceptual Innovation System Framework. *Journal of Innovation and Business Best Practices,* Vol. 2013, 768378. https://doi.org/10.5171/2013.768378

Etzkowitz, H. (2015) 'Rendezvous of the "Third Kind": Triple Helix Origins and Future Possibilities', *Industry and Higher Education*, Vol. 29, No. 4, pp. 243-247. https://doi.org/10.5367/ihe.2015.0267

Etzkowitz, H. (2011) 'The Triple Helix: Science, Technology, and the Entrepreneurial Spirit', *Journal of Knowledge-Based Innovation in China*, Vol. 3, No. 2, pp. 76-90. https://doi.org/10.1108/17561411111138937

Etzkowitz, H and Leydesdorff, L. (2000) 'The Dynamics of Innovation: From National Systems and 'Mode 2' to Triple Helix of University-Industry-Government Relations', *Research Policy*, Vol. 29, No. 2, pp. 109-123. https://doi.org/10.1016/S0048-7333(99)00055-4

Faberger, J. (2017) 'Innovation Policy: Rationales, Lessons, and Challenges', *Journal of Economic Surveys*, Vol. 31, No. 2, pp. 497-512. https://doi.org/10.1111/joes.12164

Fan, P. (2014) 'Innovation in China', *Journal of Economic Surveys*, Vol. 28, No. 4, pp. 725-745 https://doi.org/10.1111/joes.12083

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

Godin, B. (2009). National Innovation System: The System Approach in Historical Perspective. Science, Technology, & Human Values, vol. 34, no. 4, pp. 476-501. https://doi.org/10.1177/0162243908329187

Godin, B. (2017). Models of Innovation. History of an Idea. Cambridge: The MIT Press.

Greene, W. (2005 a) 'Fixed and random effects in stochastic frontier models', *Journal of Productivity Analysis*, Vol. 23, No. 1, pp. 7-32. https://doi.org/10.1007/s11123-004-8545-1

Greene, W. (2005 b) 'Reconsidering heterogeneity in panel data estimators of the stochastic frontier model', *Journal of Econometrics*, Vol. 126, No. 2, pp. 269-303. https://doi.org/10.1016/j.jeconom.2004.05.003

Guan, J., and Chen, K. (2012) 'Modeling The Relative Efficiency of National Innovation Systems', *Research Policy*, Vol. 41, No. 1, pp. 102-115. https://doi.org/10.1016/j.respol.2011.07.001

Heritage Foundation Index of Economic Freedom (2023) *Heritage Foundation Index of Economic Freedom*, [Online], Available: https://www.heritage.org/index/download [11 Jul 2023]

Heller, M. A. and Eisenberg, R. (1998) 'Can Patents Deter Innovation? The Anticommons in Biomedical Research', *Science*, Vol. 280, No. 5364, pp. 698-701. https://doi.org/10.1126/science.280.5364.698

Hu, J-L., Yang, C.-H., and Chen, C.-P. (2014) 'R&D Efficiency and The National Innovation System: An International Comparison Using the Distance Function Approach', *Bulletin of Economic Research*, Vol. 66, No. 1, pp. 3307-3378. https://doi.org/10.1111/j.1467-8586.2011.00417.x

Kotsemir, M. (2013). 'Measuring National Innovation Systems Efficiency - A Review of DEA Approach Basic', *Research Program Working Papers Series: Science, Technology, and Innovation*. https://doi.org/10.2139/ssrn.2304735

Kumbhakar, S. C. (1990) 'Production frontiers, panel data, and time-varying technical inefficiency', *Journal of Econometrics*, Vol. 46, No. 1-2, pp. 201-211. https://doi.org/10.1016/0304-4076(90)90055-X

Kumbhakar, S. C., and Heshmati, A. (1995) 'Efficiency Measurement in Swedish Dairy Farms: An Application of Rotating Panel Data', *American Journal of Agricultural Economics*, Vol. 77, No. 3, pp. 660-674. https://doi.org/10.2307/1243233

Kumbhakar, S. C., Lien, G., and Hardaker, J. B. (2014) 'Technical efficiency in competing panel data models: A study of Norwegian grain farming', *Journal of Productivity Analysis*, Vol. 41, No. 2, pp. 321-337. https://doi.org/10.1007/s11123-012-0303-1

Lacy, P., Long, J., and Spindler, W. (2019) *The Circular Economy Handbook: Realizing the Circular Advantage*. Berlin: Springer Nature.

Lundvall, B-Å. (2005) 'National Innovation Systems - Analytical Concept and Development Tool' *DRUID Tenth Anniversary Summer Conference 2005 on Dynamics of Industry and Innovation: Organizations, Networks and Systems*, Vol. 14, No. 1, pp. 95-119. https://doi.org/10.1080/13662710601130863

Lundvall, B-Å. (2007 a) 'Innovation System Research. Where it came

from and where it might go', *The Global Network for Economics of Learning, Innovation, and Competence Building System (GLOBELICS)*. Oslo: pp. 1-51.

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

Manzini, S. T. (2012) 'The National System of Innovation Concept: An Ontological Review and Critique', *South African Journal of Science*, Vol. 108, No. 9/10. http://dx.doi.org/10.4102/sajs.v108i9/10.1038

Matei, M. M., and Aldea, A. (2012) 'Ranking National Innovation Systems According to Their Technical Efficiency', *Procedia - Social and Behavioral Sciences*, Vol. 62, pp. 968 - 974. https://doi.org/10.1016/j.sbspro.2012.09.165

Nasierowski, W. (2010) 'Technical Efficiency of Efforts to Enhance Innovativeness in the European Union', *International Journal of Innovation and Technology Management*, Vol. 07, No. 04, pp. 389-404. https://doi.org/10.1142/S0219877010002045

Nasierowski W., and Arcelus, F. J. (2003) 'On The Efficiency of National Innovation Systems', *Socio-Economic Planning Sciences*, Vol. 37, No. 3, pp. 215-234. https://doi.org/10.1016/S0038-0121(02)00046-0

OECD (1997). *National Innovation Systems. Organization for Economic Cooperation and Development,* [Online], Available: https://www.oecd.org/mena/47563588.pdf [10 Jul 2023]

Packalen, M. (2019) 'Edge Factors: Scientific Frontier Positions of Nations', *Scientometrics*, Vol. 118, No. 3, pp. 787-808. https://doi.org/10.1007/s11192-018-2991-4

Pan, T.-W., Hung, S.-W., and Lu, W.-M. (2010) 'DEA Performance Measurement of The National Innovation System in Asia and Europe', *Asia-Pacific Journal of Operational Research*, Vol. 27, No. 03, pp. 369-392. https://doi.org/10.1142/S0217595910002752

Pitt, M. M., and Lee, L. F. (1981) 'The measurement and sources of technical inefficiency in the Indonesian weaving industry', *Journal of Development Economics*, Vol. 9, No. 1, pp. 43-64. https://doi.org/10.1016/0304-3878(81)90004-3

Sábato, J. (1975) El pensamiento latinoamericano en la problemática ciencia-tecnología-desarrollo-dependencia. Paidós, Buenos Aires.

Sábato, J., and Mackenzie, M. (1982) La Producción de Tecnología. Autónoma o Transnacional. Nueva Imagen, México.

Scimago Journal & Country Rank (2022) *Country Rankings,* [Online], Available: https://www.scimagojr.com/countryrank.php [11 Jul 2023].

Schot, J., and Steinmueller, W. E. (2018) 'Three Frames for Innovation Policy: R&D, Systems of Innovation and Transformative Change', *Research Policy*, Vol. 47, No. 9, pp. 1554-1567. https://doi.org/10.1016/j.respol.2018.08.011

UNESCO (2023) *Institute for Statistics (UIS),* [Online], Available: http://data.uis.unesco.org/ [11 Jul 2023]

WIPO (2023) *Information Resources on Patents,* [Online], Available: https://www.wipo.int/patents/en/ [11 Jul 2023]

World Bank (2010). *Innovation Policy. A Guide for Developing Countries* [Online], Available: http://hdl.handle.net/10986/2460 [11 Jul 2023]

World Bank Open Data (2023) *World Bank Open Data*, [Online], Available: https://data.worldbank.org/ [11 Jul 2023]

PUPILS' SCHOOL PERFORMANCE AND THEIR COGNITIVE ABILITIES TO SOLVE PROBLEMS

Eva Hejnová✉
Petr Eisenmann
Lucie Loukotová
Jiří Přibyl

Faculty of Science, University of Jan Evangelista Purkyně in Ústí nad Labem, Czech Republic

✉ eva.hejnova@ujep.cz

ABSTRACT

The paper describes the results of a study whose main aim was to find the interrelationship between pupils' school grades in Czech language (native), mathematics, and physics and pupils' cognitive predispositions to problem-solving in science and mathematics diagnosed by the Lawson Classroom Test of Scientific Reasoning and the Culture of Problem Solving test. A total of 180 pupils from the Czech Republic aged 14-15 took part in this study.

The results show that pupils with better grades in the monitored subjects achieve better results in both tests. It also turns out that there are generally statistically insignificant differences between the results of pupils assessed by grades 1 or 2 and between those assessed by grades 3 or 4. Pupils' performance in the two tests might help to strengthen the objectivization of grading at school. They might also help identify the indicators important for developing problem-solving skills. The research specifically points to the need to develop algebraic thinking, the conception of infinity, spatial imagination, geometric imagination in the plane, proportional reasoning, and the ability to control variables.

KEYWORDS

Culture of problem solving, lower secondary school, mathematics education, school performance, scientific reasoning

HOW TO CITE

Hejnová E., Eisenmann P., Loukotová L., Přibyl J. (2024) 'Pupils' School Performance and Their Cognitive Abilities to Solve Problems', *Journal on Efficiency and Responsibility in Education and Science*, vol. 17, no. 1, pp. 54-65. http://dx.doi.org/10.7160/eriesj.2024.170105

Article history
Received
September 13, 2022
Received in revised form
June 25, 2023
Accepted
September 25, 2023
Available on-line
March 31, 2024

Highlights

- *Pupils' classroom assessment shows a connection to their performance in cognitive skills tests.*
- *Pupils with better grades (classification) achieve better results in the test of predispositions to solving mathematical problems and in Lawson's Test of Scientific Reasoning.*
- *There are almost no statistically significant differences in the performance of pupils assessed by grades 1 or 2 and in the performance of pupils assessed by grades 3 or 4.*

INTRODUCTION

The paper is one of the outcomes of the research project concerning mutual relations between two constructs that are related to problem-solving (the Culture of problem solving and Scientific reasoning) and school performance in the Czech language, mathematics, and physics. The construct *Culture of problem solving* (CPS) was introduced by some authors of this paper as a tool for describing a pupil's ability to solve mathematical problems (Eisenmann et al., 2015). The other construct is *Scientific reasoning* (SR), which includes the thinking and reasoning skills involved in systematically exploring a problem, formulating and testing hypotheses, evaluating experimental outcomes, etc. (Bao et al., 2009; Gormally et al., 2012). In (Cihlář et al., 2017), we present the results of a small research conducted among 23 pupils aged 14-15 in the Czech Republic in 2016. The results of this research proved the legitimacy of the idea of exploring mutual relations between individual components of CPS and SR dimensions. This research was then followed up with extensive research in 2017 to describe the dependency between all components of the CPS and the SR dimensions (Cihlář et al., 2020). The objective of the article is to provide a new perspective on the investigation issue based on the data obtained and to answer the question about the relationship between school performance (classification) and the constructs of CPS and SR.

The following sections describe two constructs (CPS and SR). This is discussed in detail in (Hejnová et al., 2018). The following text is a slightly shortened version essential for the purpose of this paper.

The Culture of Problem Solving

The CPS construct was developed to better describe the pupils' dispositions to solve mathematical problems. In creating the CPS, we tried to make this description independent of the problem itself and equally independent of the pupils' knowledge. The development of the construct is well described in (Eisenmann et al., 2015). At this point, we should mention that in its development, we drew mainly from the work of (Schoenfeld, 1982; Sriraman, 2005; Wu and Adams, 2006).

The CPS consists of four components: mathematical intelligence, reading comprehension, creativity, and the ability to use existing knowledge.

In the development of *mathematical intelligence*, we focused on the perception and understanding of six selected mathematical phenomena. We chose Gardner (1993) and Juter and Sriraman (2011) as our inspiration, but our primary goal was not to seek out mathematically gifted pupils. Our focus was to determine the level of disposition of all pupils in the areas that we consider important for successful mathematical problem-solving. We emphasize that we are not concerned with mathematical content per se but only with content to determine the level of the specific phenomenon. These phenomena are logical reasoning, the conception of infinity, spatial imagination, geometric imagination in the plane, algebraic thinking, and arithmetic patterns.

In accordance with the PISA framework, we understand *reading comprehension* as a functional literacy, that is, as a set of knowledge, skills, strategies, and attitudes needed to understand, use, and evaluate all kinds of texts in different contexts. Reading comprehension is one of the key competencies needed to successfully solve problems, especially in the case of word problems (Akbaşl et al., 2016; Pape, 2004; Vilenius-Tuohimaa et al., 2008).

As far as *creativity* is concerned, we focused only on the part of it that relates to divergent thinking (Guilford, 1967). We recognize that convergent thinking also plays a vital role in the task solving process, but it is the production of different ideas that we see as very important in this process, as confirmed by (Kwon et. al., 2006). By divergent thinking, we mean the production of diverse but appropriate answers to an open-ended question or problem. It can be said that the higher the level of creativity, the more difficult the problems can be presented to pupils (Chamberlin and Moon, 2005).

Although some parts of the construct CPS are either investigated from a domain perspective (mathematical intelligence) or are only found in specific school contexts (reading comprehension), creativity per se has been the focus of attention in other areas of human endeavour and is seen as a prerequisite for problem-solving across different domains (Zhou, 2012).

The *ability to use the existing knowledge* is the last component of the CPS. This ability is a necessary condition for the successful solution of non-routine problems. We have developed it to determine the level of formalism in the pupils. Eisenmann et al. (2015) have shown one connection with general intelligence: Those pupils who have a higher indicator of intelligence also show a higher indicator of the ability to use the existing knowledge.

Scientific Reasoning

SR can be regarded as a complex process that is widely defined as "the skills involved in inquiry, experimentation, evidence evaluation, and inference that are done in the service of conceptual change or scientific understanding" (Zimmerman, 2007, p. 172). These general skills, referred to as science process skills (Padilla, 1990), are considered crucial components of STEM education (Science, Technology, Engineering, and Math). As Coletta and Phillips (2005) and Han (2013) have shown, their development is closely related to the cognitive abilities of the pupils and their prior knowledge of the content. In our research, we focus on measuring scientific reasoning skills. That is why we deal with the operational definition of SR. Similarly, to Lawson (1978, 1982, 2004, 2005), we suppose that its structure is determined by the hypothetical-deductive nature of science and includes dimensions such as proportional thinking, identification, and control of variables, probabilistic thinking, correlational thinking and inductive and deductive reasoning (Han, 2013). However, other skills could be involved because of the multidimensional structure of SR (Opitz et al., 2017).

Classroom Assessment in Czech Schools

We understand classroom assessment to be the process of measuring performance and gaining evaluation information in lessons. This information is primarily intended for pupils. This assessment relates to the specific learning objectives and is an integral part of the teaching (different types of assessment of pupils' knowledge and skills, oral and written testing, peer assessment, training of pupils in self-assessment) or directly supports and complements it (on-the-fly diagnostic information for parents about pupils' school performance).

Assessment is regarded as the key tool for the improvement of school performance (see, e.g., Black et al., 2003, 2004; Naylor et al., 2005). Together with the process of enriching and updating the objectives of education, also the educational content that has to be assessed is modified. The quality and efficiency of the educational process are significantly influenced by the way they are evaluated. That is why teachers from developed countries ask for assessments that will be well thought out with respect to the effects they will have on school practice. Concerns that assessment of only some (well-measurable) educational objectives will result in deformation of teaching brought about the need to verify educational goals that had not been previously assessed (critical thinking, problem-solving, social skills, etc.). This puts new demands on assessment methods and used tools (Chvál et al., 2015).

The current conception of mathematics and science education that should be acquired by all students at school emphasizes, in particular, the development of a general understanding of important concepts, understanding of methods by which science gets evidence to support its claims, understanding of strengths of science and its limitations in the real world; the ability to draw correct and well-founded conclusions from presented facts and information, to critically assess people's statements on the basis of the evidence presented and to distinguish opinions from evidence-based claims.

Tests used by teachers in the Czech Republic often target lower-

level skills and knowledge of facts. However, international research and surveys (e.g., PISA[1]) show a clear tendency to set to students the so-called context tasks that refer to some situation and show the applicability of mathematics and science in real-life situations or tasks that test such knowledge and skills of students that they will need in their future lives. Tasks also often focus on overall student abilities that permeate all subjects (not only mathematics and science), on the use of their own way of thinking and understanding in specific life situations.

In our research, we focus primarily on summative assessment, the purpose of which is to obtain an overall overview of students' performance. The aim of this assessment is to diagnose a student and evaluate their performance with respect to the assessed group of students.

Classification (assessment by grades) is still predominantly used in Czech schools as the standard for summative assessment of students. Its advantages are simplicity and systematicity when used in practice, as well as its long tradition and comprehensibility to parents and the public. In Czech schools, students are evaluated by grades 1 to 5. Grade 1 corresponds to the best performance, and 5 describes insufficient performance and failure.

Research on classification shows a relative stability of achievement during school attendance, which is attributed to a relatively stable and evenly developing dispositional basis, which determines a student's school performance (Hrabal, 1989). The existence of a grade has a constant influence on the character, intensity, and focus of a student's learning activities, affects the classroom atmosphere, and also represents simple informative feedback for the student and their parents (Slavík, 1999).

One of the tools for assessing students in science and mathematics is problem-solving. Suurtamm et al. (2016) highlight the relationship between students' problem-solving success and their grades. Tasks on whose basis students should be assessed, according to (Swan and Burkhardt, 2012), should present a balanced view of the curriculum in terms of all aspects of performance that the curriculum wants to encourage. Each student has specific cognitive predispositions to problem-solving. Students with better cognitive predispositions can be expected to have better grades in mathematics and physics. As expected, it was confirmed (Česká školní inspekce (ČŠI), 2019), for example, that students who had a better grade in mathematics on the school final report from the previous school year achieved higher average success in the mathematical literacy test, with a slightly more pronounced effect in case of ninth-graders than in case of 2nd-year students of upper secondary schools.

Objectives

In this paper, we focus on the relationship between classroom assessment (classification) and cognitive predispositions of lower secondary 14 to 15-year-old students to solve problems in the area of mathematics and science. Their predispositions were diagnosed using the Lawson Classroom Test of Scientific Reasoning (LT) and the CPS test. With respect to the focus of these two tests, we were working with grades in three major subjects - Czech language, mathematics, and physics as we believe these play a significant role in the development of the ability to solve problems.

We asked the following research question: To what extent are the variables determined by the CPS and LT tests related to grades in Czech, mathematics, and physics?

MATERIALS AND METHODS

Culture of Problem Solving

The test to determine the values of all four CPS components was designed to be administered in one teaching lesson and took 45 minutes. The individual components were timed as follows: mathematical intelligence - 13 minutes, reading comprehension - 13 minutes, creativity - 9 minutes, and the ability to use the existing knowledge - 9 minutes. Each student worked independently during the assessment and could only use a simple calculator in addition to writing tools. All parts of the test were evaluated by the authors of this paper.

The test of *mathematical intelligence* consisted of 8 problems determining sensitivity to the above-mentioned six specific phenomena: logical reasoning, a conception of infinity, spatial imagination, geometric imagination in a plane, algebraic thinking, and arithmetic patterns. All test problems, with one exception, were closed multiple-choice tasks with one correct answer. The tested student could get 2 points for each problem. The sum of all points formed the total score.

The *reading comprehension* test was created on the same principle as the tests in the PISA research. The test proceeded as follows: the students were presented with a text of 15 lines about kangaroos. The students then had to answer 4 closed and 2 open questions. The text was available to them all the time.

Since we chose to measure *creativity* by divergent thinking, we chose Guilford's Alternative Uses Test as our instrument. This test is based on the fact that the student is presented with a word that expresses a certain object, and the student is asked to suggest as many different and unusual uses of this object as possible, counting only feasible uses. In the test, four words were presented to the students to rule out a certain inclination towards a subject and to observe phenomena such as fluency, originality, and elaborateness. The qualitative assessment of each part of the test was then converted into scores, and the total score indicated the creativity index.

The test of *the ability to use the existing knowledge* consisted of four simple, non-routine problems. At the beginning of each problem, the students revised the knowledge that was necessary to solve the task (How to find the part from the whole with percentage, the circumference of a circle, the surface of a cuboid, and the lowest common multiple). For each task, the student could get one point, and the total sum informs the overall score in this test.

Scientific Reasoning

SR was tested by the Czech version of LT including 24 items (Dvořáková, 2011) with small corrections in items 8a and 8b according to (Han, 2013), allowing one to examine six

1 Programme for International Student Assessment

dimensions: conservation of matter and volume (items 1 to 4), proportional reasoning (items 5 to 8), control of variables (items 9 to 14), probability reasoning (items 15 to 18), correlation reasoning (items 19, 20), and hypothetical-deductive reasoning (items 21 to 24) (Lawson, 1978).

LT is a two-tier multiple-choice test with items of increasing difficulty, each of the two-tier items including a question offering answers and possible reasons for the response to the question. A student could get two points for questions 1 through 22 if he chose the correct answer and concurrently its correct justification. Only the answers to questions 23 and 24 were evaluated separately; that is, the student received one point for each correctly answered question or its correct justification. The students had a maximum of 45 minutes to solve the test.

The results of the test allow us to determine what level of scientific reasoning a particular student has achieved (Dvořáková, 2011). The first developmental level is concrete-level reasoning. This stage consists of students who get 0-8 points in the LT test. The second level is transitional and is made up of students who get 9-16 points. The highest level is formal-level reasoning. This stage is achieved by students who get 17-24 points.

Sample

A total of 180 students (76 girls and 104 boys) aged 14-15 from eight classes from six lower secondary schools and one class of an eight-year secondary grammar school took part in our study in May and June 2017. All schools were located in three towns in the Ústí nad Labem Region in the Czech Republic. None of the classes were specialized and integrated for pupils with physical or mental disabilities or low socioeconomic status. All the pupils were native Czech speakers.

Statistical Evaluation

As part of the statistical processing, the pupils' grades and the results of the CPS and LT tests were first analyzed with respect to the pupils' scores.

Also, the relationships between grades in the Czech language (CZECH), mathematics (MATH), and physics (PHYS) and the overall scores in the CPS and LT tests were examined. The Kruskal-Wallis test was used to evaluate the dependency between grades and overall scores in these tests, and partial differences among different classes of grades were studied using Wilcoxon tests. The effect size coefficient ξ^2 was used to determine the strength of the association. The guidelines of Field (2013) were followed to interpret the size of the effect. For the purpose of evaluation, the category of pupils with grade 5 was merged with the category of pupils with grade 4. This modification was necessary because of the low number of pupils with grade 5 in the sample. In the case of the CPS test, the results in individual subtests were normed in such a way that all subtests had the same weight.

The calculation was realized by STATISTICA 12.0 (StatSoft, Inc.). The level of significance $\alpha = 0.05$ was used in all tests.

RESULTS

The pupils' school performance was assessed according to their grades in Czech language, mathematics, and physics on the final school report in the school year 2016/2017. The structure of grades in the monitored subjects is given in Figure 1. The structure of grades in Czech language and mathematics is very similar; the proportion of grades 1 to 4 is roughly the same in both subjects. About 60% of pupils have a grade 1 or a grade 2. In contrast, a greater proportion of grades 1 can be observed in physics (about 30% of pupils in contrast to 20% in Czech language and mathematics) and a lower proportion of grades 4. The proportion of grades 3 is about the same in all three subjects (slightly more than 20%). Grades 5 only appeared in mathematics. Average grades in individual subjects correspond to the above. The average grade in Czech language and mathematics was 2.30, and in physics, 2.06.

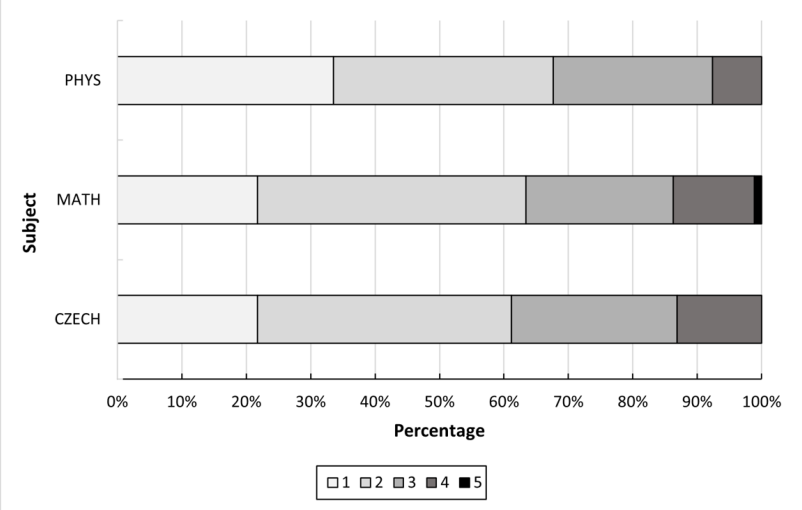

Figure 1: Structure of grades, 2017 (source: own calculation)

As far as the LT test is concerned, the average score was 7.9 points (out of 24 points) with a standard deviation of 5.3 points. Figure 2 illustrates that the distribution of the achieved score does not show a maximum in the proximity of the average gained score. On the contrary, the pupils' scores are scattered on the whole scale of possible scores.

Printed ISSN 2336-2375 Electronic ISSN 1803-1617
ERIES Journal volume 17 issue 1

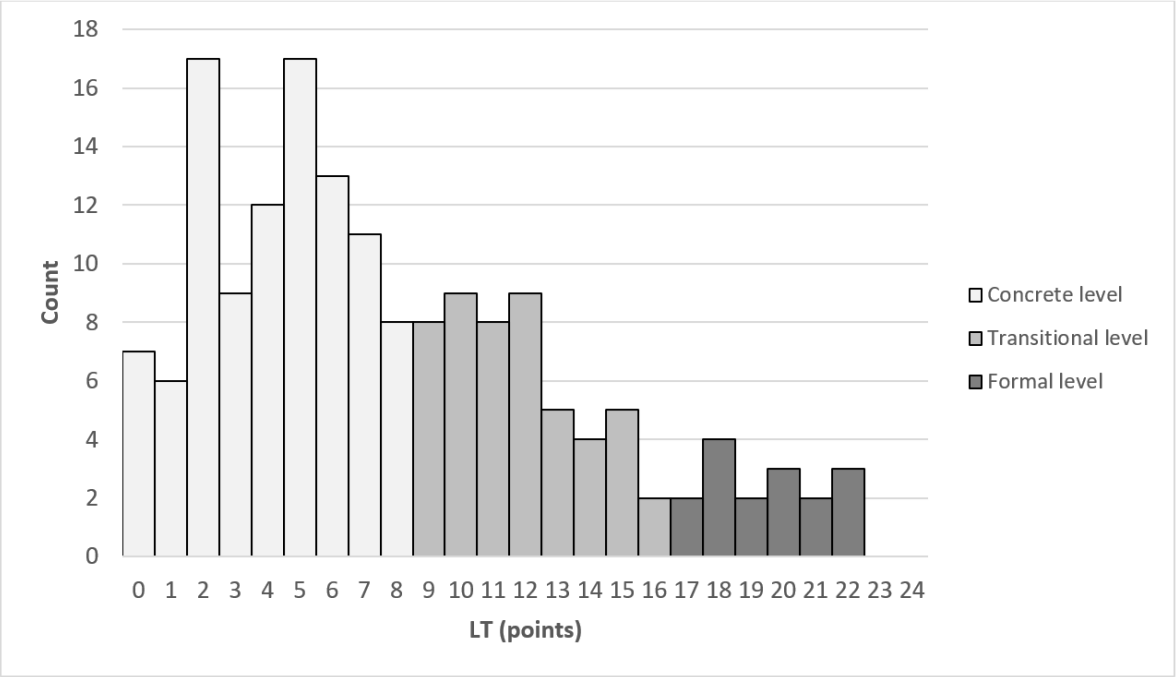

Figure 2: Points from Lawson test (LT), 2017, (source: own calculation)

Pupils most often had two to seven points in the LT test, namely 48% of them. 45% of the pupils had an above-average score, that is, 8 points and more, and 19% of the pupils gained more than half of the possible points (i.e., more than 12). Roughly 8% of the pupils had only one or no point. None of the pupils achieved the maximum number of points. 60% of the pupils are on the lowest, which is concrete-level reasoning. 30% of the pupils are on the transitional level of scientific reasoning. Only 10% of the pupils have reached the highest, i.e., formal-level reasoning.

The maximum number of points was 24, also in the CPS test. The average score achieved by pupils was 12.1, with a standard deviation of 4.4 points. The score on the CPS test is distributed normally (see Figure 3).

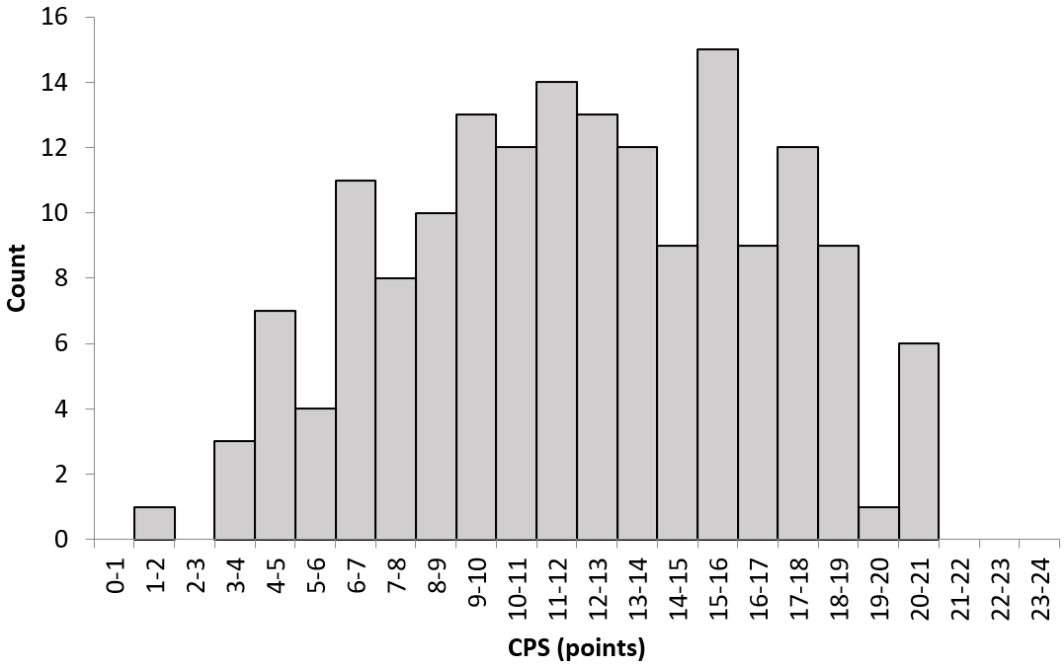

Figure 3: Points from CPS test, 2017, (source: own calculation)

The coefficient of variation in LT is 67.9%, which means that the scores of individual pupils have a moderate range. In contrast, the coefficient of variation in the CPS test is only 36.6%, which means that the scores of the CPS test are much more consistent.

Kruskal-Wallis tests show a significant association between the overall score on the CPS test and grades in Czech language, mathematics, and physics; the effect of the grade on CPS and LT tests is large. The results of the tests are presented in Table 1.

Pairs of variables	Kruskal-Wallis test		Effect size	
	H	*p*	ξ^2	size of effect
CPS & CZECH	68.87	< .001	.409	large
CPS & MATH	44.99	< .001	.261	large
CPS & PHYS	50.42	< .001	.304	large

Table 1: Relationship between the overall score in the CPS tests and grades in Czech, mathematics and physics, 2017 (source: own calculation)

The results of Kruskal-Wallis tests in the case of the search for relations between the score in LT and grades in Czech, mathematics, and physics are very similar. The results of the tests are presented in Table 2.

Pairs of variables	Kruskal-Wallis test		Effect size	
	H	*p*	ξ^2	size of effect
LT & CZECH	68.87	< .001	.208	large
LT & MATH	44.99	< .001	.189	large
LT & PHYS	50.42	< .001	.171	large

Table 2: Relationship between overall score in the LT test and grades in Czech, mathematics and physics, 2017 (source: own calculation)

Pupils with better grades generally achieve better scores both in the CPS and LT tests. The grade has a larger effect on the CPS score than on the LT score.

A more detailed look at the relationship between the grades in the selected subjects and the result in the tests is very interesting. Wilcoxon subtests show that in the case of grades in Czech and physics, the scores in CPS and LT tests are not significantly different for pupils with grades 1 or 2 (Czech: $p = .10$ for CPS, $p = .06$ for LT, physics: $p = 1.00$ for both tests), and for pupils with grades 3 or 4 ($p = 1.00$ for both of subjects and tests). This allows us to define, for the case of grades in Czech and physics, a group of "more successful pupils" (with grades 1 or 2) and of "less successful pupils" (with grades 3 or 4).

Basic descriptive statistics for the results of CPS and LT tests in relation to grades in the Czech language are shown in Figure 4 and Figure 5.

Similarly, Figures 6 and 7 present the basic descriptive statistics for the results of CPS and LT tests in relation to the grades in physics.

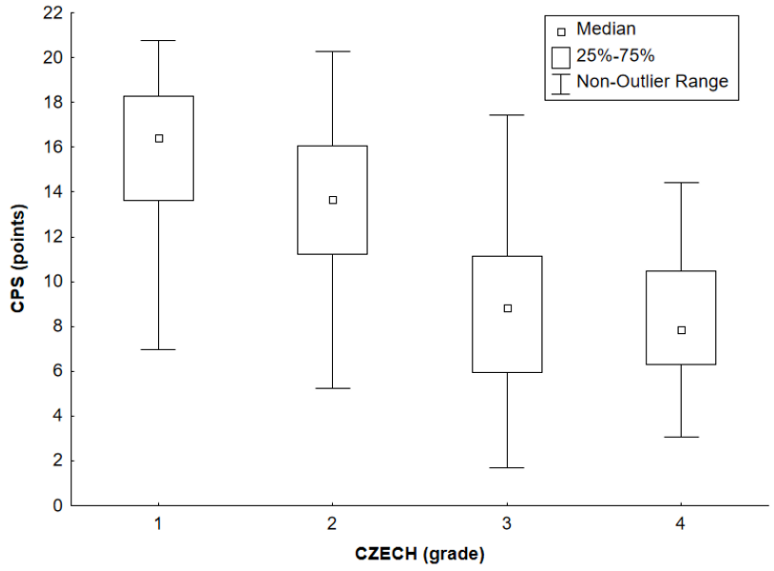

Figure 4: Results of CPS test with respect to grades in Czech, 2017 (source: own calculation)

Printed ISSN 2336-2375 Electronic ISSN 1803-1617
ERIES Journal volume 17 issue 1

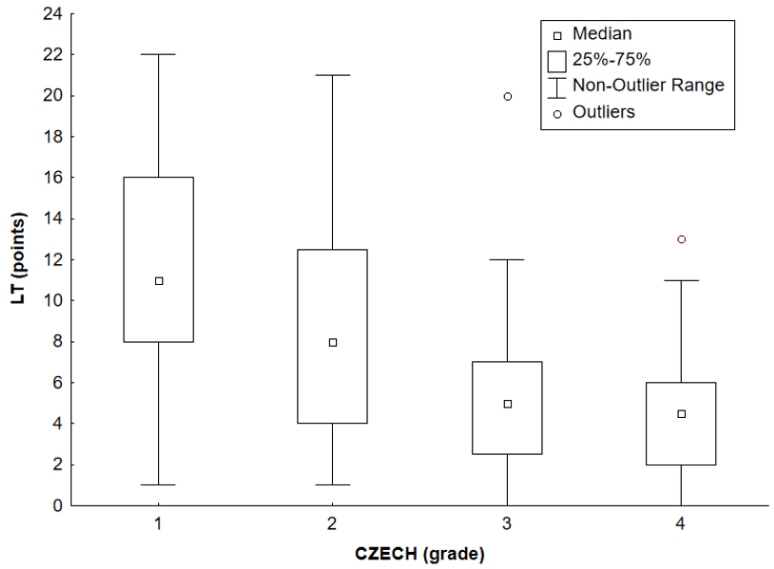

Figure 5: Results of LT with respect to grades in Czech, 2017 (source: own calculation)

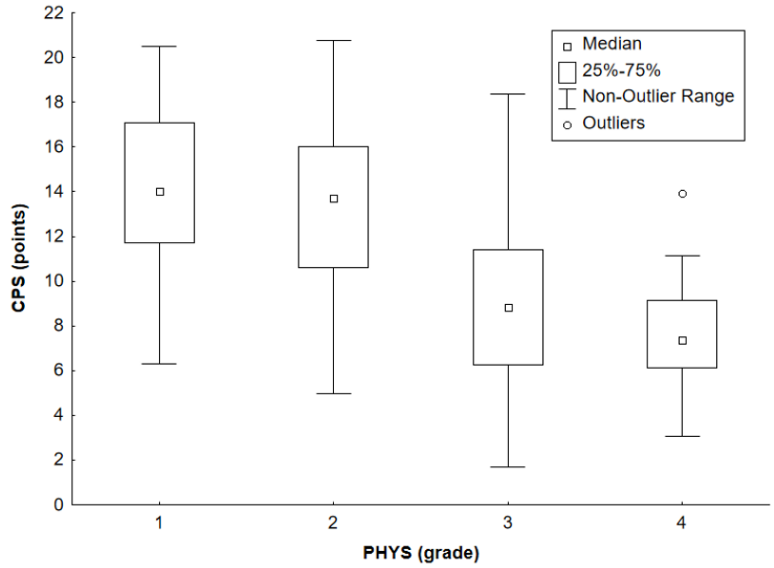

Figure 6: Results of CPS test with respect to grades in physics, 2017 (source: own calculation)

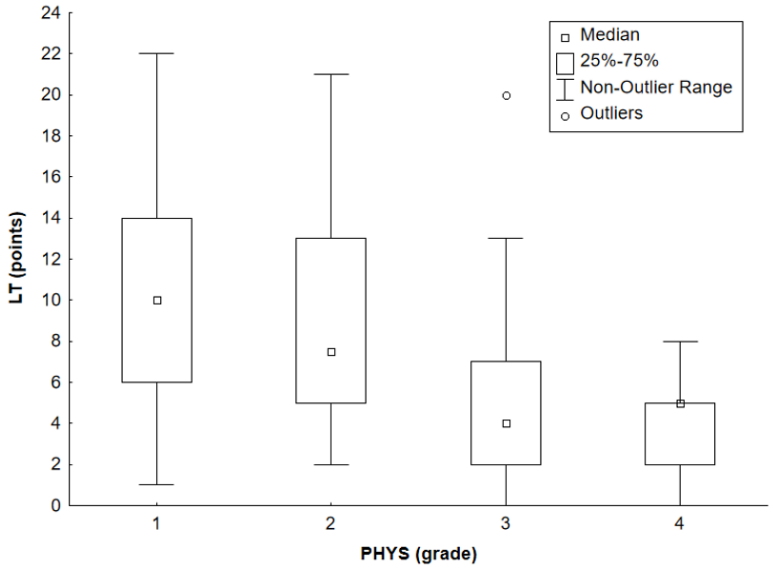

Figure 7: Results of LT with respect to grades in physics, 2017 (source: own calculation)

In the case of grades in mathematics, there are no statistically significant differences between pupils with grades 3 or 4 in CPS and LT tests ($p = 1.00$ for both tests; see Figures 8 and 9 for basic descriptive statistics). However, unlike in the case of Czech and physics, pupils with grades 1 in mathematics are significantly better than all other groups both in CPS and LT tests ($p < .02$ in all cases). Moreover, in the case of LT, pupils with grades 2, 3, or 4 show similar results (the result of pupils with grades 2 or 3 is borderline).

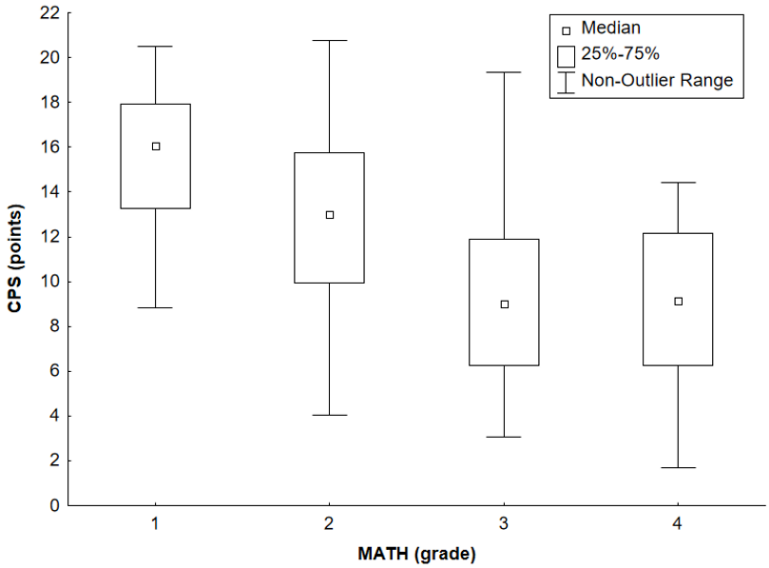

Figure 8: Results of CPS test with respect to grades in mathematics, 2017 (source: own calculation)

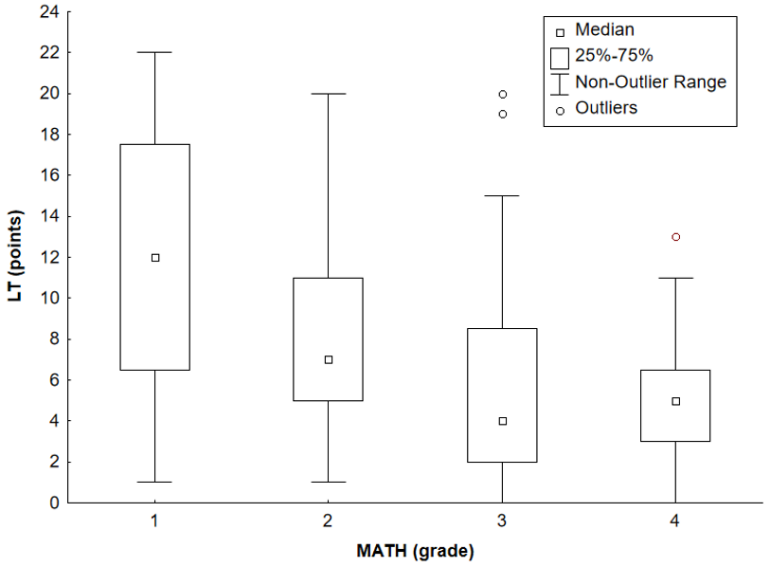

Figure 9: Results of LT tests with respect to grades of mathematics, 2017 (source: own calculation)

DISCUSSION

The study was designed to generate data relevant to the question of the relationship between the variables determined by the CPS and the LT tests and the grades in Czech, mathematics, and physics among a cohort of Czech pupils aged 14-15. All statistical tests showed a significant association between the overall score of the CPS and LT tests and the grades where the association of the score of the CPS test with the grades was found to be stronger than in the case of LT.

The dependency between variables measured by the CPS and LT tests and school performance given by pupils' grades is not much described in the literature. Valanides (1997) examined the relationship between performance on the Test of Logical Thinking (TOLT) and gender, a section of the study, and measures of school achievement of 12th-grade students. He concluded that gender, section of study, achievement in mathematics, and grade point average, but not achievement in science and Greek language, contributed significantly to predicting performance on TOLT. Generally, a pupil with a high IQ is expected to have good results at school (see, e.g., Jalili et al., 2018), and a learner with higher levels of SR could be expected to be a superior problem solver (Tajudin and Chinnappan, 2015). Research also has shown (e.g., Hand et al., 2001; Samková et al., 2021) that reasoning skills represent a set of broadly transferable skills across science and mathematics, and teaching scientific reasoning has a lasting

impact on general learning ability not only in science but also in mathematics. Hilbert et al. (2019) exposed in their research with Austrian students that the grade in mathematics was best predicted by reasoning at the age of 11-12. The reasoning was assessed using the standard progressive matrices (Raven et al., 2000). The test requires participants to complete matrices based on visual patterns. Such tasks were used by us in the test of mathematical intelligence (one component of the CPS structure) as well.

As stated in the chapter Results, there are almost no statistically significant differences in the performance in the CPS and LT test of pupils assessed by grades 1 or 2 in Czech and physics and in the performance of pupils assessed by grades 3 or 4 in these subjects. We can talk about a group of "more successful pupils" (with grades 1 and 2) and of "less successful pupils" (with grades 3 and 4). A similar grouping of pupils can be found in Tajudin and Chinnappan's (2015) study, where groups are referred to as high-achievement and low-achievement groups. Also, in PISA research (Potužníková et al., 2019), all pupils are divided into two basic groups: pupils with critical reading skills (they do not reach basic level 2 according to PISA) and others.

Teachers who diagnose their pupils using the CPS and LT tests can use these results to divide pupils into two groups and thus better account for pupil heterogeneity in instruction. Dividing pupils and then treating the two groups individually during a lesson can be a powerful tool that allows the teacher to better take into account the different cognitive backgrounds of the pupils in the class and to work with each group in a slightly different way. Pupils who are likely to have difficulties in understanding the material can be given more attention and given, for example, appropriate preparatory tasks or exercises to practice the material. Pupils who have been more successful on the CPS and LT tests can work more independently in the lessons or receive more challenging tasks. For both groups, this approach can be effective in increasing their motivation to learn.

However, this division into two groups may not always be 100% effective and may have limitations, for example, due to differences in pupils' individual abilities or their social interactions in groups. It is also important to note that the CPS and LT tests alone may not fully reveal all aspects of the abilities and learning needs of the pupils. In addition, some pupils may not perceive the division into two groups positively and may feel, for example, a certain sense of injustice or stigma. If a teacher decides to create non-heterogeneous groups for the above reasons, for example, to implement group learning, this can effectively strengthen pupils' social bonds and social skills. Non-heterogeneous groups can also help create an environment in which weaker pupils feel accepted and included regardless of their individual abilities or skills, which can increase their self-esteem and motivation to learn. However, each teacher must always individually weigh the advantages and disadvantages of both approaches before deciding on a particular teaching method.

If teachers want to eliminate a pupil's reading comprehension deficit, they can be inspired by the so-called Singapore mathematics based on the triad: concrete manipulation - image representation - abstract model (Kaur and Yeap, 2009; Wong and Lee, 2009). The image representation of the initial situation is a key tool that Singaporean students consciously learn to deal with. Drawing an image converts verbal information into a visual one. This process helps the student to realize the relationships between the individual pieces of information in the task assignment. The structure of the task parameters is analyzed and converted into a visual representation. This principle is also in accord with our research (Eisenmann et al., 2015) or, for example, with (Nunes and Bryant, 2015). Similarly, if we want to compensate for a pupil's deficiency in mathematical intelligence, we can successfully present him with mathematical tasks of different types. We must now note that each of the six named phenomena is differently sensitive to stimuli. For example, the perception of infinity develops very slowly and matures only with the individual. In contrast, arithmetic patterns and geometric imagination in a plane are phenomena that develop quite well (Alsina and Nelsen, 2006; Rezaie and Gooya, 2011). In the case of logical reasoning, it is sometimes difficult to find out what is causing the problems. In fact, the cause may be a low level of reading literacy. In a situation where we have ruled out this cause, we present the problems to the students and go through the problem statement with them to see if they are able to identify the relationships between objects. For these tasks, it is recommended that they also plot these relationships. Pólya and Conway (2004) and Boaler and Dweck (2016) clearly recommend this approach.

In the present study, we found out that a large part of the pupils (60%) had acquired only concrete-level reasoning, a third of the pupils were in the transitional (30%), and a smaller part of the pupils were in formal (10%) reasoning levels. However, in the 9th grade, most pupils should optimally be on the transitional level of reasoning (Han, 2013). Concrete-level reasoning refers to thinking patterns that allow pupils to grasp concepts and statements that directly refer to well-known actions. At this level, pupils can follow instructions step by step, provided each step is fully specified. On the transitional level of reasoning, pupils also remain limited to being only capable of partial formal reasoning. Given the positive association between the grades of the three subjects and the scores on the LT test, the "more successful pupils" are expected to achieve good results on the LT test. However, the scores on the LT test were satisfactory for only 40% of the students who achieved transitional or formal reasoning levels, which does not correlate with the number of pupils with grades 1 or 2, who represented approximately 60% of all study participants. A possible explanation for this is that the mastery of the individual dimensions of the SR is not reflected in the grades to a sufficient extent.

One of the strategic lines of the Strategy for the Education Policy of the Czech Republic up to 2030+ is to move away from a broad body of expected knowledge and to foster the ability to understand issues in a deeper context (Fryč et al., 2020). Assessment practice should, therefore, mirror the curriculum we want to develop: its goals, objectives, content, and instructional approaches. The Framework Education Programme for Basic Education (Ministerstvo školství, mládeže a tělovýchovy,

2017), which is a basic curricular document in the Czech Republic, emphasizes the development of key competencies, which also includes problem-solving competencies. Evaluation of such general competencies requires a much broader and more holistic view of student performance, e.g., also in mathematics and science. In contrast to knowledge assessment, however, assessment of progress toward competencies is more difficult, and teachers need to receive useful support in this regard.

At the end of this chapter, we would like to mention that in our study, we tested only pupils from the Ústí nad Labem Region due to the practical feasibility of the research. However, we believe that this fact does not have a significant impact on the generality of our conclusions, as schools were selected in such a way as to minimize possible bias caused by regional limitations, and the sample of pupils was sufficiently representative of the entire population in the age category 14-15 years. Thus, although our research was limited to one region, we believe that our findings may be relevant to the education of pupils in other regions of the Czech Republic and bring new insights into the overall context of educational research.

If we could generalize our research question, we could also discuss whether the three subjects we have chosen (Czech language, mathematics, and physics) really play a significant role in the development of students' ability to solve problems.

We can say that a certain unifying element of both CPS and SR is critical thinking. Indeed, this mode of reasoning is largely present in the background of both constructs (Dowd et al., 2018; Syafril et al., 2020) but the given triad of CPS, SR and critical thinking has not yet been examined together. Research that addresses the constructs of CPS, SR, and critical thinking can be perceived as a possible challenge for the future.

CONCLUSION

Our study proved that pupils with better classroom assessment achieve better results both in CPS and LT tests. This conclusion refers to all three monitored subjects. At the same time, it has been shown that there were no statistically significant differences between the results of pupils with grades 1 or 2, and also between the results of pupils with grades 3 or 4. This finding seems to point to the fact that when assessed by grades, pupils are evaluated not only for their cognitive skills but also for other competences and for implementation of other learning objectives that are not targeted by the CPS and LT tests. However, we believe that the results of pupils in tests such as CPS and LT can help to reinforce the objectification of grades within the summative assessment of pupils. Evaluation by grades still has a significant impact on the degree of effort in learning. Thus, it makes sense to develop those skills in pupils who are involved in problem-solving.

These skills correspond to the variables that were subject to testing using the CPS test. In particular, we mean reading comprehension and some variable components of mathematical intelligence that can be developed: algebraic thinking, the conception of infinity, spatial imagination, and geometric imagination in the plane.

In the case of the LT test, it turned out that nearly two-thirds of the pupils only reached concrete-level reasoning. Pupils should be able to solve problems on the conservation of matter and volume already at the end of the primary level. This should be followed by practicing problems on proportional reasoning at both primary and lower secondary school levels. With respect to the development of scientific reasoning, we believe that the development of the ability to control variables is essential at the lower secondary school level. A thorough acquisition of the above-mentioned skills is a prerequisite to the development of other skills, such as probability reasoning, correlation, and hypothetical-deductive reasoning.

Pupils who perform better on the CPS and Lawson tests are more likely to develop problem-solving competences: they are better equipped to perceive a variety of problem situations both in and out of school, to recognize and understand the problem, and to think through and plan how to solve it. Additionally, they may be more effective in searching for information that is suitable for solving a problem and identifying its commonalities and differences. These pupils are then also more likely to succeed in entrance examinations to schools where science and engineering are taught.

REFERENCES

Akbaşl, S., Şahin, M. and Yaykiran, Z. (2016) 'The Effect of Reading Comprehension on the Performance in Science and Mathematics', *Journal of Education and Practice*, Vol. 7, No. 16, pp. 108-121.

Alsina, C. and Nelsen, R. B. (2006) *Math made visual: creating images for understanding mathematics*, Washington DC: Mathematical Association of America.

Bao, L., Cai, T., Koenig, K., Fang, K., Han, J., Wang, J., Liu, Q., Ding, L., Cui, L., Luo, Y, Wang, Y., Li, L. and Wu, N. (2009) 'Learning and Scientific Reasoning', *Science*, Vol. 323, No. 5914, pp. 586-587. https://doi.org/10.1126/science.1167740

Black, P., Harrison, C., Lee, C., Marshall, B. and Wiliam, D. (2003) *Assessment for learning: putting it into practice*, Maidenhead: Open Univ. Press.

Black, P., Harrison, C., Lee, C., Marshall, B. and Wiliam, D. (2004) 'Working inside the Black Box: Assessment for Learning in the Classroom', *Phi Delta Kappan*, Vol. 86, No. 1, pp. 8-21. https://doi.org/10.1177/003172170408600105

Boaler, J. and Dweck, C. (2016) *Mathematical mindsets: unleashing students' potential through creative math, inspiring messages, and innovative teaching*, San Francisco: Wiley

Česká školní inspekce (ČŠI) (2019) *Rozvoj matematické gramotnosti na základních a středních školách ve školním roce 2017/2018 [Development of mathematical literacy in primary and secondary schools in the school year 2017/2018]*, Praha: Česká školní inspekce.

Chamberlin, S. A. and Moon, S. M. (2005) 'Model-Eliciting Activities as a Tool to Develop and Identify Creatively Gifted Mathematicians', *Journal of Secondary Gifted Education*, Vol. 17, No. 1, pp. 37-47. https://doi.org/10.4219/jsge-2005-393

Chvál, M., Procházková, I. and Straková, J. (2015) *Hodnocení výsledků vzdělávání didaktickými testy [Evaluation of learning outcomes by didactic tests]*, Praha: Česká školní inspekce.

Cihlář, J., Eisenmann, P., Hejnová, E., and Přibyl, J. (2017). 'Science reasoning and culture of problem solving', *Proceedings of the 14th International Conference (ERIE 2017)*, Prague, pp. 33-40.

Cihlář, J., Eisenmann, P., Hejnová, E., and Přibyl, J. (2020). 'Problem Solving in Mathematics and Scientific Reasoning', *The New Educational Review*, Vol. 41, No. 3, pp. 97-108. https://doi.org/10.15804/tner.20.61.3.08

Hejnová, E., Eisenmann, P., Cihlář, J. and Přibyl, J. (2018) 'Relations between Scientific Reasoning, Culture of Problem Solving and Pupil's School Performance', *Journal on Efficiency and Responsibility in Education and Science*, Vol. 11, No. 2, pp. 38-44. https://doi.org/10.7160/eriesj.2018.110203

Coletta, V. P. and Phillips, J. A. (2005) 'Interpreting FCI scores: Normalized gain, preinstruction scores, and scientific reasoning ability', *American Journal of Physics*, Vol. 73, No. 12, pp. 1172-1182. https://doi.org/10.1119/1.2117109

Dowd, J. E., Thompson, R. J., Schiff, L. A. and Reynolds, J. A. (2018) 'Understanding the Complex Relationship between Critical Thinking and Science Reasoning among Undergraduate Thesis Writers', *CBE Life Sciences Education*, Vol. 17, No. 1, pp. ar4. https://doi.org/10.1187/cbe.17-03-0052

Dvořáková, I. (2011) 'Scientific reasoning of pupils - How to develop and test it' [Online], Available: https://kdf.mff.cuni.cz/lide/dvorakova/Plzen_prispevek_Dvorakova.pdf [17 Jun 2023]

Eisenmann, P., Novotná, J., Přibyl, J. and Břehovský, J. (2015) 'The development of a culture of problem solving with secondary students through heuristic strategies', *Mathematics Education Research Journal*, Vol. 27, No. 4, pp. 535-562. http://doi.org/10.1007/s13394-015-0150-2

Field, A. (2013) *Discovering Statistics Using IBM SPSS Statistics*, Sage: London.

Fryč, J., Matušková, Z., Katzová, P., Kovář, K., Beran, J., Valachová, I., Seifert, L., Běťáková, M. and Hrdlička, F. (2020) '*Strategy for the Education Policy of the Czech Republic up to 2030*', Praha: Ministerstvo školství, mládeže a tělovýchovy. [Online], Available: https://www.msmt.cz/uploads/Brozura_S2030_online_CZ.pdf [20 Jun 2023]

Gardner, H. (1993) *Frames of mind: the theory of multiple intelligences*, New York: BasicBooks.

Gormally, C., Brickman, P. and Lutz, M. (2012) 'Developing a Test of Scientific Literacy Skills (TOSLS): Measuring Undergraduates' Evaluation of Scientific Information and Arguments', *CBE—Life Sciences Education*, Vol. 11, No. 4, pp. 364-377. https://doi.org/10.1187/cbe.12-03-0026

Guilford, J. P. (1967) *The nature of human intelligence*, New York: McGraw-Hill.

Han, J. (2013) *Scientific Reasoning: Research, Development, and Assessment,* [Ph.D. thesis], Ohie: The Ohio State University.

Hand, B. M., Prain, V. and Yore, L. (2001) 'Sequential Writing Tasks' Influence on Science Learning', in Tynjälä, P., Mason, L. and Lonka, K. (ed.) *Writing as a Learning Tool*, Dordrecht: Springer. https://doi.org/10.1007/978-94-010-0740-5_7

Hilbert, S., Bruckmaier, G., Binder, K., Krauss, S. and Bühner, M. (2019) 'Prediction of elementary mathematics grades by cognitive abilities', *European Journal of Psychology of Education*, Vol. 34, No. 3, pp. 665-683. https://doi.org/10.1007/s10212-018-0394-9

Hrabal, V. (1989) *Pedagogicko-psychologická diagnostika žáka [Pedagogical and psychological diagnosis of the pupil]*, Praha: Státní pedagogické nakladatelství.

Jalili, A., Hejazi, M., Fomani, G. E. and Morovvati, Z. (2018) 'The Correlation between IQ with Educational Performance and Problem Solving Mediation', *Journal of Health Promotion Managment*, Vol. 7, No. 1, pp. 1-8. https://doi.org/10.21859/jhpm-08011

Juter, K. and Sriraman, B. (2011) 'Does High Achieving in Mathematics = Gifted and/or Creative in Mathematics?', in Sriraman, B. and Lee, K. H. (ed.) *The Elements of Creativity and Giftedness in Mathematics*, Rotterdam: SensePublishers. https://doi.org/10.1007/978-94-6091-439-3_4

Kaur, B. and Yeap, B. H. (2009) *Pathways to reasoning and communication in the primary school mathematics classroom: A resource for teachers by teachers*, Singapore: Centre for Research in Pedagogy and Practice, National Institute of Education.

Kwon, O. N., Park, J. H. and Park, J. S. (2006) 'Cultivating divergent thinking in mathematics through an open-ended approach', *Asia Pacific Education Review*, Vol. 7, No. 1, pp. 51-61. https://doi.org/10.1007/BF03036784

Lawson, A. E. (1978) 'The development and validation of a classroom test of formal reasoning', *Journal of Research in Science Teaching*, Vol. 15, No. 1, pp. 11-24. https://doi.org/10.1002/tea.3660150103

Lawson, A. E. (1982) 'The nature of advanced reasoning and science instruction', *Journal of Research in Science Teaching*, Vol. 19, No. 9, pp. 743-760. https://doi.org/10.1002/tea.3660190904

Lawson, A. E. (2004) 'The Nature and Development of Scientific Reasoning: A Synthetic View', *International Journal of Science and Mathematics Education*, Vol. 2, No. 3, pp. 307-338. https://doi.org/10.1007/s10763-004-3224-2

Lawson, A. E. (2005) 'What is the role of induction and deduction in reasoning and scientific inquiry?', *Journal of Research in Science Teaching*, Vol. 42, No. 6, pp. 716-740. https://doi.org/10.1002/tea.20067

Ministerstvo školství, mládeže a tělovýchovy (2017) 'Rámcový vzdělávací program pro základní vzdělávání [Framework Education Programme for Basic Education]', [Online], Available: https://www.msmt.cz/file/43792_1_1/ [10 June 2023]

Naylor, S., Keogh, B. and Goldsworthy, A. (2005) *Active Assessment for Science: Thinking, Learning and Assessment in Science*, London: Routledge.

Nunes, T. and Bryant, P. (2015) *Handbook of Child Psychology and Developmental Science: Volume II - Cognitive Processes*, Hoboken: John Wiley & Sons, Inc. https://doi.org/10.1002/9781118963418.childpsy217

Opitz, A., Heene, M. and Fischer, F. (2017) 'Measuring scientific reasoning - a review of test instruments', *Educational Research and Evaluation*, Vol. 23, No. 3-4, pp. 78-101. https://doi.org/10.1080/13803611.2017.1338586

Padilla, M. J. (1990) *The Science Process Skills*, [Online], Available: https://narst.org/research-matters/science-process-skills [20 June 2022]

Pape, S. J. (2004) 'Middle School Children's Problem-Solving Behavior: A Cognitive Analysis from a Reading Comprehension Perspective', *Journal for Research in Mathematics Education*, Vol. 35, No. 3, pp. 187. https://doi.org/10.2307/30034912

Pólya, G. and Conway, J. H. (2004) *How to solve it: a new aspect of mathematical method,* Princeton: Princeton University Press.

Potužníková, E., Janotová, Z. and Blažek, R. (2019) *Mezinárodní šetření PISA 2018: Koncepční rámec hodnocení čtenářské gramotnosti [PISA 2018: conceptual framework for reading literacy]*, Praha: Česká školní inspekce.

Raven, J., Raven, J. C. and Court, J. H. (2000) *Manual for Raven's Progressive Matrices and Vocabulary Scales: Section 3: The standard progressive matrices (including the parallel and plus versions)*, Oxford: Oxford Psychologists Press.

Rezaie, M. and Gooya, Z. (2011) 'What do I mean by combinatorial thinking?', *Procedia - Social and Behavioral Sciences*, Vol. 11, pp. 122-126. https://doi.org/10.1016/j.sbspro.2011.01.046

Samková, L., Rokos, L. and 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. https://doi.org/10.7160/eriesj.2021.140406

Schoenfeld, A. H. (1982) 'Measures of Problem-Solving Performance and of Problem-Solving Instruction', *Journal for Research in Mathematics Education*, Vol. 13, No. 1, pp. 31. https://doi.org/10.2307/748435

Slavík, J. (1999) *Hodnocení v současné škole: Východiska a nové metody pro praxi [Evaluation in the present school: Starting points and new methods for practice]*, Praha: Portál.

Sriraman, B. (2005) 'Are Giftedness and Creativity Synonyms in Mathematics?', *Journal of Secondary Gifted Education*, Vol. 17, No. 1, pp. 20-36. https://doi.org/10.4219/jsge-2005-389

Suurtamm, C., Thompson, D. R., Kim, R. Y., Moreno, L. D., Sayac, N., Schukajlow, S., Silver, E., Ufer, S. and Vos, P. (2016) *Assessment in Mathematics Education*, Cham: Springer. https://doi.org/10.1007/978-3-319-32394-7

Swan, M. and Burkhardt, H. (2012) 'A Designer Speaks: Designing Assessment of Performance in Mathematics', *Educational Designer*, Vol. 2, No. 5.

Syafril, S., Aini, N. R., Netriwati, Pahrudin, A., Yaumas, N. E., and Engkizar (2020) 'Spirit of Mathematics Critical Thinking Skills (CTS)', *Journal of Physics: Conference Series*, Lampung, Vol. 1467, No. 1, pp. 012069. https://doi.org/10.1088/1742-6596/1467/1/012069

Tajudin, N. M. and Chinnappan, M. (2015) 'Exploring relationship between scientific reasoning skills and mathematics problem solving', in Marshman, M., Geiger, V. and Bennison, A. (ed.) *Mathematics education in the margins: Proceedings of the 38th annual conference of the Mathematics Education Research Group of Australasia*, Australia: Mathematics Education Research Group of Australasia.

Valanides, N. (1997) 'Cognitive Abilities among Twelfth-grade Students: Implications for Science Teaching', *Educational Research and Evaluation*, Vol. 3, No. 2, pp. 160-186. https://doi.org/10.1080/1380361970030204

Vilenius-Tuohimaa, P. M., Aunola, K. and Nurmi, J. (2008) 'The association between mathematical word problems and reading comprehension', *Educational Psychology*, Vol. 28, No. 4, pp. 409-426. https://doi.org/10.1080/01443410701708228

Wong, K. Y. and Lee, N. H. (2009) 'Singapore Education and Mathematics Curriculum', in Wong, K. Y., Lee, P. Y., Kaur, B., Foong, P. Y. and Ng, S. F. (ed.) *Series on Mathematics Education*, Singapore: World Scientific. https://doi.org/10.1142/9789812833761_0002

Wu, M. and Adams, R. (2006) 'Modelling mathematics problem solving item responses using a multidimensional IRT model', *Mathematics Education Research Journal*, Vol. 18, No. 2, pp. 93-113. https://doi.org/10.1007/BF03217438

Zhou, C. (2012) 'Teaching engineering students creativity: a review of applied strategies', *Journal on Efficiency and Responsibility in Education and Science*, Vol. 5, No. 2, pp. 99-114. https://doi.org/10.7160/eriesj.2012.050205

Zimmerman, C. (2007) 'The development of scientific thinking skills in elementary and middle school', *Developmental Review*, Vol. 27, No. 2, pp. 172-223. https://doi.org/10.1016/j.dr.2006.12.001

Full research paper

COLLABORATIVE LEARNING BASED ON SOPHISTICATED THINKING LABORATORY (STB-LAB) AND GATHER TOWN AS GAMIFICATION TOOL FOR BLENDED LABORATORY ON SCIENCE UNDERGRADUATE STUDENTS

Rena Denya Agustina✉
Riki Purnama Putra
Milla Listiawati

UIN Sunan Gunung Djati Bandung, Indonesia

✉ renadenya@uinsgd.ac.id

ABSTRACT

The evolution of learning design continues, focusing on blended laboratory approaches incorporating technology. The Sophisticated Thinking Laboratory (STB-LAB) and Gather Town are key tools in implementing collaborative learning in this context. This study assesses the effectiveness of blended laboratory implementation using STB-LAB and Gather Town as a gamification tool, utilizing the Assessment Based on Teaching and Learning Trajectory (AABTLT) with Student Activity Sheets (SAS). The results reveal the successful execution of STB-LAB syntax in blended laboratory activities. Additionally, STB-LAB and Gather Town significantly enhance students' collaborative skills, as indicated by a substantial Cohen's D Effect Size. For physics education majors, the effect size is 1.736 in the experiment group and 0.754 in the control group, while for biology education majors, it is 1.522 in the experiment group and 0.541 in the control group. This study highlights the positive impact of blended laboratories with STB-LAB and Gather Town on collaborative skill development, further emphasizing the role of technology in contemporary learning design.

KEYWORDS

Blended laboratory, collaborative learning, collaborative skills, gather town, STB-LAB

HOW TO CITE

Agustina R. D., Putra R. P., Listiawati M. (2024) 'Collaborative Learning Based on Sophisticated Thinking Laboratory (STB-LAB) and Gather Town as Gamification Tool for Blended Laboratory on Science Undergraduate Students', *Journal on Efficiency and Responsibility in Education and Science*, vol. 17, no. 1, pp. 66-77. http://dx.doi.org/10.7160/eriesj.2024.170106

Article history
Received
January 23, 2023
Received in revised form
March 26, 2023
Accepted
October 19, 2023
Available on-line
March 31, 2024

Highlights

- *The implementation of blended laboratory activities using the STB-LAB model assisted by Gather Town as a gamification tool has achieved good results.*
- *The results of the hypothesis testing show that there is an increasing effect on collaborative skills when using the STB-LAB model.*
- *Gather Town shows features that can be used to carry out blended laboratory activities.*
- *The enhancement effect shows that there is a large effect for the experiment group in the physics education major and biology education major.*

INTRODUCTION

21st-century skills are skills needed to support life in the future. These 21st-century skills are an important component in the world of education, especially for the next generation of successors who will enable them to build and bring about change on a national or global scale (Van Laar et al., 2017). The importance of 21st-century skills is demonstrated by the rapid development of science and technology, where advances in various technologies require qualified skills to control or create technology (Jacobson-Lundeberg, 2016).

21st-century skills, characterized by 4C as the main skills, include; (1) Critical thinking skills; (2) Creative thinking skills; (3) Communication skills; and (4) Collaborative skills. The 4C skills are expected to facilitate students in supporting their lives later when they enter the world of work so that these students are able to compete on a global scale (Chalkiadaki, 2018). However, fulfilling the 4C skills requires a process, and, of course, there is one skill that is difficult to practice (Punya Mishra and Mehta, 2017). One example of a skill that is difficult to train is collaborative skills because, in collaborative skills,

it is sometimes difficult for educators to monitor all activities carried out by students (Rahman et al., 2019). In addition, collaborative skills are difficult to train because sometimes students feel indifferent to what is given by their educators because they feel that they are not being cared for by educators (Hur, Shen and Cho, 2020).

The difficulty in training 21st-century skills is due to various background factors, one of which is the condition factor that changes habits or transformations in the world of education (Bell, 2016). Changes in habits in the world of education that are often felt in the 21st century are when a pandemic occurs, which limits human activities such as learning and teaching. One of the difficulties when carrying out virtual learning is monitoring students in collaborative learning (Koşar, 2021). We believe that the pandemic will provide new changes or transformations in the world of education, in which many new innovations will emerge, such as the opinion expressed by Chick (2020), which states that innovation occurs due to the pressure of circumstances that make people think of seeking other alternatives as one of the conveniences in carrying out activities. In addition, Kang (2021) argues that innovation in education during a pandemic is shown by the many innovative learning models and methods that make it possible to do without or with minimal physical contact. In addition, we need media that can accommodate virtual, real, or blended learning activities. Because a medium used in learning activities is felt to be able to bring interest, even ease in the process of transferring knowledge in learning (Williamson, Eynon and Potter, 2020).

Many innovations have sprung up and been developed in the world of education, especially in learning models; of course, they must be studied more deeply by analyzing the syntax of the learning model. In the learning model innovations that emerged during the pandemic, one of them was carried out by Agustina, Putra, and Listiawati (2022), where they developed a laboratory activity-based learning model that combines virtual and real called the Sophisticated Thinking Blended Laboratory (STB- LAB). STB-LAB has a syntax that combines virtual and real activities into a unified learning design, in which STB-LAB combines LOTS and HOTS as a skill foundation to fulfill 21st-century skills. The syntax owned by STB-LAB is deemed qualified to support the skills requirements of the 21st Century, where the syntax is; (1) Disposition stages; (2) Argumentation stages; (3) Verification stages; (4) Laboratory stages; and (5) Communication stages. The five stages of the STB-LAB are certainly felt to be able to train collaborative skills in collaborative learning because there is two-way learning between educators and students, both virtually and in real terms. Collaborative learning at STB-LAB is, of course, focused on the argumentation stage, up to the laboratory stage because these stages have two-way characteristics, where when students give arguments against a problem, there will be comments from educators so that students design their arguments well and perfect where of course the ability to argue is properly trained using the STB-LAB model (Agustina and Putra, 2022). Furthermore, at the verification stage, the students are expected to be able to seek verification of the arguments formed per group so that the group can exchange ideas with one another, but still, at the verification stage, it is monitored by the educator so that misunderstandings do not occur. At the laboratory stage, all groups that have been formed conduct real and virtual experiments to test their arguments.

Using the STB-LAB model in blended laboratories certainly requires a media that supports students in carrying out laboratory activities. Of course, the media used must have uniqueness and adequate features, such as media that can be used as a gamification tool (Sailer and Homner, 2020). This gamification will certainly make it interesting for students to carry out learning activities because an attractive visual appearance will stimulate students. As if they are playing a game, but in fact, they are carrying out learning activities (Majuri, Koivisto and Hamari, 2018). In addition, gamification in blended learning or blended laboratory activities must have several features, including being able to display screen videos and face videos and open other media applications (Hallifax et al., 2019).

One possible media is Gather Town, which in Gather Town has features that support carrying out blended learning or blended laboratory because Gather Town has qualified features such as accommodating up to 40 users in a free account (Zhao and McClure, 2022). In addition, Gather Town is also able to display user screens as share screens for each user, so not only one share screen can be displayed, but all participants can share screens simultaneously, which, of course, makes it easier for students to discuss and carry out laboratory activities (McClure and Williams, 2021). Also, Gather Town has pixel visuals like the appearance of a game in the 90s, which has a certain appeal with users being able to change the characters' appearance and decorate the place that will be used as a certain room (Fitria, 2021). The decorations that can be used are very diverse and have their own functions in their features; for example, there is a blackboard that can be used as a feature for writing like a virtual whiteboard, and there are also posters or televisions to display static images or display videos that have buttons to trigger the video when other users want to see what's in the video (Latulipe and De Jaeger, 2022; Lee et al., 2023).

With the syntax that is owned by STB-LAB and the features that Gather Town owns, it is felt that it is possible to make collaborative learning work well, so this research has two objectives, namely to find out how well the implementation of blended laboratory activities based STB-LAB with the help of Gather Town as a gamification tool for blended laboratory and finding out how STB-LAB influences with the help of Gather Town in the implementation of collaborative learning can improve students collaborative skills while carrying out blended laboratory activities. The hypothesis designed in this study is that there is a positive difference in the average score between the pretest and post-test for the ability to collaborate in the experimental group or the group using the STB-LAB model, so the STB-LAB can improve students' collaborative skills. The hypothetical design is applied to major physics education and biology education.

METHOD

This research uses a quasi-experimental method, where it focuses on finding quantitative results, which will later be described as a whole from a series of learning activities using

the STB-LAB model assisted by Gather Town as a gamification tool. The Blended Laboratory is used because the STB-LAB model has the characteristics of combining two laboratory activities virtually or in real terms. This laboratory activity was carried out for three meetings.

PARTICIPANTS

This study used 122 students as subjects from two different majors at the same university. The majors used in this research are physics education and biology education. The details of the participants used by the researcher can be seen in Table 1.

Major	Participants	Gender	
		Male	Female
Physics Education	62	36	26
Biology Education	60	22	38
Total	122	58	64

Table 1: Participant description

All participants used in this study were in the same semester, namely in the second semester of the 2021/2022 academic year. Participants were not informed that research would be carried out, so all participants went naturally, without any contrivance. Furthermore, the participants were divided into two groups in their respective departments, with the group design used as two pretest-posttest groups. This means that each department has two different groups, namely the control and experimental groups. In total, there are four groups, namely the control group in Physics Education major and Biology Education. At the same time, the experimental class also has the same grouping as the control group, as the distribution of the control and experimental groups can be seen in Table 2.

Major	Group	Number of Participants
Physics Education	Experiment	31
	Control	31
Biology Education	Experiment	30
	Control	30

Table 2: Group description

In detail, the groups in the physics education major were divided into two groups of a total of sixty-two participants, with an experimental group of thirty-one subjects and a control group of thirty-one subjects. In addition, the division of the experimental and control groups in the biology education major was the same as the physics education major, where a total of sixty participants were divided into two. In this case, the experimental and control groups use different subjects, so the subjects in the experimental and control groups are not the same. In contrast, in the physics and biology education majors, there are three classes in each generation, and each major uses one class for the experimental and control classes. In detail, when major physics education has three classes, namely A, B, and C, in the second semester, class a will be the experimental group, and class B will be the control group. The determination of the experimental and control groups in the biology education major is the same as described in the determination of the experimental and control groups in the physics education major.

Laboratory Activities Description

Laboratory activities are carried out in two conditions, namely using a virtual laboratory and a real laboratory, for the activity model using the STB-LAB model, with gamification media using Gather Town as a means of communication between participants who carry out real laboratory activities, and real laboratory activities that can be seen in Figure 1.

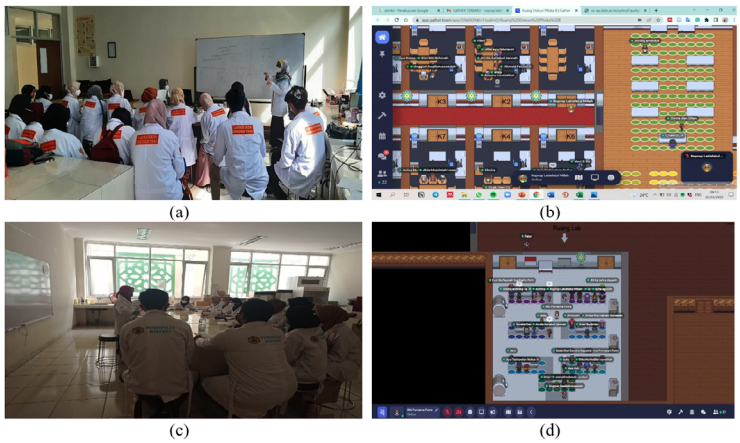

Figure 1: (a) Real laboratory class on physics education; (b) Virtual laboratory class on physics education; (c) Real laboratory class on biology education; and (d) Virtual laboratory class on biology education

As for the Physics Education major, it uses refraction material with tools in the form of a traditional spectrometer for a real laboratory and a traditional virtual spectrometer using a virtual laboratory made by Amrita (https://vlab.amrita.edu/) for a virtual laboratory. In addition, the Biology Education major uses drosophila material, with the tools used in the form of stereo microscopes and monoculars for real laboratories and the use of virtual observations made by FlyLab JS (https://www.sciencecourseware.org/) for virtual laboratories.

Instruments

The instruments used in this study were used to observe the implementation of collaborative learning when using STB-LAB media assisted by Gather Town as a gamification tool. The implementation of collaborative learning is based on the assessment of the implementation presented by several researchers, such as (1) Knowledge Construction (KC); (2) Instructional Activities (IA); (3) Anti-competition (AC); (4) Group Participation (GP); and (5) Social Interaction (SI) (Le, Janssen and Wubbels, 2018; Strijbos, 2016). The entire process of laboratory activities is measured for its implementation using an assessment of the implementation of the Assessment Based on Teaching and Learning Trajectory (AABTLT) with Student Activity Sheet (SAS) developed by Rochman (2017), where AABTLT is observed and assessed by observers other than the teacher/assistant who supervises the activities the laboratory. Students will fill in questions in each series of blended laboratory activities in a short time, which later results from AABTLT and SAS will be given a score range of 1 - 5. The average value of the two assessments is sought, so that the assessment will be graded objectively and in detail. The results of implementing blended laboratory activities can be interpreted according to the percentages obtained, which can be seen in Table 3.

Percentage (%)	Performance Interpretation
80 - 100	Very good
60 - 79	Well
40 - 59	Pretty good
20 - 39	Bad
< 20	Very bad

Table 3: AABTLT with SAS interpretation

In addition, the instrument for measuring the results of the collaborative skills of students uses assessment instruments from several researchers, with the aspects used namely: (1) Task focus; (2) Participation; (3) Knowledge sharing; (4) Reliability; and (5) Socio-Cultural.

For the Task focus (TF), the assessment consists of three items based on the assessment conducted by Häkkinen (2017), where the assessment is assessed to determine how consistent the students are in staying focused on the task at hand. An example of the assessment items is: "Students are not distracted by other assignments, so students only do what is assigned".

For Participation (PC), the assessment consists of three assessment items based on the assessment presented by Hesse (2015), where the researcher takes an assessment to find out how students participate in their group individually or in their role as leaders, with this assessment assessed then will know the participation of students in group discussions. An example of the assessment items is: "Students are able to lead the discussion so that there are no distractions during the discussion".

For Knowledge sharing (KS), the assessment consists of four assessment points, which are based on the assessment presented by Care (2016), in which the assessment is assessed to find out how active the student is in sharing opinions according to their capacity. With the benchmark of student activity, it can be seen when these students can explain their knowledge, accept other people's opinions, and correct other people's opinions so that their group can accept them. An example of the assessment items is: "Students are able to give opinions in accordance with the concepts, theories, and problems being faced".

For Reliability (RB), the assessment consists of four assessment points, which are based on the assessment presented by Widana (2018), where the assessment is assessed to find out how the student does not rely too much on others in his work. The assessment is also assessed in a personal review, which allows it to describe whether the student is independent or able to be invited to collaborate with mutually beneficial work with one another. An example of the assessment items is: "Students are able to independently carry out their duties without making it difficult for others".

For socio-cultural (SC), the assessment consists of three assessment items based on the assessment described by de Hei (2020), where the assessment is used to find out how students listen, think, and discuss further in their groups. This assessment is also assessed in a review of interpersonal interactions, allowing it to provide an overview of whether students are actively interacting with their group. An example of the assessment items is: "Students are able to discuss well and have ethics in discussions".

In summary, how the instruments for assessing student collaborative skills per indicator are shown in Table 4.

All assessments to assess collaborative abilities are used by lecturers to observe and assess students' collaborative abilities by providing assessments using a Likert scale of 1 - 5, which is then interpreted as a percentage per indicator or whole indicator.

Indicators	Description	Example for observation
TF	How can the students focus on their tasks and consistently do their role in group	Students are not distracted by other assignments, so students only do what is assigned
PC	How the students can participate in the discussion in their group and how they can lead the group	Students are able to lead the discussion so that there are no distractions during the discussion
KS	How the students share their opinions, knowledge, and ideas with the groups	Students are able to give opinions in accordance with the concepts, theories, and problems being faced
RB	How can the students be independent or help other members of their group to do the task	Students are able to independently carry out their duties without making it difficult for others
SC	How the students interact and respect the members of their groups	Students are able to discuss well and have ethics in discussions

Table 4: Collaborative skills indicator description and example of observation

Meeting Description and Data Collection

This study took data from the experimental and control groups in major physics and biology education. Data were collected using the pretest and posttest, where the instruments used in the pretest and posttest used the instruments previously described. In practice, major physics education and biology education were carried out on the same day, with details of the experimental group being given the STB-LAB treatment model, where the implementation was carried out in the morning class, and the control group was carried out in the afternoon class. The meeting was held in five meetings within one week of one meeting, where the first meeting was a pretest for the researcher to observe the initial collaborative results of the subjects. Then, the second to fourth meeting became a treatment, with details in the experimental class in both major physics education and biology education, carrying out disposition sessions to arguments at the second meeting. The third meeting held a verification and laboratory session, then the fourth meeting held a communication session. In the fifth meeting, a posttest was carried out for the researcher to observe the final collaborative results of the subjects.

Based on the control class and the details of each major physics and biology education meeting, a pretest was conducted for the researcher to observe the students' initial collaborative results. In the second meeting, the presentation of the theory was carried out, followed by the third meeting, where students carried out trials; then, in the fourth meeting, the presentation of the results of trials by students was carried out. In the fifth meeting, a posttest was carried out for the researcher to observe the final collaborative results of the subjects. All observations in both the experimental and control groups in major physics education and biology education were assisted in observing student collaborative assessments with the assistance of three laboratory assistants.

All the implementation of activities is assessed using authentic assessment with AABTLT with SAS, where the implementation of the teacher is assessed by the observer in accordance with the implementation of the syntax. The implementation of students is assessed by giving one or two short questions regarding activities according to the syntax carried out to students which students then carry out. All the work given to students regarding brief questions regarding activities according to the syntax is assessed using a scale of 1 - 5 which will later be used as a percentage. The pretest and posttest regarding a collaborative that researchers value use an assessment with a scale of 1 - 5, which will later be used as a percentage.

DATA ANALYSIS

The data was processed using quantitative analysis to find out how big the percentage of implementation is using the percentage equation as described by Listiawati (2022), which later on this percentage will be analyzed per aspect and as a whole. Where to analyze the implementation of blended laboratory activities based on STB-LAB with the help of Gather Town as a gamification tool, the percentage of implementation is calculated by means of a data review based on the implementation value in each syntax, which is then averaged and made into a percentage of the average results, so that by calculating the percentage implementation with the AABTLT with SAS assessment will answer the first goal with an interpretation of the percentage gain, namely: (1) < 59% (Not Good); (2) 60% - 69% (Less); (3) 70% - 79% (Well); (4) > 80% (Very Good) (Zakwandi, Yuningsih and Setya, 2020).

Then, in testing the hypothesis and to answer the second objective, this study used a paired sample *t*-test, which in testing, because the data used uses different populations in the major and different samples in each group, so that in one major, there is interference in the control class (Kim, 2015). Also, this hypothesis requires a pre-assumption test, which includes a normality test, and a homogeneity test. The normality test used in this study was tested as a whole for data acquisition with the number of samples above status, so the test used two tests, namely Liliefors, and Shapiro-Wilk, which were able to describe normality well in acquisition values that had control and experimental classes with a total subject range of thirty to forty (Razali and Wah, 2011). Furthermore, a homogeneity test using the Levene test was carried out to test whether the data is homogeneous when the samples used between research groups have differences in the number of samples. This will reduce the risk of invalidity of the data to test the hypothesis (Prabhaker Mishra et al., 2019). Meanwhile, testing homogeneity is not mandatory if the hypothesis test is carried out using the paired sample *t*-test because the paired *t*-test only requires a normality test to fulfill the classic assumption test. Also, the paired *t*-test must be in the same number of subjects in each group used (Lakens,

2017), but in this study, a homogeneity test will also be presented to detail the results of the acquisition of statistical data. In addition, this study will also examine the level of effectiveness of using the STB-LAB on students' collaboration skills using Cohen's D Effect Size, which in the Cohen's D Effect Size will be tested based on the mean and standard deviation obtained so that it can clearly describe the size of the effect given during treatment (Lee, 2016). The use of the effect size is based on the fact that the paired *t*-test only describes, in general, the results of decision-making but does not describe the magnitude of the influence, so as a further explanation to detail it again, tests must be used to test these effects, one of which is the Cohen's D Effect Size (Kraft, 2020).

RESULTS AND DISCUSSION

This study obtained results in the form of a percentage of implementation, primary data, normality test, homogeneity test, hypothesis test, and effect size, where the initial results that will be presented are the percentage of implementation to find out the results of the implementation of blended laboratory activities using the STB-LAB model assisted by Gather Town as a gamification tool. Following are the results of the percentage of blended laboratory activities using the AABTLT with SAS assessment, which can be seen in Table 5.

Table 5 shows that, on average, from the three meetings, the percentage of implementation was 78.59%, with the interpretation of its implementation being in the Well category. The lowest results were from the first meeting, especially in educator activities, namely at the communication stage, which obtained results of 77.39%. In contrast, the highest results were located at the laboratory stage, which obtained results of 83.08%, and the lowest results from the first meeting on student activities, namely at the communication stage as well which got a result of 71.77%, while the highest result was located in the argumentation stage which got a result of 83.06%. At the second meeting, the lowest results were for educator activities, namely at the communication stage, which obtained results of 76.72%, while the highest results were at the laboratory stage, which obtained results of 83.08%, and the lowest results from the second meeting on student activities, namely at the communication stage as well which got a result of 70.16, while the highest result was in the argumentation stage which got a result of 79.83%. At the third meeting, the lowest result was for educator activities, namely at the argumentation stage, which got results of 77.42%, while the highest results were at the disposition stage, which got results of 83.08%. The lowest result from the second meeting was on student activities, namely at the communication stage which gets a result of 71.77%, while the highest result is in the argumentation stage which gets a result of 83.06%.

Blended Laboratory Activities		Percentage of Activity Implementation (%)		Total Percentage (%)	Interpretation
Meeting	Syntax	Lecturer	Students		
1	Disposition	82.22	80.64	81.43	Very Good
	Argumentation	78.19	83.06	80.62	Very Good
	Verification	82.78	79.03	80.90	Very Good
	Laboratory	83.08	77.41	80.24	Very Good
	Communication	77.39	71.77	74.58	Well
2	Disposition	78.27	76.61	77.44	Well
	Argumentation	80.62	79.83	80.22	Very Good
	Verification	77.42	75.80	76.61	Well
	Laboratory	83.03	72.58	77.80	Well
	Communication	76.72	70.16	73.44	Well
3	Disposition	83.08	79.83	81.30	Very Good
	Argumentation	77.42	83.06	80.24	Very Good
	Verification	82.22	76.61	79.41	Well
	Laboratory	76.72	78.22	77.47	Well
	Communication	82.78	71.77	77.27	Well
Average				78.59	Well

Table 5: Results of the implementation of blended laboratory activities for three meetings based on AABTLT with SAS

Description of research findings data on student collaborative assessment, both in major physics education and biology education, which includes the average pretest and posttest scores in each group used in the study, and all assessments are presented per indicator of collaborative ability, which can be seen in Table 6.

Table 6 shows the average value of the experimental and control groups based on their indicators; the experimental group in major physics education obtained the difference in scores between the pretest and posttest of 15.645 points, with a pretest gain of 68.387 and a posttest gain of 84.032. The highest score obtained on the posttest in the major physics education experimental group was on the RB indicator, with a score of 86.329. The control group in major physics education obtained a difference between the pretest and posttest of 8.065 points, with a pretest gain of 68,064 and a posttest gain of 76.129.

The highest score obtained in the post-test control group was on the PC indicator, with a score of 80.115. Based on the results of the difference in the pretest and posttest average scores between the control and experimental classes in major physics education, a difference of 0.323 points in the pretest and 7.903 points in the posttest, with the experimental class being superior to the control class.

Table 6 also provides information on the average scores of the experimental and control groups based on their indicators in major biology education, with the results of the difference in scores between the pretest and posttest in the experimental group of 16.500 points, with pretest gains of 65.666 and posttest gains of 82.166. The highest score obtained in the posttest of the major biology education experimental group was on the RB indicator, with a score of 85.685. The control group in major biology education obtained a difference between the pretest and posttest of 15.533 points, with a pretest gain of 63.633 and a posttest gain of 79.166. The highest score obtained in the post-test control group was on the PC indicator, with a score of 82.905. Based on the difference in the pretest and posttest average scores between the control and experimental classes in major biology education, a difference of 2.033 points was obtained in the pretest and 3.000 points in the posttest, with the experimental class being superior to the control class.

Major	Group	Indicator	Pretest	Posttest	Pretest Total	Posttest Total
Physics Education	Experiment	TF	70.528	79.449	68.387	84.032
		PC	71.629	85.702		
		KS	62.276	82.232		
		RB	66.115	86.329		
		SC	71.388	86.450		
	Control	TF	70.648	76.232	68.064	76.129
		PC	70.635	80.115		
		KS	62.777	74.777		
		RB	65.115	73.405		
		SC	71.148	76.115		
Biology Education	Experiment	TF	65.232	82.105	65.666	82.166
		PC	67.555	78.245		
		KS	60.227	84.528		
		RB	72.135	85.685		
		SC	63.184	80.270		
	Control	TF	65.425	77.227	63.633	79.166
		PC	72.115	82.905		
		KS	63.777	76.135		
		RB	62.343	78.449		
		SC	54.505	81.115		

Table 6: Primary data descriptive

Then, the classic assumption test is carried out by carrying out the normality and homogeneity tests. The first test to be carried out is the normality test based on the major, so this review of the normality test is not carried out thoroughly. Still, it is divided based on the subject in order to detail and detail the normality results, where the normality test can be seen in Table 7.

Major	Data Type	Liliefors			Saphiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Physics Education	Pretest (Experiment)	.147	31	.086	.964	31	.361
	Posttest (Experiment)	.133	31	.173	.935	31	.060
	Pretest (Control)	.128	31	.200	.958	31	.258
	Posttest (Control)	.156	31	.054	.935	31	.059
Biology Education	Pretest (Experiment)	.126	30	.200	.963	30	.378
	Posttest (Experiment)	.156	30	.059	.937	30	.078
	Pretest (Control)	.141	30	.134	.936	30	.072
	Posttest (Control)	.151	30	.079	.955	30	.224

Table 7: Normality test of collaborative skills data in physics education major and biology education major

Table 7 shows that the use of all data is normal, with an alpha (α) used, namely 5%, or 0.05; it can be concluded that if the results obtained for normality show > 0.05, then the decision taken is that the data is normal. None of the data is abnormal from the data used, so the primary data used to test the hypothesis can be used to test the paired sample *t*-test. Still, before that, a homogeneity test will be carried out using the Levene test based on major and broken down into classes per class. The homogeneity test results can be seen in Table 8.

Major	Type	Levene Statistic	df1	df2	Sig.
Physics Education	Pretest	1.303	1	60	.258
	Posttest	1.499	1	60	.226
Biology Education	Pretest	1.640	1	57	.205
	Posttest	.017	1	57	.896

Table 8: Homogeneity test of collaborative skills data in physics education major and biology education major

Table 8 shows that the use of all data is homogeneous, with α used, which is 5% or 0.05; it can be concluded that if the results obtained for homogeneity show > 0.05, then the decision taken is that the data is homogeneous. None of the data is homogeneous from the data used. The results of this homogeneity aim to strengthen the data used so that none of the data is inaccurate in statistical calculations. Furthermore, when the normality and homogeneity tests have been fulfilled, the paired sample *t*-test is carried out to test the hypotheses taken, with the interpretation of decision-making rejecting H_0 if the value of Sig. (2-tailed) obtains a value smaller than α (5%) or does t reject H_0 if the value is Sig. (2-tailed) obtains a value greater than α (5%). As a guideline for making hypothetical decisions, the hypothetical decisions and their description can be seen in Table 9. The results of hypothesis testing using the paired sample *t*-test can be seen in Table 10.

Decision	Description
Sig. (2-tailed) < 0.05	Reject H_0: There is a difference in the average score between the pretest and posttest scores after the treatment
Sig. (2-tailed) > 0.05	Do not reject H_0: There is no difference in the average score between the pretest and posttest scores after the treatment

Table 9: Research hypothesis

Major	Group	Pair Differences					*t*	df	Sig. (2-tailed)
		Mean	Std. Dev	Std. Error	95% Confidence Interval of the Diff				
					Lower	Upper			
Physics Education	Experiment	15.645	9.013	1.619	12.339	18.951	9.665	30	< .001
	Control	8.065	10.699	1.922	4.140	11.989	4.197	30	< .001
Biology Education	Experiment	16.500	10.840	1.979	12.452	20.584	8.337	29	< .001
	Control	5.867	10.849	1.981	1.815	9.918	2.962	29	.006

Table 10: Paired sample t-test for hypothesis decision

Table 10 shows that the paired *t*-test results can be seen in the acquisition of Sig. (2-tailed), where with an α of 5%, or 0.05, the H_0 will be rejected if the value of Sig. (2-tailed) < α, where the results obtained show the value of Sig. (2-tailed) obtained a total value of < .001 for experimental and control groups in the physics education major, which means that in the experiment group, the application of the STB-LAB model assisted by Gather Town as a gamification tool can affect the improvement of students' collaborative skills, and the control group is also able to influence the improvement of students' collaborative abilities. The results obtained show the value of Sig. (2-tailed) for biology education major obtained a total value of < .001 for the experiment group, while the control group obtained a value of 0.006, which means that in the experiment group, the application of the STB-LAB model assisted by Gather Town as a gamification tool can affect the improvement of students' collaborative skills, and the control group is also able to influence the improvement of students' collaborative abilities. Furthermore, to find out the magnitude of the effect, an effect size test is carried out with Cohen's D Effect Size test, with an interpretation of the effect size gain that is obtained if the value is 0.2. The effect is small, 0.5, then the effect is medium, and > 0.8 effect is large, the results of the Cohen's D Effect Size test can be seen in Table 11.

Major	Group	*N*	Mean	Std. Deviation	*t*	Effect Size	Interpretation
Physics Education	Experiment	31	15.645	9.013	9.665	1.736	Large
	Control	31	8.065	10.699	4.197	0.754	Medium
Biology Education	Experiment	30	16.500	10.840	8.337	1.522	Large
	Control	30	5.867	10.849	2.962	0.541	Medium

Table 11: Cohen's D Effect Size results

Table 11 shows that the results of the Cohen's D Effect Size test obtained from the physics education major are equal to 1.736 for the experiment group, where the interpretation obtained is that there is a large improvement effect and for the control group to get a value of 0.754, where the interpretation obtained is that there is a medium improvement effect. The effect size in biology education major is equal to 1.522 for the experiment group, where the interpretation obtained is an effect large improvement, and for the control group to get a value of 0.541, where the interpretation obtained is that there is a medium improvement effect. The results of the two majors show that the effect of using the STB-LAB model with the help of Gather Town as a gamification tool to improve students' collaborative skills is relatively large in terms of improvement.

Based on the results from Table 5, the implementation of blended laboratory activities using the STB-LAB model assisted by Gather Town as a gamification tool was obtained; an average of 78.59% was obtained, with a good implementation category. A good implementation category was obtained because, based on their activities, students usually like laboratory activities more than studying theory. In line with research conducted by Rashidovna (2020), which states that students in the 21st century tend to have an interest in laboratory activities because laboratory activities students feel they have experience as researchers, Estriegana (2019) states that laboratory activities have their characteristics, where students are able to learn theory and also practice based on direct observation or acquisition from experimental results which then obtained experimental data results will be synthesized based on the applicable theory. In its implementation, blended laboratory activities are felt to have their challenges in their implementation because educators carrying out blended laboratory activities must have two focuses; of course, these focuses are sometimes divided, or it could be that one of the activities is not carried out properly (Ożadowicz, 2020). But in STB-LAB, of course, this can be handled well because STB-LAB has a syntax carried out in parallel between virtual and real, allowing these two activities to run properly.

In collaborative learning, activities between educators and students must be related to one another, where students and educators must jointly solve common problems to obtain results that are in accordance with theory (Jeong and Hmelo-Silver, 2016). The STB-LAB model is certainly able to facilitate collaborative learning well, shown from the implementation of blended laboratory activities between educators and students, which is getting very good results because STB-LAB, in its activities, requires educators to always provide feedback and related suggestions. Research will be carried out by students in blended laboratory activities. However, blended laboratory activities require a medium that can facilitate between virtual and real activities in real-time (Dangwal, 2017); this can be resolved by using the right technology because if a technology can be used properly, a problem can occur. Resolved, as in a blended laboratory, a technology is needed that can unite the virtual and real (Melis et al., 2019). The use of Gather Town is considered very appropriate in carrying out blended laboratory activities, where Gather Town can facilitate virtual and real activities on one server at the same time. Users who carry out virtual and real laboratory activities together carry out laboratory activities in the same room according to the design. That has been designed by educators. Gather Town plays an important role in the implementation of bled laboratory because all of Gather Town's features are deemed appropriate for the need to carry out blended laboratory, such as sharing screens or activating the camera together, and has its own discussion room according to the group without requiring a break-out room.

In terms of improving students' collaborative skills, based on tracking the results of statistical tests carried out, it shows that the STB-LAB model is able to improve students' collaborative skills with the effect obtained, namely the large effect on physics education major and biology education major. This acquisition was obtained because the disposition stage until communication requires educators to monitor and direct students in each activity. Students carry out the disposition stage to discuss a given problem based on the educator's design; from this disposition stage, students must collaborate well with their friends to narrow down a broad problem into a detailed one. Based on the narrative from previous researchers, which stated that giving a problem to be solved to students where students will consciously carry out discussions with their friends to understand the problem to be solved (Eyisi, 2016). In addition, group discussions to understand problems can hone students' collaborative skills because there is an interaction between friends in the group (Delamont, 2017).

The next stage is the argumentation stage, which in this argumentation stage does not really form the collaborative skills of students because the argumentation stage focuses on interactions between educators and students in this argumentation. Stage it focuses on forming collaborative learning, where students describe their arguments against a problem, which then the educator will provide enlightenment or provide input on what steps must be taken by these students in getting the answer with laboratory activities. This interaction can create good collaborative learning because educators act as facilitators for students in building their cognitive abilities (Hadwin, Järvelä and Miller, 2017). In addition, providing the best directions for students and educators in laboratory activities will minimize misconceptions that occur so that students will be very well-formed (Wismath and Orr, 2015).

In the verification stage and in the laboratory stage, interaction will be created between students in their groups, where students will conduct a review by conducting a literature study to deepen again regarding the steps of laboratory activities to be carried out, which, of course, requires further discussion in providing various opinions of these students. Sharing opinions in exploring knowledge, of course, will have the impact of good collaboration, where good collaboration is when students are able to express their opinions well and are able to accept other people's opinions well or reject them wisely so that all of these opinions will produce conclusions fast (Bower, Lee and Dalgarno, 2017). The verification activities carried out by students, resulting in a discussion period for verification of ten to twenty-five minutes for each group, show that students are able to collaborate well with their groups. The good in collaboration can also be determined based on the completion

time of the discussion, where previous researchers explained that good collaboration is when you can streamline your time in discussions, but back in a quality perspective, sometimes fast collaboration does not always show good quality (Schneider and Pea, 2017). The quality of the discussion results will be proven by concrete evidence (Rozo et al., 2016). Real evidence of the results of the quality of the discussion is evidenced by the laboratory stage, where students will carry out experiments based on the results of discussions with their group mates.

The experimental results obtained by students will be communicated through the stages of communication. Based on the laboratory stages, students were seen to be very active in carrying out experiments in accordance with the division of tasks from the results of previous discussions; not a few of these students worked together with other friends to achieve their goals, for example, when using a traditional spectrometer they were divided into two observers, one leader to find the right color spectrum, and two people to record the results. However, from the division of tasks, it can be seen that students take turns as data is exchanged so that all students feel and know how to use tools, observe, and sort data. Of course, this is also known as a chain role, where the chain role in this laboratory activity is defined as exchanging resources to gain experience on an equal footing with others (Kirschner et al., 2018). This illustrates good collaboration so that students are able to understand the meaning of collaboration, which, of course, will increase the students' collaborative skills.

The obstacles that occur during blended learning using the STB-LAB model are interactions between students virtually and students in real terms. These obstacles focus on the internet connection, which is sometimes interrupted, thus hindering the course of activities. Internet connection is very important in smoothly running blended laboratory activities because internet connection is the most important component in smooth communication between students (Heflin, Shewmaker and Nguyen, 2017). Sometimes, miscommunication occurs between educators, virtual students, and real students, so the students' understanding is sometimes not conveyed properly. Of course, poor delivery due to an internet connection will hinder students' cognitive development and collaboration between friends, so good collaboration in a blended laboratory must be balanced with a good internet connection (Duţă and Martínez-Rivera, 2015).

The limitation of this research is that it is limited to the majors used, where the science education major must cover physics, chemistry, and biology education majors, or those that concentrate specifically on science education. Still, this study only uses physics education majors and biology education. In addition, this study was limited in terms of the number of subjects used because the Gathertown platform was limited to forty-five participants, and researchers did not have the extra budget to design this study on a larger scale.

CONCLUSION

Based on the results obtained from this study, there were several direct findings and unexpectedly, were direct findings for the first goal, namely the implementation of blended laboratory activities using STB-LAB with the help of Gather Town as a gamification tool for blended laboratory assessed using AABTLT with SAS obtained good implementation results so that the activities can run well. Then, for direct findings on the second goal, there was a difference in scores between the pretest and posttest for the experimental class in both the physics and biology education major, so this shows that STB-LAB with the help of Gather Town as a gamification tool for blended laboratory can improve collaborative skills. The magnitude of the influence can be known by showing an effect size for the physics education major in the experimental group, which got a score of 1.736 and 0.754 in the control group. In addition, the biology education major in the experimental group received a score of 1.522 and 0.541 in the control group. This shows that students' collaborative skills can be improved greatly by using STB-LAB. Indirectly, this study obtained results in the form of several obstacles that occurred in blended laboratory activities using the Gather Town-assisted STB-LAB model as a gamification tool, namely paying more attention to collaborative skills assessment, where collaborative assessment must be designed in detail so that all collaborative aspects can be assessed, besides that, the major used must be readjusted, because STB-LAB focuses on science blended laboratory activities, so it must require a complete major from science, or pure science itself.

REFERENCES

Agustina, R. D. and Putra, R. P. (2022) 'Sophisticated Thinking Blended Laboratory (STB-LAB) Learning Model: Implications on Virtual and Real Laboratory for Increasing Undergraduate Student's Argumentation Skills', *Jurnal Pendidikan IPA Indonesia*, Vol. 11, No. 4, pp. 657-671. https://doi.org/10.15294/jpii.v11i4.38772

Agustina, R. D., Putra, R. P. and Listiawati, M. (2022) 'Development of Sophisticated Thinking Blending Laboratory (STB-LAB) to Improve 4C Skills for Student as Physics Teacher Candidate', *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, Vol. 8, No. 1, pp. 65-82. https://doi.org/doi.org/10.21009/1.08107

Bell, D. V. J. (2016) 'Twenty-first century education: Transformative education for sustainability and responsible citizenship', *Journal of Teacher Education for Sustainability*, Vol. 18, No. 1, pp. 48-56. https://doi.org/10.1515/jtes-2016-0004

Bower, M., Lee, M. J. W. and Dalgarno, B. (2017) 'Collaborative learning across physical and virtual worlds: Factors supporting and constraining learners in a blended reality environment', *British Journal of Educational Technology*, Vol. 48, No. 2, pp. 407-430. https://doi.org/10.1111/bjet.12435

Care, E., Scoular, C. and Griffin, P. (2016) 'Assessment of collaborative problem solving in education environments', *Applied Measurement in Education*, Vol. 29, No. 4, pp. 250-264. https://doi.org/10.1080/08957347.2016.1209204

Chalkiadaki, A. (2018) 'A systematic literature review of 21st century skills and competencies in primary education.', *International Journal of Instruction*, Vol. 11, No. 3, pp. 1-16. https://doi.org/10.12973/iji.2018.1131a

Chick, R. C., Clifton, G. T., Peace, K. M., Propper, B. W., Hale, D. F., Alseidi, A. A. and Vreeland, T. J. (2020) 'Using technology to maintain the education of residents during the COVID-19 pandemic', *Journal of Surgical Education*, Vol. 77, No. 4, pp. 729-732. https://doi.org/10.1016/j.jsurg.2020.03.018

Dangwal, K. L. (2017) 'Blended learning: An innovative approach.', *Universal Journal of Educational Research*, Vol. 5, No. 1, pp. 129-136. https://doi.org/10.13189/ujer.2017.050116

de Hei, M., Tabacaru, C., Sjoer, E., Rippe, R. and Walenkamp, J. (2020) 'Developing intercultural competence through collaborative learning in international higher education', *Journal of Studies in International Education*, Vol. 24, No. 2, pp. 190-211. https://doi.org/10.1177/1028315319826226

Delamont, S. (2017) *Interaction in the classroom: Contemporary sociology of the school*, London: Routledge.

Duţă, N. and Martínez-Rivera, O. (2015) 'Between theory and practice: the importance of ICT in Higher Education as a tool for collaborative learning', *Procedia-Social and Behavioral Sciences*, Vol. 180, pp. 1466-1473. https://doi.org/10.1016/j.sbspro.2015.02.294

Estriegana, R., Medina-Merodio, J.-A. and Barchino, R. (2019) 'Student acceptance of virtual laboratory and practical work: An extension of the technology acceptance model', *Computers & Education*, Vol. 135, pp. 1-14. https://doi.org/10.1016/j.compedu.2019.02.010

Eyisi, D. (2016) 'The usefulness of qualitative and quantitative approaches and methods in researching problem-solving ability in science education curriculum.', *Journal of Education and Practice*, Vol. 7, No. 15, pp. 91-100. https://eric.ed.gov/?id=EJ1103224

Fitria, T. N. (2021) 'Creating sensation of learning in classroom: using'gather town'platform video game-style for virtual classroom', *Education and Human Development Journal*, Vol. 6, No. 2, pp. 30-43. https://doi.org/10.33086/ehdj.v6i2.2106

Hadwin, A., Järvelä, S. and Miller, M. (2017) *Handbook of self-regulation of learning and performance*. London: Routledge.

Häkkinen, P., Järvelä, S., Mäkitalo-Siegl, K., Ahonen, A., Näykki, P. and Valtonen, T. (2017) 'Preparing teacher-students for twenty-first-century learning practices (PREP 21): a framework for enhancing collaborative problem-solving and strategic learning skills', *Teachers and Teaching*, Vol. 23, No. 1, pp. 25-41. https://doi.org/10.1080/13540602.2016.1203772

Hallifax, S., Serna, A., Marty, J.-C. and Lavoué, É. (2019) 'Adaptive gamification in education: A literature review of current trends and developments', *European conference on technology enhanced learning*, Delft, pp. 294-307. https://doi.org/10.1007/978-3-030-29736-7_22

Heflin, H., Shewmaker, J. and Nguyen, J. (2017) 'Impact of mobile technology on student attitudes, engagement, and learning', *Computers & Education*, Vol. 107, pp. 91-99. https://doi.org/10.1016/j.compedu.2017.01.006

Hesse, F., Care, E., Buder, J., Sassenberg, K. and Griffin, P. (2015) 'A framework for teachable collaborative problem solving skills', in Griffin, P. and Care, E. (ed.) *Assessment and teaching of 21st century skills*, New York: Springer.

Hur, J. W., Shen, Y. W. and Cho, M.-H. (2020) 'Impact of intercultural online collaboration project for pre-service teachers', *Technology, Pedagogy and Education*, Vol. 29, No. 1, pp. 1-17. https://doi.org/10.1080/1475939X.2020.1716841

Jacobson-Lundeberg, V. (2016) 'Pedagogical Implementation of 21st Century Skills.', *Educational Leadership and Administration: Teaching and Program Development*, vol. 27, pp. 82-100.

Jeong, H. and Hmelo-Silver, C. E. (2016) 'Seven affordances of computer-supported collaborative learning: How to support collaborative learning? How can technologies help?', *Educational Psychologist*, Vol. 51, No. 2, pp. 247-265. https://doi.org/10.1080/00461520.2016.1158654

Kang, B. (2021) 'How the COVID-19 pandemic is reshaping the education service', in Jungwoo, L (ed.) *The Future of Service Post-COVID-19 Pandemic,*Singapore: Springer Singapore

Kim, T. K. (2015) 'T test as a parametric statistic', *Korean Journal of Anesthesiology*, Vol. 68, No. 6, pp. 540-546. https://doi.org/10.4097/kjae.2015.68.6.540

Kirschner, P. A., Sweller, J., Kirschner, F. and Zambrano R, J. (2018) 'From cognitive load theory to collaborative cognitive load theory', *International Journal of Computer-Supported Collaborative Learning*, Vol. 13, No. 2, pp. 213-233. https://doi.org/10.1007/s11412-018-9277-y

Koşar, G. (2021) 'Online collaborative learning: does it improve college students' critical reading skills?', *Interactive Learning Environments*, Vol. 31, No. 8, pp. 5114-5126. https://doi.org/10.1080/10494820.2021.1998137

Kraft, M. A. (2020) 'Interpreting effect sizes of education interventions', *Educational Researcher*, Vol. 49, No. 4, pp. 241-253. https://doi.org/10.3102/0013189X20912798

Lakens, D. (2017) 'Equivalence tests: A practical primer for t tests, correlations, and meta-analyses', *Social Psychological and Personality Science*, Vol. 8, No. 4, pp. 355-362. https://doi.org/10.1177/1948550617697177

Latulipe, C. and De Jaeger, A. (2022) 'Comparing Student Experiences of Collaborative Learning in Synchronous CS1 Classes in Gather. Town vs. Zoom', *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education,* New York, Vol. 1, pp. 411-417. https://doi.org/10.1145/3478431.3499383

Le, H., Janssen, J. and Wubbels, T. (2018) 'Collaborative learning practices: teacher and student perceived obstacles to effective student collaboration', *Cambridge Journal of Education*, Vol. 48, No. 1, pp. 103-122. https://doi.org/10.1080/0305764X.2016.1259389

Lee, D. K. (2016) 'Alternatives to P value: confidence interval and effect size', *Korean Journal of Anesthesiology*, Vol. 69, No. 6, pp. 555-562. https://doi.org/10.4097/kjae.2016.69.6.555

Lee, Y., Jung, J.-H., Kim, H., Jung, M. and Lee, S.-S. (2023) 'Comparative Case Study of Teamwork on Zoom and Gather. Town', *Sustainability*, Vol. 15, No. 2, pp. 1629. https://doi.org/10.3390/su15021629

Listiawati, M., Hartati, S., Agustina, R. D., Putra, R. P. and Andhika, S. (2022) 'Analysis of the Use of LabXChange as a Virtual Laboratory Media to Improve Digital and Information Literacy for Biology Education Undergraduate Students', *Scientiae Educatia: Jurnal Pendidikan Sains*, Vol. 11, No. 1, pp. 56. https://doi.org/10.24235/sc.educatia.v11i1.10278

Majuri, J., Koivisto, J. and Hamari, J. (2018) 'Gamification of education and learning: A review of empirical literature', *Proceedings of the 2nd international GamiFIN conference, GamiFIN 2018*, Pori, pp. 11-19.

McClure, C. D. and Williams, P. N. (2021) 'Gather. town: An opportunity for self-paced learning in a synchronous, distance-learning environment', *Compass: Journal of Learning and Teaching*, Vol. 14, No. 2. https://doi.org/10.21100/compass.v14i2.1232

Melis, L., Song, C., De Cristofaro, E. and Shmatikov, V. (2019) 'Exploiting unintended feature leakage in collaborative learning', *2019 IEEE Symposium on Security and Privacy (SP)*, San Francisco, pp. 691-706. https://doi.org/10.1109/SP.2019.00029

Mishra, Prabhaker, Singh, U., Pandey, C. M., Mishra, P. and Pandey, G. (2019) 'Application of student's t-test, analysis of variance, and covariance', *Annals of Cardiac Anaesthesia*, Vol. 22, No. 4, pp. 407. https://doi.org/10.4103/aca.ACA_94_19

Mishra, P. and Mehta, R. (2017) 'What we educators get wrong about 21st-century learning: Results of a survey', *Journal of Digital Learning in Teacher Education*, Vol. 33, No. 1, pp. 6-19. https://doi.org/10.1080/21532974.2016.1242392

Ożadowicz, A. (2020) 'Modified blended learning in engineering higher education during the COVID-19 lockdown—Building automation courses case study', *Education Sciences*, Vol. 10, No. 10, pp. 292. https://doi.org/10.3390/educsci10100292

Rahman, H., Roy, S. B., Thirumuruganathan, S., Amer-Yahia, S. and Das, G. (2019) 'Optimized group formation for solving collaborative tasks', *The VLDB Journal*, Vol. 28, No. 1, pp. 1-23. https://doi.org/10.1007/s00778-018-0516-7

Rashidovna, M. F. (2020) 'Educational laboratory as a modern form of educational activity organization', *International Scientific Review*, no. LXXII, pp. 41-42.

Razali, N. M. and Wah, Y. B. (2011) 'Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests', *Journal of Statistical Modeling and Analytics*, vol. 2, no. 1, pp. 21-33.

Rochman, C., Nasrudin, D. and Kariadinata, R. (2017) 'Authentic Assessment Based on Teaching and Learning Trajectory', 2nd International Conference on Sociology Education, Bandung, pp. 174-177. https://doi.org/10.5220/0007094901740177

Rozo, L., Calinon, S., Caldwell, D. G., Jimenez, P. and Torras, C. (2016) 'Learning physical collaborative robot behaviors from human demonstrations', *IEEE Transactions on Robotics*, Vol. 32, No. 3, pp. 513-527. https://doi.org/10.1109/TRO.2016.2540623

Sailer, M. and Homner, L. (2020) 'The gamification of learning: A meta-analysis', *Educational Psychology Review*, Vol. 32, No. 1, pp. 77-112. https://doi.org/10.1007/s10648-019-09498-w

Schneider, B. and Pea, R. (2016) 'Real-time mutual gaze perception enhances collaborative learning and collaboration quality', in Orey, M. and Branch, R. M. (ed.) *Educational media and technology yearbook*, New York: Springer. https://doi.org/10.1007/978-3-319-45001-8_7

Strijbos, J. W. (2016) *Handbook of Human and Social Conditions in Assessment*, New York: Routledge

Van Laar, E., Van Deursen, A. J. A. M., Van Dijk, J. A. G. M. and De Haan, J. (2017) 'The relation between 21st-century skills and digital skills: A systematic literature review', *Computers in Human Behavior*, vol. 72, pp. 577-588. https://doi.org/10.1016/j.chb.2017.03.010

Widana, I. W., Parwata, I. and Sukendra, I. K. (2018) 'Higher order thinking skills assessment towards critical thinking on mathematics lesson', *International Journal of Social Sciences and Humanities*, Vol. 2, No. 1, pp. 24-32. https://doi.org/10.29332/ijssh.v2n1.74

Williamson, B., Eynon, R. and Potter, J. (2020) 'Pandemic politics, pedagogies and practices: digital technologies and distance education during the coronavirus emergency', *Learning, Media and Technology*, Vol. 46, No. 2, pp. 107-114. https://doi.org/10.1080/17439884.2020.1761641

Wismath, S. L. and Orr, D. (2015) 'Collaborative Learning in Problem Solving: A Case Study in Metacognitive Learning', *Canadian Journal for the Scholarship of Teaching and Learning*, Vol. 6, No. 3, pp. 10. https://doi.org/10.5206/cjsotl-rcacea.2015.3.10

Zakwandi, R., Yuningsih, E. K. and Setya, W. (2020) 'Implementasi pembelajaran berbasis praktikum pada konsep taraf intensitas bunyi untuk meningkatkan penguasaan konsep Peserta Didik', *Jurnal Penelitian Pembelajaran Fisika*, Vol. 11, No. 1, pp. 75-82. https://doi.org/10.26877/jp2f.v11i1.4522

Zhao, X. and McClure, C. D. (2022) 'Gather. Town: A Gamification Tool to Promote Engagement and Establish Online Learning Communities for Language Learners', *RELC Journal*, pp. 00336882221097216. https://doi.org/10.1177/00336882221097216

INDIVIDUAL INTEREST OF STUDENTS IN PHYSICAL EDUCATION AND SCHOOL ENGAGEMENT IN FOSTERING PHYSICAL CULTURE INSIDE THE CAMPUS: THE CASE OF TWO PROMINENT LOCAL COLLEGES IN PAMPANGA, PHILIPPINES

Joseph Lobo[1]✉
Genesis Dimalanta[2]

[1]Bulacan State University, Philippines

[2]Mabalacat City College, Philippines

✉ jtldlobo@gmail.com

ABSTRACT

Interest has been globally established as a powerful predictor that triggers engagement, especially in fostering physical culture among students. Scholarly works in PE have already shown that situational interest triggers individual interest, which can lead to highly engaged students. However, little is known about the impact of individual interest alone on students' school engagement. In this regard, this study is focused on determining the significant difference between the sex and institution of students concerning individual interest and school engagement. Moreover, it aimed to assess the relationship and direct influence of individual interest on school engagement. After obtaining data from 1659 samples of undergraduate students from the most prominent local colleges in Pampanga, it was revealed that there is no significant difference concerning individual interest and school engagement between sex and institution. Fascinatingly, a significant association between students' individual interests and school engagement was observed. Lastly, individual interest, along with its three factors, predicts school engagement. It can be concluded that individual interest has a direct positive influence on school engagement. Recommendations for future research direction and PE teachers to fully translate physical culture to students to improve their overall well-being are hereby presented.

KEYWORDS

Individual interest, local colleges, physical culture, physical education, school engagement

HOW TO CITE

Lobo J., Dimalanta G. (2024) 'Individual Interest of students in Physical Education and School Engagement in fostering Physical Culture inside the campus: The case of two prominent local colleges in Pampanga, Philippines', *Journal on Efficiency and Responsibility in Education and Science*, vol. 17, no. 1, pp. 78-90. http://dx.doi.org/10.7160/eriesj.2024.170107

Article history
Received
November 23, 2022
Received in revised form
June 2, 2023
Accepted
October 9, 2023
Available on-line
March 31, 2024

Highlights

- *The level of individual interest and school engagement has no significant difference with respect to sex and institution/college in the case of the two local colleges.*
- *There is a significant relationship and direct influence of individual interest on school engagement in Physical Education to foster Physical culture among students.*
- *Individual interest and school engagement are two important variables to foster Physical culture inside the campus.*

INTRODUCTION

Promoting Physical Culture inside the campus

Instilling physical culture inside the campus and in students will only be possible with the promotion and conservation of the culture through Physical Education (PE) and those significant people behind it, the teachers (Gądek, 2020; Madejski, Jaros and Madejski, 2019). In this regard, to fully translate physical culture, students' interest in the content and engagement are vital and need to be considered. Furthermore, to boost the participation of the students to be physically active, the course content and exercises should

be carefully and adequately selected to provide satisfaction and meet their expectations (Teixeira et al., 2012). For a more in-depth explanation, Physical Culture studies various domains pertaining to human movements, such as physical education through various physical activities (Brown, 2019). It provides an avenue for the students to improve and strengthen their bodies, and prevent such illnesses and conditions (Dróżdż et al., 2022; Iqboljon, 2021; Kokoulina et al., 2021; Lutkovskaya et al., 2021). Indeed, fostering physical culture inside the campus through physical education is equally important as other academic courses as it plays a significant role in students' overall growth (Zhang, 2021). Ergo, fostering physical culture develops physical fitness such as improving motor skills and abilities, sports achievement, ethics, esthetic, and cognitive development (Fang, Teng and Wang, 2021; Sierra-Díaz et al., 2019; Tainio, 2019; Yalgashevich, Shermamatovich and Zayniddinovich, 2021).

Fascinatingly, many positive findings were discovered based on previous studies concerning students' attitudes toward physical education, leading to activated interest and engagement. According to Viva and Limbo (2021) and Li et al. (2014), undergraduate students from the University of Eastern Philippines and four Chinese universities reported a moderate to positive attitude toward Physical Education. The result is harmonized with the hard work of teachers in presenting the course properly and efficiently to students. Additionally, Mohamed and Kamil (2020) unraveled that there is a positive and high level of attitude toward implementing Physical Education among secondary school students. The data displayed that respondents are comfortable teaching today's subject. Furthermore, it was found that the professional qualification of teachers is associated with students' attitudes toward the subject (Carcamo-Oyarzun et al., 2022; Sucuoglu and Atamturk, 2020; Zalech, 2021). On the other hand, the findings differ from what most people expected after examining other published scholarly papers. Based on the study of Solomonko et al. (2022), it was revealed that traditional physical culture is not highly important to be considered based on the survey conducted on college students taking a law degree. The findings of Mutlu et al. (2021) also displayed a negative attitude toward Physical Education among third-graders from the Menteşe district of Muğla province in Turkey. A study by Iconomescu et al. (2018) also uncovered a negative attitude from Romanian students toward the subject because of poor communication with teachers. In line with the aforementioned findings, studies have confirmed that teacher factors may also affect student attitudes toward Physical Education (González-Peño, Franco and Coterón, 2021; Malinauskas, Emeljanovas and Valantine, 2018; Scrabis-Fletcher and Silvermanv, 2010). Additionally, physical and sports activities were also identified as predictors of students' attitudes (Lobo, Dimalanta and Bautista, 2022; Sağın, 2022; Tagare and Villaluz, 2021). Hence, it can be concluded that physical education teachers are the people behind solidifying students' interest and engagement in the subject through carefully selected activities promoting physical culture.

Individual Interest in Physical Education and School Engagement

Over the past years, interest has strongly influenced in-school engagement (Chen and Wang, 2017). It has been intellectualized as a motivational process that bolsters learning and is crucial to students' academic success (Harackiewicz, Smith and Priniski, 2016). A student can be identified as interested in the content if (s)he is determined to hone the skills that were just newly introduced, play sports assiduously, or engage vigorously in an activity. As Renninger and Hidi (2022) recently reiterated, interest is a motivator that triggers all human activities. Furthermore, various empirical studies were already executed focusing on the outcome of situational interest and school engagement of students in Physical Education (Allard-Latour, Rannou and Kermarrec, 2022; Roure et al., 2019; Roure and Pasco, 2018a; Wang, Shen and Bo, 2022); such as physical activities (Pasco and Roure, 2022). This is because educational researchers have argued that situational interest is far substantial in terms of its potential in motivating students when compared to individual interest, especially in a daily pedagogical setting. Moreover, Physical Education instructors can easily manipulate learning tasks targeting situational interest (Roure and Pasco, 2018b). In an in-depth elucidation, interest has been described as a multifaceted construct with three distinct features according to Roure, Lentillon-Kaestner and Pasco (2021): a mental state contrasted with a steady attribute; specific on the content; and a structure with multiple dimensions. In essence, situational interest is the individual's increased attention on a concept, disposing oneself to learn mixed with a constructive solicitude about it (Fastrich and Murayama, 2020; Schmidt and Rotgans, 2021; Wong et al., 2020). On the other hand, individual interest is defined as propensity to reacquaint oneself vis-à-vis to specific content over again (Hong et al., 2019; Knekta et al., 2020; Quinlan and Renninger, 2022; Shin and Kim, 2019). Moreover, interest is always intellectualized as content-specific (Kang, Keinonen and Salonen, 2021; Wild, 2022), which can be postulated an individual may have a keen interest towards Physical Education, but not with other courses or contrariwise. Last of all, interest has been contextualized as a structure with multiple dimensions which encompasses two domains: affective (e.g., excitement, positive feeling) and cognitive (e.g., perceived importance, usefulness) (Aslan et al., 2021; Rowland et al., 2019; Svenningsson et al., 2022).

Meanwhile, according to Pasco and Roure (2022), individual interest encompasses three distinct factors namely Positive affect and willingness to reengage (PAWR), Stored Utility Value (SUV) and Stored Attainment Value and Knowledge-seeking Intentions (SAVKSI). For a more in-depth explanation, PAWR refers to a student's positive state of pleasure or exhilaration when interacting with certain content, such as Physical Education. It can be posited that students will continue to reconnect themselves when they have a positive experience in the course (O'Keefe and Linnenbrink-Garcia, 2014; Renninger and Hidi, 2022). SUV is usually determined by how a specific bustle or concept can be related to the current and eminent goals of a student. To illustrate, besides from the physical benefits of Physical Education, when students perceived that it

may bring forth health and lifelong benefits, their interest may be triggered and would consider it valuable. Lastly, SAVKSI are demarcated as the significance of a content on a personal level vis-à-vis the relevant aspects to a student's concept of self and personal desire. As an emphasis, self-schema is derived from a cognitive theory which describes presumptions about one self and capacities (Hovelius et al., 2021) and is defined as multiple representations in memory of the self (Scott et al., 2022). Moreover, self-schema are classes of knowledge that echo in what manner a person expects to reflect and act in a specific setting or situation.

Furthermore, the method to envisage academic achievement that became a significant concept and piqued the interest of educational researchers is called School Engagement (Liu et al., 2021). Students display a degree of attention, inquisitiveness, interest, buoyancy, and desire when being imparted vis-à-vis a concept, such as various physical activities in Physical Education. Students' engagement in school extends to the impetus they have to acquire and advance in their education (Charkhabi et al., 2019). In addition, this concept has been considered to encompass various domains encompassing cognitive (e.g., investment in mastering tasks, flexible problem-solving), affective (e.g., sense of identification, positive outlook toward school) and behavioral domains (e.g., students' participation in schooling, positive practices in schooling) (Benito Mundet et al., 2021). Based from these three domains, school engagement has three distinct features: Dedication (DE), Absorption (ABS), and Vigor (VI) (Jaya and Ariyanto, 2021). Firstly, dedication refers to an individual's engrossment in various academic responsibilities with high-sense of commitment and enthusiasm (Listau, Christensen and Innstrand, 2017; Teuber, Nussbeck and Wild, 2021). It has been described by Widlund, Tuominen and Korhonen (2021) as a student with a highly constructive cognitive attitude toward school and extremely devoted to the learning procedures and outcomes. Moreover, an intensive student occupied in various academic endeavors is called absorption (Koob et al., 2021). It can be implied that possessing a very high feeling of competence toward studying the content is described by this domain. Lastly, a student pigeonholed by unmeasurable energy, resiliency, exuberance, and flexibility in the pursuit of excellence is referred to as vigor (Demirbatır, 2020; Pulido-Martos et al., 2020). In other words, amidst the difficulties students face due to various academic tasks, (s)he still displays a positive attitude toward these challenges. These three facets of school engagement are distinctly different but highly interconnected (Estévez et al., 2021). Previously conducted studies have accentuated the interrelationship of situational interest and school engagement (Hui et al., 2019; Mallari and Tayag, 2022; Park and Han, 2021; Upadyaya et al., 2021). It can be postulated that the situational interests of students indeed have a significant role in their school engagement.

Moreover, issues of inequalities regarding sex have been predominantly evident worldwide, especially when dealing with the interest of students in Physical Education activities being introduced to them (e.g., physical activities and sports events). Grounded on the findings of various scholars, higher interest and positive perception scores toward Physical Education are predominantly evident for males compared to their counterparts (Perić et al., 2020; Sofi, Waseem and Padder, 2019). Additionally, findings exposed that male students are highly more interested and engaged in organized contact sports (e.g., basketball), while female students are more on into individual sports, dancing, and exercise (Cowley et al., 2021; Cruz, 2022; Deaner, Balish and Lombardo, 2016; Pituk and Cagas, 2019; Ricardo et al., 2022). Indeed, it can be postulated that the issue has never been solved for the past years, and addressing these problems should be targeted bull's-eyes, especially fostering higher engagement and participation from female students in various activities of Physical Education in Higher Education. Additionally, no studies were not yet performed analyzing the difference on the individual interest and school engagement of students from different higher education institutions in the Philippines, and even in a global context. Surprisingly, a single study was performed by Lobo (2023) focusing on the difference in terms of individual interest of students toward Philippine Traditional Dances and school engagement, which revealed that there is no significant difference was observed in relation to the individual interest of students from Pampanga State Agricultural University, Mabalacat City College, and City College of Angeles. On the one hand, a significant variance was observed concerning students' school engagement to among these schools to which PSAU respondents are highly engaged compared to MCC and CCA. However, this particular study is not related to Physical Education. Hence, conducting a similar study in the context of Physical Education in the local colleges sector is highly recommended.

As mentioned earlier, situational interest causes the amplification of individual interest leading to higher students' school engagement (Hong et al., 2019; Palmer, 2019) and vice-versa (Kahu, Nelson and Picton, 2017; Romine et al., 2020; Rotgans and Schmidt, 2018). The aforementioned studies have claimed that students' situational interest could strengthen individual interest, leading students to engage and enjoy the content. In other words, the current environment accounts for a higher situational interest vis-à-vis their individual interest. However, more established information about how individual interest may affect engagement is needed. This paper agrees with what Roure et. al. (2021) have stated: numerous documentations were already accrued concerning situational interest and its motivational roles, but studies focusing on individual interest are still unexplored. Surprisingly, there have been few studies on these variables—for example, the study of Lobo (2023) have revealed that individual interest is highly related to school engagement in order to prepare future educators of Physical Education in the promotion of intangible cultural heritage of the Philippines. Additionally, findings reported that individual interest in Physical Education is significantly related to and triggers students' school engagement (Bautista, De Dios and Lobo, 2023). However, these studies are not on the context of local college for the promotion of physical culture inside the campus. After humungous efforts dedicated to looking for published scholarly papers concerning this present investigation, it can be beseeched that conducting a study focusing on the association and direct effect of individual interest in students' school engagement is highly needed. This

is due to the insufficiency of works conducted and published, especially in the context of Physical Education, due to its lesser importance to educational researchers.

Aims and hypotheses

This present study is focused on the following objectives:

1. Determine the significant difference between sex and institution concerning individual interest and school engagement;
2. Assess the significant association of individual interest to school engagement;
3. Evaluate the direct influence of individual interest to school engagement; and
4. Examine the direct influence of the three factors of individual interest to school engagement.

Aside from determining the significant difference of sex and institution/college relating to individual interest and school engagement, this investigation is highly focused on the association of the two variables and the effect of the independent variable/s (individual interest-PAWR, SUV, SAVKSI) on the dependent variable (SE). Since the topic is relatively new and there were only few studies that were known to be conducted concerning these two variables, therefore, this present study tested the following hypotheses (Figures 1 and 2 illustrates the conceptual framework for the study):

H_1 II has no significant relationship to SE
H_2 II has a no direct influence to SE
H_3 PAWR has no significant impact to SE
H_4 SUV has no significant effect to SE
H_5 SAVKSI has no significant influence to SE

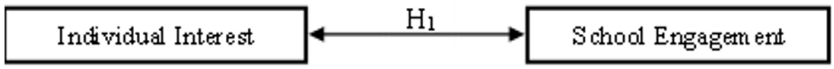

Figure 1: Conceptual Framework for correlation analysis

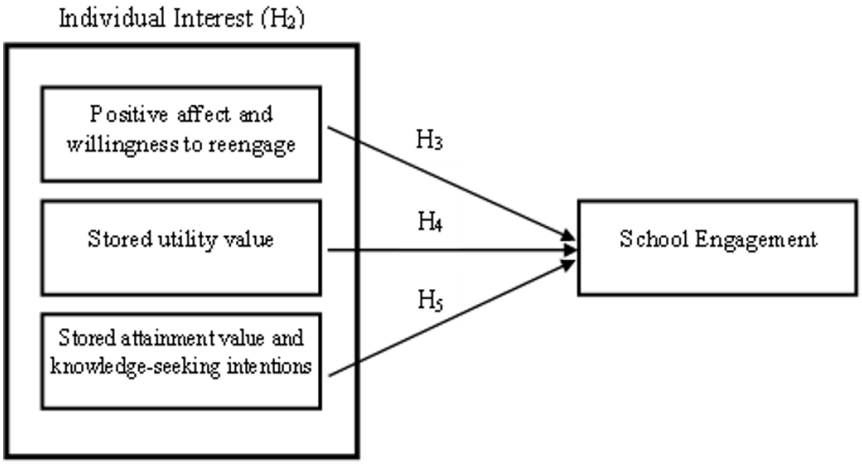

Figure 2: Conceptual Framework for multiple regression analysis

METHODS AND MATERIALS

Participants and Sampling Technique

Participating respondents comprise of first- and second-year undergraduate students currently taking two minor Physical Education courses (Physical Education 1 [PATH-Fit1] and Physical Education 3 [PATH-Fit3]) during the first semester, A.Y. 2022-2023 from the two prominent local colleges in Pampanga, City College of Angeles and Mabalacat City College. These local colleges are under the governance of two different local government units (LGU), which are situated in the City of Angeles and Mabalacat, respectively. Both local colleges' overall student population ranges from approximately 5,000-6,000 across various degree programs. Furthermore, two (2) sampling procedures were performed to identify the respondents for the investigation. The *Purposive Sampling Technique* was performed first due to the specificity of respondents needed for the study. It is a non-parametric technique that allows researchers to identify participants based on the qualities needed for the investigation; and for this study, undergraduate students enrolled in PE1 and PE3. After identifying the needed respondents, the *Convenience Sampling Technique* was also employed. It is a procedure where gathering of data from the respondents is based on their own convenience (Frey, 2018). Specifically, obtainability of data is vis-à-vis the topographical propinquity that may also include other variants of accessibility. The respondents for the study are highly suitable for this type of research because minor courses in physical education are only offered during the first- and second-year level for both colleges. Hence, promoting physical culture through these courses can be made possible. The demographic characteristics of the respondents are displayed in Table 1. Overall, 1659 undergraduate students volunteered to answer the online survey and accepted for scrutiny. The findings which are shown in Table 1 illustrates that most students who answered are females (N = 1036), in lieu of 62.4% of the overall sample population, while males (N = 623) are 37.6%. Lastly, most of the respondents are from CCA (N = 1072) denoting 64.6%, while MCC (N = 587) is 35.4%.

Item		*f*	Percentage
Sex	Male	623	37.6%
	Female	1036	62.4%
College/Institution	City College of Angeles (CCA)	1072	64.6%
	Mabalacat City College (MCC)	587	35.4%
	Total	1659	100.0%

Table 1: Demographic profile

Instruments and data gathering

Obtaining of data was performed via online survey through the use of Google Forms. Surveying online has a great potential to amass massive amount of data efficiently, cost-effective, and within relatively short time frames (Li, Shamsuddin and Braga, 2021; Regmi et al., 2017). There are two instruments adopted and utilized to gather data from the respondents. The *Students' Individual Interest in Physical Education questionnaire* newly developed by Roure et al. (2021), a 14-item multidimensional questionnaire that measures students' individual interest based on three (3) domains namely: positive affect and willingness to reengage [PAWR], stored utility value [SUV], and stored attainment value and knowledge-seeking intentions [SAVKSI]. It is a 5-point Likert Scale where responses are recorded as 1- "strongly disagree" to 5- "strongly agree." the reliability value (Cronbach's Alpha) is .95 for the overall scale, and for its subscales is between .87 and .94. Lastly, the *Utrecht Work Engagement Scale for Students* (UWES-9S) adapted by Carmona-Halty et al. (2019) is a 5-point Likert Scale was used to measure students' overall engagement in school. This particular instrument is segmented into three parts: vigor [VI], dedication (DE) and absorption (ABS). Responses are encoded from 1- "never" to 5- "always." the reliability value (Cronbach's Alpha) of the scale is .93, ranging from .79 to .90 for its subscales.

Data analysis

A test of normality, reliability test, and bivariate correlation for inter-variable relationship were performed. Table 2 typifies the results from the normality test across various subscales. Based on the findings, the skewness and kurtosis values did not obtain the threshold value [-2, 2] across all subscales (Romano et al., 2021). Hence, it can be construed that the distribution of data is non-normal. In relation to this, a distribution-free test is applicable to examine the variance between groups concerning individual interest and school engagement, and the interrelatedness of the two constructs. Moreover, the table explicates the results from the reliability test of each subscale of personal interest and school engagement. Grounded on the findings, it can be postulated that all subscales are highly reliable with Cronbach's Alpha value between .79 to .94. Lastly, the bivariate correlations for each subscale are also exhibited in the table which displayed a significant relationship across all variables ($p < .01$).

Moreover, Mann-Whitney *U* Test was presumed that can be used examine the variance between groups (i.e., sex and institution/college) concerning individual interest and school engagement. This specific non-parametric test is focused on comparing the means of the two independent groups with the assumption that the distribution of data are non-normal (Sundjaja, Shrestha and Krishan, 2022). To further determine if the specified statistical analysis may be used, a non-parametric version of the Levene's test of Homogeneity of Variances was performed, and *p*-values should be >.05 to test the assumption of Mann-Whitney *U*. Based on the findings shown in Table 3, only the Individual Interest [institution/college] ($p > .05$) did not violate the assumption. Therefore, the Mann Whitney *U* test can be used for this specific variable. On the other hand, Individual interest ($p < .05$) and both School engagement [sex] ($p < .05$) [college] ($p < .05$); therefore, significant; indicating that the assumption was violated. Instead, the *Independent Samples T-Test* may be used, since the sample size for this study is significantly large.

Furthermore, to examine the correlation of individual interest to school engagement, Spearman's *Rho* (r_s) assessment was used. It is a non-parametric measure of correlation using ranks (Akoglu, 2018). In this specific assessment, the study focused on determining the association of the two variables regardless of their latent constructs. Hence, obtaining a global score through computing for their composted score was used. Lastly, Multiple Regression analysis was utilized predicting the association and direct influence of individual interest to school engagement. This specific modeling technique can be performed assessing two or more variables predicting one dependent variable (Kang and Zhao, 2020). For this specific analysis, the three factors or latent constructs will be used as a whole model to predict its direct influence to school engagement, and to which these latent constructs will be analyzed individually (as an independent variable) right after.

	M ± SD	Skewness	Kurtosis	1	2	3	4	5	6
PAWR	3.58 ±.85	-3.67	-1.33	(.91)					
SUV	3.65 ±.85	-4.21	-2.23	.74**	(.87)				
SAVKSI	4.08 ±.86	-16.57	-7.40	.68**	70**	(.94)			
VI	3.59 ±.77	-1.37	-0.3	.57**	63**	.61**	(.79)		
DE	3.87 ±.82	-9.7	0.61	.53**	.57**	.81**	.81**	(.84)	
AB	3.61 ±.83	-0.89	-2.01	.48**	54**	.74**	.77**	.77**	(.89)

** Statistically significant at p < .05.*

*** Statistically significant at p < .01.*

Table 2: Descriptive statistics, normality estimates, internal consistency coefficients, and bivariate correlations

	Levene Statistic	df1	df2	Sig.
Individual Interest (sex)	18.950	1	1647.166	.000
School Engagement (sex)	14.622	1	1636.415	.000
Individual Interest (institution/college)	1.839	1	1643.387	.175
School Engagement (institution/college)	9.960	1	1640.986	.002

Table 3: Non-parametric version of Levene's test of Homogeneity of Variances results

Ethical Considerations

The respondents were required to provide their consent by agreeing to the statement attached in the Google Forms. Additionally, respondents were advised about the objectives of the study, its instruments, and the variables being measured for the study. Minor risks answering the online survey were also enumerated. Respondents are free to withdraw or to ask for debriefing anytime.

RESULTS AND DISCUSSION

Table 4 displays the results from the *Independent Samples T-Test* analysis between individual interest and school engagement being independent of gender and school engagement being independent of institution/college. According to Levene's test for equality of variances, the Sig. Value of individual interest (sex) and both school engagement (sex and institution) is < .05, which means that the assumption of equal variances has not been violated. Based on the findings, no observation of a significant difference concerning individual interest (sex) for both male (3.78 ± .83) and female (3.79 ± .72) students [$t(1657) = -.412$, $p = .681$]. Also, there is a significant difference concerning school engagement (sex) for both male (3.66 ± .80) and female (3.70 ± .70) students [$t(1657) = -1.206$, $p = .228$]. The findings posited that regardless of gender, undergraduate students have a higher individual interest in Physical Education and school engagement. Lastly, no observation of a significant difference concerning school engagement (institution/college) between City College of Angeles (3.68 ± .78) and Mabalacat City College (3.70 ± .69), [$t(1657) = -.218$, $p = .827$]. Therefore, the results can be postulated that regardless of institution/college, all undergraduate students from these two local colleges are highly engaged in their schooling. Lastly, after performing the Mann-Whitney *U* Test, there is no observation of a significant difference concerning individual interest between college/institution [City College of Angeles (*Mdn* = 3.86) and Mabalacat City College (*Mdn* = 3.86), $U(N_{\text{City College of Angeles}} = 1072, N_{\text{Mabalacat City College}} = 587) = 313987.000$, $z = -.069$ $p = .945$, $r = -0.002$)]. The findings can be construed that regardless of college/institution, all undergraduate students are highly interested in Physical Education.

Based on the findings, it can be postulated that there is no significant variance in the level of individual interest of students in Physical Education and their school engagement in terms of gender and institution. Regardless of sex and institution, male and female undergraduate students from City College of Angeles and Mabalacat City College are all interested in Physical Education and highly engaged in their schooling. These present findings are dissimilar from previous studies from other scholars, which uncovered that male students are highly interested in Physical Education, give more importance to the content, and have a more positive attitude towards the subject compared to females (Perić et al., 2020; Sofi et al., 2019). Contrastingly, based on other published scholarly papers, it was found that conceptual knowledge in Physical Education is higher for female students compared to males (Thomas and Ti, 2021). However, these findings are not relatively specific on either situational or individual interest, but in general. Furthermore, it was found that male students are highly engaged in school concerning Physical Education compared to their counterparts (Hands and Parker, 2016), especially concerning Physical activities and sports participation (Delextrat et al., 2020; Lauderdale et al., 2015). It has been prevalent across studies in Physical Education worldwide concerning the difference in terms of interest and engagement when sex is being taken into consideration. In terms of physical activities and sports (Hsu et al., 2022), male students are more interested and highly engaged to contact sports, while female are more into individual sports, dancing or exercising to music (Frömel et al., 2022; Peral-Suárez et al., 2020; Resaland et al., 2019). The findings in terms of significant variance concerning school engagement was supported by the study of Lobo (2023), to which PSAU students are more engaged compared to MCC and CCA students. However, this study is not related to Physical Education and in a local college

context. In line with this, performing a similar comparative study is highly advised to examine if individual interest and school engagement in Physical Education varies from school to school.

	N	M ± SD	SE	df	*t*-test	Sig.	Decision
Individual Interest (sex)							
Male	623	3.78 ±.83	.033	1657	-.412	.681	Not significant
Female	1036	3.79 ±.72	.022				
School Engagement (sex)							
Male	623	3.66 ±.80	.032	1657	-1.206	.228	Not significant
Female	1036	3.70 ±.70	.021				
School Engagement (institution)							
City College of Angeles	1072	3.68 ±.78	.024	1657	-.218	.827	Not significant
Mabalacat City College	587	3.70 ±.69	.028				

Table 4: Independent Samples T-Test results

Table 5 unravels the findings after performing Spearman's *rho* (r_s) analysis. The results exhibited a positive association between overall individual interest and school engagement [$r(1657) = .67$, $p < .05$)]. The findings posited that as the personal interest of students in Physical Education increases, their engagement in school is also amplified. As what other scholars have stated, the more a student is highly interested in the content, the more a student will engage to school (Harackiewicz et al., 2016; Murayama, 2022; Rotgans and Schmidt, 2017). Based on the result of the correlational analysis, H_1 has been rejected. As mentioned earlier, there were only few studies conducted concerning the association between the two variables (Bautista et al., 2023; Lobo, 2023). Hence, this further suggests conducting a similar study to deepen the linkage between the two variables, or refute the findings.

			Personal Interest	School Engagement
Spearman's rho	Personal Interest	Correlation Coefficient	1.000	.664**
		Sig. (2-tailed)	.	< .001
		N	1659	1659
	School Engagement	Correlation Coefficient	.664**	1.000
		Sig. (2-tailed)	< .001	.
		N	1659	1659

*** Correlation is significant at the 0.01 level (2-tailed).*

Table 5: Spearman's Rho (r_s) test results

The next step, the dependent variable was regressed on the latent constructs that represent students' overall individual interest. After performing the analysis, it was discovered that the three independent variables predict school engagement [F (3, 1655) = 501.015, $p < .001$], which uncovered that the three factors have a significant impact to students' overall school engagement. Additionally, the $R_2 = .476$ displayed that the model explains 47.6% of the variance in school engagement. The result can be postulated that individual interest itself has a direct impact to students' school engagement which reverberated the previous findings of other scholars (Bautista et al., 2023; Lobo, 2023). Studies pertaining to individual interest and its relationship to school engagement in Physical Education are still inadequate, therefore, conducting a thorough investigation of these variables is highly suggested. Based from the findings above, H_2 has been rejected.

Also, coefficients were further examined to ascertain the influence of each of the individual factors on the dependent variable. The third hypothesis evaluated the impact of positive affect and willingness to reengage (PAWR) to school engagement (SE). The result yielded that PAWR predicts SE, which can be posited that PAWR has a significant impact to SE ($\beta = .097$, $t = 3.984$, $p < .001$). PAWR has been referred to as a student's attached pleasure to specific content, such as Physical Education. Based on the findings, students may reengage themselves continuously towards Physical Education as long as they perceive the content as fun and exciting. Parallel to other scholars' findings, it was unraveled that enjoyment and excitement were feelings that had the highest predictive

power in the engagement of students (Fierro-Suero et al., 2022; Leisterer & Jekauc, 2019; Powell & Ceaser-White, 2017). It can be postulated that positive emotional experiences in Physical Education may undoubtedly lead to a higher level of school engagement (Jaakkola et al., 2015; Roure, Méard, et al., 2019; Wang et al., 2010). In summary, H_3 has been rejected.

Furthermore, the fourth hypothesis evaluated the impact of stored utility value (SUV) to school engagement (SE). It was discovered that SUV predicts SE, which indicated that SUV has a significant effect to SE (β = .246, t = 9.713, p < .001). As defined earlier, SUV is how a specific content, such as Physical Education, relates to the student's present and future goals. In line with the results, it was found that SUV directly impacts school engagement. It can be construed that when students perceive that Physical Education is highly beneficial for their present and future attainments, their school engagement is bolstered. For example, the students with a higher individual interest significantly perceive the positive benefits of the content; there is a higher possibility of school engagement, which may also result in healthy activity engagement outside of school. Synonymous with what various scholars have mentioned, students' approval on physical activities throughout the process bolsters engagement, adherence to healthy régime habits in the future and to carry on with the routine outside the four walls of the campus (Curran and Standage, 2017; Polet et al., 2019; Rojo-Ramos et al., 2022). The items for this particular factor are highly related to the intention for future participation of students to various physical activities as mentioned to other studies (Gao and Xiang, 2008; Yli-Piipari et al., 2013). In this, H_4 has been rejected.

Lastly, the fifth hypothesis evaluated the impact of stored attainment value and knowledge-seeking attentions (SAVKSI) on school engagement (SE). The findings displayed that SAVKSI predicts SE, which can be postulated that SAVKSI has a substantial impact on SE (β = .319, t = 13.887, p < .001), hypothesizing that H_5 has been rejected. As mentioned earlier, SAVKSI refers to a student's perceived importance of content, such as Physical Education, on a personal level which may likely result in deepening of knowledge on a specific content that is highly relevant to their perception of self-schema and their core personal goals. Based on the findings, it can be deliberated that as long as students see Physical Education reflects their perception of self-schema and central personal goals can upsurge higher engagement. Tantamount to other published scholarly works, attainment value based on the expectancy-value theory (EVT) predicts motivation (Ding, Sun and Chen, 2013; Wang and Xue, 2022), which further leads to students' rendezvous in Physical Education (Shang, Moss and Chen, 2022; Zhu and Chen, 2013). In conclusion, the findings speculated that individual interest (II) has a direct and significant influence on school engagement (SE), and all the latent constructs of individual interest (II) are accountable for the concerned relationship. Results are illustrated in Table 6.

Based on the findings, this study further proposes conducting a comprehensive examination of the direct influence of individual interest on school engagement by taking each feature of school engagement separately. This is in line with what Widlund et al. (2021) and Salmela-Aro and Upadyaya (2017) have suggested, the three domains of school engagement may be best described as a global construct highly applicable for younger pupils in the primary and basic education; however, for young adults such as college students, these constructs may seem become separated. Concerning the three unique features of both variables being studied, the role of physical educators should be emphasized. This study accentuates that inculcating the real "value" of physical education to students is highly desirable as this will help students to understand the significance of the content in a deeper and personal level. Although teachers may not directly address the underlying external barriers to students' engagement, developing teachers' instructional techniques and self-competence may assist them maintaining a strong, positive, learning-focused and relevant school or classroom climate (Powell and Ceaser-White, 2017).

Hypothesis	Regression weights	Beta Coefficient	R^2	F	t	p	Decision
H_2	II → SE	.661	.476	501.015	-	< .001	Rejected
H_3	PAWR → SE	.097	-	-	3.984	< .001	Rejected
H_4	SUV → SE	.246	-	-	9.713	< .001	Rejected
H_5	SAVKSI → SE	.319	-	-	13.887	< .001	Rejected
R^2	.476						
F (3, 1655)	501.015						

**Sig. value p <.05. II- Individual Interest, SE- School Engagement, PAWR- Positive affect and willingness to reengage, SUV- Stored utility value, SAVKSI- Stored attainment value and knowledge-seeking attentions.*

Table 6: Multiple regression results and Hypotheses testing

CONCLUSION

New insights have been drawn from the results, which indicated that individual interest alone plays a significant role in students' school engagement without controlling the environment to amplify situational interest. To fully translate physical culture efficiently inside the campus, examining students' individual interest in Physical Education through various physical activities that lead to active and engaged learners is imperative. Moreover, this will provide a complete picture of how these constructs may strengthen active involvement in various physical movements separate from the four walls of the campus, which is well-known to be highly beneficial for their well-being

while elevating the quality of their lives. This study suggests the following proposals based on the three distinct features of individual interest. First, Physical Education teachers should consider selecting various physical activities that are highly perceived by students as enjoyable and exciting to be learned, which applies to both sexes. As mentioned in the discussion earlier, the emotional experience provided by the content can lead to a higher level of student engagement. Lastly, Physical Education teachers must consider inculcating world-related values in all the physical activities imparted to increase students' level of appreciation, which they can consider beneficial for their current and future aims. In line with the following proposals, from a professional development standpoint, this study recommends that the higher administration consider providing in-depth and extensive training to all teachers to address the needs and boost individual interest to increase school engagement. Policymakers and practitioners should provide such activities but are not limited to: coming up with an educated decision on the various physical activities that will be taught to students that are highly enjoyable, exciting, valuable, and related to the personal lives of students; and pieces of training that are highly concerned to various practical teaching strategies and techniques that are highly effective which may pique the interest of students toward Physical Education, foster life-long learning, and inculcate physical culture in the most efficient yet enjoyable way possible. Additionally, teachers' engagement with students may develop their interest and engagement in physical education over time. The following proposals align with the derived conclusion based on the earlier findings.

Furthermore, adding other exogenous and predicting variables not included in the study may be tested to deepen the understanding of what other features, aside from situational interest, may affect students' individual interests and school engagement. Also, this study suggests that future investigations may consider adopting a multi-informant approach by combining data from physical education teachers, as they may deliver much scholarly information concerning the individual interest of students and their engagement which may provide additional information on the vital role of teachers, deepening the relationship between the two variables and filling-in up the scarcity of investigations related to these. Most importantly, gathering essential data from the teachers will help develop a specific intervention that can improve the promotion and preservation of physical culture on campuses.

This present study is restricted to some limitations that are highly important to take into consideration. First of all, the respondents are controlled samples of undergraduate students from the CCA and MCC, both under the sector of local colleges and universities (LCUs). Consequently, the findings of this investigation cannot take a broad view of the entire studentry, specifically those belonging to other Higher Education Institutions (HEIs) such as institutes from the State Universities and Colleges (SUCs) and even Private Higher Education Institutions (PHEIs) in the Philippines or even from other international academies. In line with this, future scholars may find inquisitiveness in conducting a similar study by amassing reports from the HEIs as mentioned above and examining if the outcomes may support or repudiate the findings of this investigation. Finally, as mentioned by the authors from their original paper behind the newly developed and validated instrument used for this study (Students' Individual Interest in Physical Education questionnaire), the tool is suggested to be used to examine further the relationship and direct effect of students' individual interest to school engagement, which was performed by this present study. However, this study would suggest further examining the instrument by performing factor analyses and scrutinizing its convergent and discriminant validity from other sets of populations to determine if the tool may be used for further investigations.

DECLARATIONS

Conflict of Interest

The contributors hereby certify that there are no conflicts of interest.

Funding

The researchers were not provided with any financial support from any external organizations.

Data and Materials

Access to materials and data will be provided exclusively upon request.

REFERENCES

Akoglu, H. (2018) 'User's guide to correlation coefficients', *Turkish Journal of Emergency Medicine*, Vol. 18, No. 3, pp. 91-93. https://doi.org/10.1016/j.tjem.2018.08.001

Allard-Latour, E., Rannou, J. and Kermarrec, G. (2022) 'Adolescent girls' and boys' situational interest for a learning task in physical education', *Journal of Physical Education and Sport*, Vol. 22, No. 6, pp. 1356-1362. https://doi.org/10.7752/jpes.2022.06170

Aslan, S., Fastrich, G., Donnellan, E., Jones, D. J. W. and Murayama, K. (2021) 'People's naïve belief about curiosity and interest: A qualitative study', *PLOS ONE*, Vol. 16, No. 9, pp. e0256632. https://doi.org/10.1371/journal.pone.0256632

Bautista, C., De Dios, D. Al and Lobo, J. (2023) 'The Nexus between individual interest and school engagement in bolstering Physical Culture for a habitual healthy régime: a case of a state university', *Physical Education of Students*, Vol. 27, No. 1, pp. 24-35. https://doi.org/10.15561/20755279.2023.0104

Benito Mundet, H., Llop Escorihuela, E., Verdaguer Planas, M., Comas Matas, J., Lleonart Sitjar, A., Orts Alis, M., … Rostan Sánchez, C. (2021) 'Multidimensional research on university engagement using a mixed method approach', *Educación XX1*, Vol. 24, No. 2, pp. 65-96. https://doi.org/10.5944/educxx1.28561

Brown, D. (2019) 'Physical Culture', *Societies*, Vol. 9, No. 1, pp. 23. https://doi.org/10.3390/soc9010023

Carcamo-Oyarzun, J., Wydra, G., Hernandez-mosqueira, C., Martinez-salazar, C. and Souza De Carvalho, R. (2022) 'Attitudes toward Physical Education teachers from a cross-cultural perspective: German and Chilean students' viewpoints', *Cultura, Ciencia y Deporte*, Vol. 17, No. 51, pp. 15-21. https://doi.org/10.12800/ccd.v17i51.1614

Carmona-Halty, M. A., Schaufeli, W. B. and Salanova, M. (2019) 'The Utrecht Work Engagement Scale for Students (UWES-9S): Factorial Validity, Reliability, and Measurement Invariance in a Chilean Sample of Undergraduate University Students', *Frontiers in Psychology*, Vol. 10. https://doi.org/10.3389/fpsyg.2019.01017

Charkhabi, M., Khalezov, E., Kotova, T., S Baker, J., Dutheil, F. and Arsalidou, M. (2019) 'School engagement of children in early grades: Psychometric, and gender comparisons', *PLOS ONE*, Vol. 14, No. 11, pp. e0225542. https://doi.org/10.1371/journal.pone.0225542

Chen, A. and Wang, Y. (2017) 'The Role of Interest in Physical Education: A Review of Research Evidence', *Journal of Teaching in Physical Education*, Vol. 36, No. 3, pp. 313-322. https://doi.org/10.1123/jtpe.2017-0033

Cowley, E. S., Watson, P. M., Foweather, L., Belton, S., Thompson, A., Thijssen, D. and Wagenmakers, A. J. M. (2021) '"Girls Aren't Meant to Exercise": Perceived Influences on Physical Activity among Adolescent Girls—The HERizon Project', *Children*, Vol. 8, No. 1, pp. 31. https://doi.org/10.3390/children8010031

Cruz, A. B. (2022) 'Post-primary School Students' Attitudes Toward Physical Education and Physical Activity Preferences: Philippines' K-12 Program', *The Asia-Pacific Education Researcher*, Vol. 31, No. 5, pp. 507-517. https://doi.org/10.1007/s40299-021-00598-2

Curran, T. and Standage, M. (2017) 'Psychological Needs and the Quality of Student Engagement in Physical Education: Teachers as Key Facilitators', *Journal of Teaching in Physical Education*, Vol. 36, No. 3, pp. 262-276. https://doi.org/10.1123/jtpe.2017-0065

Deaner, R. O., Balish, S. M. and Lombardo, M. P. (2016) 'Sex differences in sports interest and motivation: An evolutionary perspective', *Evolutionary Behavioral Sciences*, Vol. 10, No. 2, pp. 73-97. https://doi.org/10.1037/ebs0000049

Delextrat, A., Esser, P., Beale, N., Bozon, F., Eldridge, E., Izadi, H., Johansen-Berg, H., Wheatley, C. and Dawes, H. (2020) 'Effects of gender, activity type, class location and class composition on physical activity levels experienced during physical education classes in British secondary schools: a pilot cross-sectional study', *BMC Public Health*, Vol. 20, No. 1, pp. 1590. https://doi.org/10.1186/s12889-020-09698-y

Demirbatır, R. E. (2020) 'Comparison of burnout, vigor and education satisfaction of music and art majors in department of fine arts education', *International Journal of Evaluation and Research in Education*, Vol. 9, No. 3, pp. 478. https://doi.org/10.11591/ijere.v9i3.20548

Ding, H., Sun, H. and Chen, A. (2013) 'Impact of Expectancy-Value and Situational Interest Motivation Specificity on Physical Education Outcomes', *Journal of Teaching in Physical Education*, Vol. 32, No. 3, pp. 253-269. https://doi.org/10.1123/jtpe.32.3.253

Dróżdż, R., Pasek, M., Zając, M. and Szark-Eckardt, M. (2022) 'Physical Culture and Sports as an Educational Basis of Students' Healthy Physical Activities during and Post-Lockdown COVID-19 Restrictions', *International Journal of Environmental Research and Public Health*, Vol. 19, No. 18, pp. 11663. https://doi.org/10.3390/ijerph191811663

Estévez, I., Rodríguez-Llorente, C., Piñeiro, I., González-Suárez, R. and Valle, A. (2021) 'School Engagement, Academic Achievement, and Self-Regulated Learning', *Sustainability*, Vol. 13, No. 6, pp. 3011. https://doi.org/10.3390/su13063011

Fang, J.-D. D., Teng, P.-C. and Wang, F.-J. (2021) 'The Impact of Physical Education Classes on Health and Quality of Life during the COVID-19', *Applied Sciences*, Vol. 11, No. 19, pp. 8813. https://doi.org/10.3390/app11198813

Fastrich, G. M. and Murayama, K. (2020) 'Development of Interest and Role of Choice During Sequential Knowledge Acquisition', *AERA Open*, Vol. 6, No. 2, pp. 233285842092998. https://doi.org/10.1177/2332858420929981

Frey, B. B. (2018) 'Convenience Sampling', *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*, Thousand Oaks: SAGE Publications, Inc. https://doi.org/10.4135/9781506326139.n155

Frömel, K., Groffik, D., Kudláček, M., Šafář, M., Zwierzchowska, A. and Mitáš, J. (2022) 'The Differences in Physical Activity Preferences and Practices among High versus Low Active Adolescents in Secondary Schools', *Sustainability*, Vol. 14, No. 2, pp. 891. https://doi.org/10.3390/su14020891

Gądek, A. (2020) 'Primary school students' attitudes towards physical culture, physical education lessons and exercises during a pandemic', *Health Promotion & Physical Activity*, Vol. 13, No. 4, pp. 40-45. https://doi.org/10.55225/hppa.177

Gao, Z. and Xiang, P. (2008) 'College Students' Motivation Toward Weight Training: An Application of Expectancy-Value Model', *Journal of Teaching in Physical Education*, Vol. 27, No. 3, pp. 399-415. https://doi.org/10.1123/jtpe.27.3.399

González-Peño, A., Franco, E. and Coterón, J. (2021) 'Do Observed Teaching Behaviors Relate to Students' Engagement in Physical Education?', *International Journal of Environmental Research and Public Health*, Vol. 18, No. 5, pp. 2234. https://doi.org/10.3390/ijerph18052234

Hands, B. and Parker, H. (2016) 'Male and Female Differences in Health Benefits Derived from Physical Activity: Implications for Exercise Prescription', *Journal of Womens Health, Issues and Care*, Vol. 5, No. 4. https://doi.org/10.4172/2325-9795.1000238

Harackiewicz, J. M., Smith, J. L. and Priniski, S. J. (2016) 'Interest Matters: The Importance of Promoting Interest in Education', *Policy Insights from the Behavioral and Brain Sciences*, Vol. 3, No. 2, pp. 220-227. https://doi.org/10.1177/2372732216655542

Hong, J.-C., Chang, C.-H., Tsai, C.-R. and Tai, K.-H. (2019) 'How situational interest affects individual interest in a STEAM competition', *International Journal of Science Education*, Vol. 41, No. 12, pp. 1667-1681. https://doi.org/10.1080/09500693.2019.1624992

Hovelius, E., Lindén, E., Bengtsson, H. and Håkansson, A. (2021) 'Self-Schema, Attachment Style, and Treatment Outcome of Patients in an Opiate Maintenance Treatment Unit', *Frontiers in Psychology*, Vol. 12, 595883. https://doi.org/10.3389/fpsyg.2021.595883

Hsu, S.-H., Chepyator-Thomson, J. R., Lackman, J. and Byron, K. (2022) 'Gendered Preferences in Selecting Physical Education Activity Skills Courses: A Study of a Taiwanese and a United States University', *International Research in Higher Education*, Vol. 7, No. 1, pp. 1. https://doi.org/10.5430/irhe.v7n1p1

Hui, Y. K., Li, C., Qian, S. and Kwok, L. F. (2019) 'Learning engagement via promoting situational interest in a blended learning environment', *Journal of Computing in Higher Education*, Vol. 31, No. 2, pp. 408-425. https://doi.org/10.1007/s12528-019-09216-z

Iconomescu, T.-M., Mindrescu, V. and Popovici, I.-M. (2018) 'A comparative study regarding secondary school students' satisfaction degree regarding the physical education class in Romanian and in Turkey', *SHS Web of Conferences*, Vol. 48, pp. 01028. https://doi.org/10.1051/shsconf/20184801028

Iqboljon, T. J. (2021) 'The Role of Physical Culture in Human Life', *Eurasian Journal of Humanities and Social Sciences*,Vol. 2, pp. 20-22. https://doi.org/10.46554/Russian.science-2019.10-1-359/362

Jaakkola, T., Wang, C. K. J., Soini, M. and Liukkonen, J. (2015) 'Students' Perceptions of Motivational Climate and Enjoyment in Finnish Physical Education: A Latent Profile Analysis', *Journal of Sports Science & Medicine*, Vol. 14, No. 3, pp. 477-83.

Jaya, L. H. S. and Ariyanto, E. (2021) 'The Effect of Vigor, Dedication and Absorption on the Employee Performance of PT Garuda Indonesia Cargo', *European Journal of Business and Management Research*, Vol. 6, No. 4, pp. 311-316. https://doi.org/10.24018/ejbmr.2021.6.4.1006

Kahu, E., Nelson, K. and Picton, C. (2017) 'Student interest as a key driver of engagement for first year students', *Student Success*, Vol. 8, No. 2, pp. 55-66. https://doi.org/10.5204/ssj.v8i2.379

Kang, H. and Zhao, H. (2020) 'Description and Application Research of Multiple Regression Model Optimization Algorithm Based on Data Set Denoising', *Journal of Physics: Conference Series*, Vol. 1631, No. 1, pp. 012063. https://doi.org/10.1088/1742-6596/1631/1/012063

Kang, J., Keinonen, T. and Salonen, A. (2021) 'Role of Interest and Self-Concept in Predicting Science Aspirations: Gender Study', *Research in Science Education*, Vol. 51, No. S1, pp. 513-535. https://doi.org/10.1007/s11165-019-09905-w

Knekta, E., Rowland, A. A., Corwin, L. A. and Eddy, S. (2020) 'Measuring university students' interest in biology: evaluation of an instrument targeting Hidi and Renninger's individual interest', *International Journal of STEM Education*, Vol. 7, No. 1, pp. 23. https://doi.org/10.1186/s40594-020-00217-4

Kokoulina, O. P., Davydova, Y. A., Kargapolova, E. V. and Glazynov, O. N. (2021) 'Physical Culture and Sports Activities in Self-Preserving Behavior of Students', *BIO Web of Conferences*, Vol. 29, 01002. https://doi.org/10.1051/bioconf/20212901002

Koob, C., Schröpfer, K., Coenen, M., Kus, S. and Schmidt, N. (2021) 'Factors influencing study engagement during the COVID-19 pandemic: A cross-sectional study among health and social professions students', *PLOS ONE*, Vol. 16, No. 7, pp. e0255191. https://doi.org/10.1371/journal.pone.0255191

Lauderdale, M. E., Yli-Piipari, S., Irwin, C. C. and Layne, T. E. (2015) 'Gender Differences Regarding Motivation for Physical Activity Among College Students: A Self-Determination Approach', *The Physical Educator*, Vol. 72, pp. 153-172. https://doi.org/10.18666/TPE-2015-V72-I5-4682

Li, B., Shamsuddin, A. and Braga, L. H. (2021) 'A guide to evaluating survey research methodology in pediatric urology', *Journal of Pediatric Urology*, Vol. 17, No. 2, pp. 263-268. https://doi.org/10.1016/j.jpurol.2021.01.009

Li, F., Chen, J. and Baker, M. (2014) 'University Students' Attitudes Toward Physical Education Teaching', *Journal of Teaching in Physical Education*, Vol. 33, No. 2, pp. 186-212. https://doi.org/10.1123/jtpe.2012-0187

Listau, K., Christensen, M. and Innstrand, S. T. (2017) 'Work Engagement: A Double-Edged Sword? A Study of the Relationship between Work Engagement and the Work-Home Interaction Using the ARK Research Platform', *Scandinavian Journal of Work and Organizational Psychology*, Vol. 2, No. 1, pp. 1-13. https://doi.org/10.16993/sjwop.20

Liu, F., Gai, X., Xu, L., Wu, X. and Wang, H. (2021) 'School Engagement and Context: A Multilevel Analysis of Adolescents in 31 Provincial-Level Regions in China', *Frontiers in Psychology*, Vol. 12, 724819. https://doi.org/10.3389/fpsyg.2021.724819

Lobo, J. (2023) 'Protecting Philippine Dance Traditions via Education of Tomorrow's Pedagogues: The Role of Individual Interest and School Engagement', *Journal of Ethnic and Cultural Studies*, Vol. 10, No. 1, pp. 98-124. https://doi.org/10.29333/ejecs/1527

Lobo, J., Dimalanta, G. and Bautista, C. (2022) 'An Investigation on the Factors Affecting Students' Interest in Physical Education Using Principal Component Analysis (PCA) at a Local City College in Angeles City, Pampanga, Philippines', *American Journal of Multidisciplinary Research and Innovation*, Vol. 1, No. 2, pp. 63-69. https://doi.org/10.54536/ajmri.v1i2.291

Lutkovskaya, O., Minenok, E., Antipin, N., Spashchanskaya, V., Prokopkina, S. and Yushchenko, A. (2021) '21st Century: Physical Fitness of School Age Students', *BIO Web of Conferences*, Vol. 29, 01015. https://doi.org/10.1051/bioconf/20212901015

Madejski, E., Jaros, A. and Madejski, R. (2019) 'Attitudes of secondary school students towards physical culture, physical education lessons and exercises', *Health Promotion & Physical Activity*, Vol. 7, No. 2, pp. 23-29. https://doi.org/10.5604/01.3001.0013.2756

Malinauskas, R., Emeljanovas, A. and Valantine, I. (2018) 'Retrospective attitudes towards the assessment system in physical education in the former Soviet Republic of Lithuania: Differences with regard to gender, education, age, and physical activity', *Baltic Journal of Health and Physical Activity*, Vol. 10, No. 3, pp. 38-47. https://doi.org/10.29359/BJHPA.10.3.04

Mallari, M. D. and Tayag, J. R. (2022) 'Situational Interest and Engagement of Public Junior High School Science Students in Modular Distance Learning', *International Journal of Instruction*, Vol. 15, No. 3, pp. 581-598. https://doi.org/10.29333/iji.2022.15332a

Mohamed, M. and Kamil, N. A. (2020) 'Relationship of Attitude Factors to Engagement in Physical Education among Secondary School Students', *International Journal of Academic Research in Business and Social Sciences*, Vol. 10, No. 14, pp. 171-180. https://doi.org/10.6007/IJARBSS/v10-i14/7686

Murayama, K. (2022) 'A reward-learning framework of knowledge acquisition: An integrated account of curiosity, interest, and intrinsic-extrinsic rewards.', *Psychological Review*, Vol. 129, No. 1, pp. 175-198. https://doi.org/10.1037/rev0000349

Mutlu, T. O., Senturk, H. E., Akoğlu, H. E. and Çetinkaya, A. (2021) 'The Analysis of High School Students' Attitudes towards Physical Education and Sports Class', *I-Manager's Journal on Educational Psychology*, Vol. 14, No. 3, pp. 54. https://doi.org/10.26634/jpsy.14.3.17663

O'Keefe, P. A. and Linnenbrink-Garcia, L. (2014) 'The role of interest in optimizing performance and self-regulation', *Journal of Experimental Social Psychology*, Vol. 53, pp. 70-78. https://doi.org/10.1016/j.jesp.2014.02.004

Palmer, D. (2019) 'Situational Interest and Actualized Individual Interest: Two Problematic Constructs', *Journal of Advances in Education Research*, Vol. 4, No. 3, pp. 110-112. https://doi.org/10.22606/jaer.2019.43002

Park, J. H. and Han, S. (2021) 'Exploring the factors of situational interest in learning mathematics', *The Mathematical Education*, Vol. 60, No. 4, pp. 555-580. https://doi.org/https://doi.org/10.7468/mathedu.2021.60.4.555

Pasco, D. and Roure, C. (2022) 'Situational interest impacts college students' physical activity in a design-based bike exergame', *Journal of Sport and Health Science*, Vol. 11, No. 2, pp. 172-178. https://doi.org/10.1016/j.jshs.2021.03.003

Peral-Suárez, Á., Cuadrado-Soto, E., Perea, J. M., Navia, B., López-Sobaler, A. M. and Ortega, R. M. (2020) 'Physical activity practice and sports preferences in a group of Spanish schoolchildren depending on sex and parental care: a gender perspective', *BMC Pediatrics*, Vol. 20, No. 1, pp. 337. https://doi.org/10.1186/s12887-020-02229-z

Perić, A., Rodrigues, A., Pankowska-Jurczyk, K., Batutis, O., Costa, J., Onofre, M. and Martins, J. (2020) 'Middle school students' views about physical education: A comparative study between four European countries (Opiniones de los estudiantes de secundaria sobre la educación física: Un estudio comparado entre cuatro países europeos)', *Retos*, Vol. 2041, No. 39, pp. 396-402. https://doi.org/10.47197/retos.v0i39.79755

Pituk, C. S. and Cagas, J. Y. (2019) 'Physical Activity and Physical Fitness Among Filipino University Students', *Journal of Physical Education*, Nol. 30, e3076, pp. 1-10. https://doi.org/10.4025/jphyseduc.v30i1.3076

Polet, J., Hassandra, M., Lintunen, T., Laukkanen, A., Hankonen, N., Hirvensalo, M., Tammelin, T., Hagger, M. S. (2019) 'Using physical education to promote out-of school physical activity in lower secondary school students - a randomized controlled trial protocol', *BMC Public Health*, Vol. 19, No. 1, pp. 157. https://doi.org/10.1186/s12889-019-6478-x

Powell, D. and Ceaser-White, F. (2017) 'Commentary on "Engaging Students in Physical Education: Key Challenges and Opportunities for Physical Education Teachers in Urban Settings"', *Journal of Physical Education, Recreation & Dance*, Vol. 88, No. 3, pp. 49-50. https://doi.org/10.1080/07303084.2017.1271267

Pulido-Martos, M., Cortés-Denia, D., de la Rosa-Blanca, J. J. and Lopez-Zafra, E. (2020) 'The Shirom-Melamed Vigor Measure for Students: Factorial Analysis and Construct Validity in Spanish Undergraduate University Students', *International Journal of Environmental Research and Public Health*, Vol. 17, No. 24, p. 9590. https://doi.org/10.3390/ijerph17249590

Quinlan, K. M. and Renninger, K. A. (2022) 'Rethinking employability: how students build on interest in a subject to plan a career', *Higher Education*, Vol. 84, No. 4, pp. 863-883. https://doi.org/10.1007/s10734-021-00804-6

Regmi, P. R., Waithaka, E., Paudyal, A., Simkhada, P. and Van Teijlingen, E. (2017) 'Guide to the design and application of online questionnaire surveys', *Nepal Journal of Epidemiology*, Vol. 6, No. 4, pp. 640-644. https://doi.org/10.3126/nje.v6i4.17258

Renninger, K. A. and Hidi, S. E. (2022) 'Interest: A unique affective and cognitive motivational variable that develops', *Advances in Motivation Science*, Vol. 9, pp. 179-239. https://doi.org/10.1016/bs.adms.2021.12.004

Resaland, G. K., Aadland, E., Andersen, J. R., Bartholomew, J. B., Anderssen, S. A. and Moe, V. F. (2019) 'Physical activity preferences of 10-year-old children and identified activities with positive and negative associations to cardiorespiratory fitness', *Acta Paediatrica*, Vol. 108, No. 2, pp. 354-360. https://doi.org/10.1111/apa.14487

Ricardo, L. I. C., Wendt, A., Costa, C. dos S., Mielke, G. I., Brazo-Sayavera, J., Khan, A., Kolbe-Alexander, T. L., Crochemore-Silva, I. (2022) 'Gender inequalities in physical activity among adolescents from 64 Global South countries', *Journal of Sport and Health Science*, Vol. 11, No. 4, pp. 509-520. https://doi.org/10.1016/j.jshs.2022.01.007

Rojo-Ramos, J., González-Becerra, M. J., Gómez-Paniagua, S. and Adsuar, J. C. (2022) 'Satisfaction with Physical Activity among Students in the Last Cycle of Primary Education in Extremadura', *International Journal of Environmental Research and Public Health*, Vol. 19, No. 11, 6702. https://doi.org/10.3390/ijerph19116702

Romano, L., Angelini, G., Consiglio, P. and Fiorilli, C. (2021) 'Academic Resilience and Engagement in High School Students: The Mediating Role of Perceived Teacher Emotional Support', *European Journal of Investigation in Health, Psychology and Education*, Vol. 11, No. 2, pp. 334-344. https://doi.org/10.3390/ejihpe11020025

Romine, W., Tsai, C.-L., Miller, M., Tang, N.-E. and Folk, W. (2020) 'Evaluation of a Process by which Individual Interest Supports Learning Within a Formal Middle School Classroom Context', *International Journal of Science and Mathematics Education*, Vol. 18, No. 7, pp. 1419-1439. https://doi.org/10.1007/s10763-019-10032-1

Rotgans, J. I. and Schmidt, H. G. (2017) 'The relation between individual interest and knowledge acquisition', *British Educational Research Journal*, Vol. 43, No. 2, pp. 350-371. https://doi.org/10.1002/berj.3268

Rotgans, J. I. and Schmidt, H. G. (2018) 'How individual interest influences situational interest and how both are related to knowledge acquisition: A microanalytical investigation', *The Journal of Educational Research*, Vol. 111, No. 5, pp. 530-540. https://doi.org/10.1080/00220671.2017.1310710

Roure, C., Lentillon-Kaestner, V., Méard, J. and Pasco, D. (2019a) 'Universality and Uniqueness of Students' Situational Interest in Physical Education: A Comparative Study', *Psychologica Belgica*, Vol. 59, No. 1, pp. 1-15. https://doi.org/10.5334/pb.446

Roure, C., Lentillon-Kaestner, V. and Pasco, D. (2021) 'Students' individual interest in physical education: Development and validation of a questionnaire', *Scandinavian Journal of Psychology*, Vol. 62, No. 1, pp. 64-73. https://doi.org/10.1111/sjop.12669

Roure, C., Méard, J., Lentillon-Kaestner, V., Flamme, X., Devillers, Y. and Dupont, J.-P. (2019b) 'The effects of video feedback on students' situational interest in gymnastics', *Technology, Pedagogy and Education*, Vol. 28, No. 5, pp. 563-574. https://doi.org/10.1080/1475939X.2019.1682652

Roure, C. and Pasco, D. (2018a) 'Exploring situational interest sources in the French physical education context', *European Physical Education Review*, Vol. 24, No. 1, pp. 3-20. https://doi.org/10.1177/1356336X16662289

Roure, C. and Pasco, D. (2018b) 'The Impact of Learning Task Design on Students' Situational Interest in Physical Education', *Journal of Teaching in Physical Education*, Vol. 37, No. 1, pp. 24-34. https://doi.org/10.1123/jtpe.2017-0046

Rowland, A. A., Knekta, E., Eddy, S. and Corwin, L. A. (2019) 'Defining and Measuring Students' Interest in Biology: An Analysis of the Biology Education Literature', *CBE—Life Sciences Education*, Vol. 18, No. 3, ar34. https://doi.org/10.1187/cbe.19-02-0037

Sağın, A. E. (2022) 'The role of gender in predicting life satisfaction of the interest in physical education lesson', *Pedagogy of Physical Culture and Sports*, Vol. 26, No. 2, pp. 83-92. https://doi.org/10.15561/26649837.2022.0202

Salmela-Aro, K. and Upadyaya, K. (2017) 'Co-Development of Educational Aspirations and Academic Burnout from Adolescence to Adulthood in Finland', *Research in Human Development*, Vol. 14, No. 2, pp. 106-121. https://doi.org/10.1080/15427609.2017.1305809

Schmidt, H. G. and Rotgans, J. I. (2021) 'Epistemic Curiosity and Situational Interest: Distant Cousins or Identical Twins?', *Educational Psychology Review*, Vol. 33, No. 1, pp. 325-352. https://doi.org/10.1007/s10648-020-09539-9

Scott, W. D., Penningroth, S. L., Paup, S., Li, X., Adams, D. and Mallory, B. (2022) 'The Relational Self-Schema Measure: Assessing Psychological Needs in Multiple Self-with-Other Representations', *Journal of Personality Assessment*, Vol. 104, No. 1, pp. 74-85. https://doi.org/10.1080/00223891.2021.1900207

Scrabis-Fletcher, K. and Silvermanv, S. (2010) 'Perception of Competence in Middle School Physical Education', *Research Quarterly for Exercise and Sport*, Vol. 81, No. 1, pp. 52-61. https://doi.org/10.1080/02701367.2010.10599627

Shang, C., Moss, A. C. and Chen, A. (2022) 'The expectancy-value theory: A meta-analysis of its application in physical education', *Journal of Sport and Health Science*, Vol. 12, No. 1, pp. 52-64. https://doi.org/10.1016/j.jshs.2022.01.003

Shin, D. D. and Kim, S. il (2019) 'Homo Curious: Curious or Interested?', *Educational Psychology Review*, Vol. 31, No. 4, pp. 853-874. https://doi.org/10.1007/s10648-019-09497-x

Sierra-Díaz, M. J., González-Víllora, S., Pastor-Vicedo, J. C. and López-Sánchez, G. F. (2019) 'Can We Motivate Students to Practice Physical Activities and Sports Through Models-Based Practice? A Systematic Review and Meta-Analysis of Psychosocial Factors Related to Physical Education', *Frontiers in Psychology*, Vol. 10, p. 2115. https://doi.org/10.3389/fpsyg.2019.02115

Sofi, M. A., Waseem, M. and Padder, J. (2019) 'Comparative study of attitude and interest of junior college boys towards physical education', *International Journal of Physiology, Nutrition and Physical Education*, Vol. 4, No. 1, pp. 367-368.

Solomonko, A., Zanevskyy, I., Bodnarchuk, O., Andres, A., Petryna, R. and Lapychak, I. (2022) 'Attitude of law college students towards physical culture and sports', *Journal of Physical Education and Sport*, Vol. 22, No. 3, pp. 780-785. https://doi.org/10.7752/jpes.2022.03099

Sucuoglu, E. and Atamturk, H. (2020) 'Correlation relation between professional qualifications of physical education teachers and students' attitudes towards Physical Education classes', *Pedagogy of Physical Culture and Sports*, Vol. 24, No. 1, pp. 44-47. https://doi.org/10.15561/18189172.2020.0107

Sundjaja, J. H., Shrestha, R. and Krishan, K. (2022) *McNemar And Mann-Whitney U TestsStatPearls*, St. Petersburg: StatPearls Publishing.

Svenningsson, J., Höst, G., Hultén, M. and Hallström, J. (2022) 'Students' attitudes toward technology: exploring the relationship among affective, cognitive and behavioral components of the attitude construct', *International Journal of Technology and Design Education*, Vol. 32, No. 3, pp. 1531-1551. https://doi.org/10.1007/s10798-021-09657-7

Tagare, R. J. L. and Villaluz, G. D. C. (2021) 'Activity Preferences of Generation Z Students for Tertiary Physical Education: Implications for Curriculum Enhancement', *Multidisciplinary Journal for Education, Social and Technological Sciences*, Vol. 8, No. 2, pp. 92. https://doi.org/10.4995/muse.2021.15492

Tainio, M. (2019) 'Contemporary physical activities: the aesthetic justification', *Sport in Society*, Vol. 22, No. 5, pp. 846-860. https://doi.org/10.1080/17430437.2018.1430483

Teixeira, P. J., Carraça, E. V, Markland, D., Silva, M. N. and Ryan, R. M. (2012) 'Exercise, physical activity, and self-determination theory: A systematic review', *International Journal of Behavioral Nutrition and Physical Activity*, Vol. 9, No. 1, pp. 78. https://doi.org/10.1186/1479-5868-9-78

Teuber, Z., Nussbeck, F. W. and Wild, E. (2021) 'The Bright Side of Grit in Burnout-Prevention: Exploring Grit in the Context of Demands-Resources Model among Chinese High School Students', *Child Psychiatry & Human Development*, Vol. 52, No. 3, pp. 464-476. https://doi.org/10.1007/s10578-020-01031-3

Thomas, M. and Ti, M. (2021) 'Gender differences in physical education conceptual knowledge on physical education teacher education (PETE) students', *International Journal of Physical Education, Sports and Health*, Vol. 8, No. 5, pp. 99-105.

Upadyaya, K., Cumsille, P., Avalos, B., Araneda, S., Lavonen, J. and Salmela-Aro, K. (2021) 'Patterns of situational engagement and task values in science lessons', *The Journal of Educational Research*, Vol. 114, No. 4, pp. 394-403. https://doi.org/10.1080/00220671.2021.1955651

Viva, E. B. and Limbo, C. B. (2021) 'Motivation, Attitude, And Competence of Physical Education Students', *Canadian Journal of Educational and Social Studies*, Vol. 1, No. 2, pp. 1-12. https://doi.org/10.53103/cjess.v1i2.9

Wang, H., Shen, B. and Bo, J. (2022) 'Examining Situational Interest in Physical Education: A New Inventory', *Journal of Teaching in Physical Education*, Vol. 41, No. 2, pp. 270-277. https://doi.org/10.1123/jtpe.2020-0218

Wang, J. C. K., Liu, W. C., Chatzisarantis, N. L. D. and Lim, C. B. S. (2010) 'Influence of Perceived Motivational Climate on Achievement Goals in Physical Education: A Structural Equation Mixture Modeling Analysis', *Journal of Sport and Exercise Psychology*, Vol. 32, No. 3, pp. 324-338. https://doi.org/10.1123/jsep.32.3.324

Wang, Q. and Xue, M. (2022) 'The implications of expectancy-value theory of motivation in language education', *Frontiers in Psychology*, Vol. 13, pp. 1-8. https://doi.org/10.3389/fpsyg.2022.992372

Widlund, A., Tuominen, H. and Korhonen, J. (2021) 'Development of school engagement and burnout across lower and upper secondary education: Trajectory profiles and educational outcomes', *Contemporary Educational Psychology*, Vol. 66, p. 101997. https://doi.org/10.1016/j.cedpsych.2021.101997

Wild, S. (2022) 'Trajectories of subject-interests development and influence factors in higher education', *Current Psychology*, Vol. 42, No. 15, pp. 12879-12895. https://doi.org/10.1007/s12144-021-02691-7

Wong, L.-H., Chan, T.-W., Chen, W., Looi, C.-K., Chen, Z.-H., Liao, C. C. Y., King, R. B., Wong, S. L. (2020) 'IDC theory: interest and the interest loop', *Research and Practice in Technology Enhanced Learning*, Vol. 15, No. 1, pp. 1-16. https://doi.org/10.1186/s41039-020-0123-2

Yalgashevich, K. S., Shermamatovich, M. M. and Zayniddinovich, N. I. (2021) 'The Role Of Physical Culture In The Life Of Students', *International Journal of Scientific and Technology Research*, Vol. 10, No. 03, pp. 1-4.

Yli-Piipari, S., Jaakkola, T., Liukkonen, J. and Nurmi, J.-E. (2013) 'The effect of physical education students' beliefs and values on their physical activity: A growth mixture modelling approach', *International Journal of Sport and Exercise Psychology*, Vol. 11, No. 1, pp. 70-86. https://doi.org/10.1080/1612197X.2012.731191

Zalech, M. (2021) 'Student perception of pe teachers and its effect on their participation in pe classes and sports: A new perspective on teacher competencies', *Journal of Physical Education and Sport*, Vol. 21, No. 2, pp. 1106-1111. https://doi.org/10.7752/jpes.2021.s2139

Zhang, J. (2021) 'The Lack of Culture in School Physical Education and its Countermeasures', *International Journal of New Developments in Education*, Vol. 3, No. 3, pp. 66-70. https://doi.org/10.25236/IJNDE.2021.030315

Zhu, X. and Chen, A. (2013) 'Adolescent Expectancy-Value Motivation, Achievement in Physical Education, and Physical Activity Participation', *Journal of Teaching in Physical Education*, Vol. 32, No. 3, pp. 287-304. https://doi.org/10.1123/jtpe.32.3.287

Full research paper

DEVELOPMENT OF COMPUTER-BASED CHEMICAL FIVE-TIER DIAGNOSTIC TEST INSTRUMENTS: A GENERALIZED PARTIAL CREDIT MODEL

Achmad Rante Suparman[1]✉
Eli Rohaeti[2]
Sri Wening[2]

[1]Universitas Papua, Indonesia

[2]Universitas Negeri Yogyakarta, Indonesia

✉ a.rante@unipa.ac.id

ABSTRACT

This study focuses on developing a five-tier chemical diagnostic test based on a computer-based test with 11 assessment categories with an assessment score from 0 to 10. A total of 20 items produced were validated by education experts, material experts, measurement experts, and media experts, and an average index of the Aiken test > 0.70 was obtained. The validation results were tested on 580 respondents and analyzed using the Generalized Partial Credit Model (GPCM) Item Response Theory (IRT) type. The results of the analysis show that all of the items meet the requirements to be said to be valid for the model; the evidence of the value this: RMSEA < 0.08, CFI > 0.87, SRMR < 0.10, GFI > 0.90, NFI > 0.90, NNFI > 0.90, IFI > 0.90, TLI > 0.90, and RFI > 0.90, and all items were obtained has a p.S_X2 value greater than 0.05 which indicates that all items developed are fit and by the GPCM model. The construct reliability (CR) value is 0.99, which suggests the construct is reliable. The most challenging item is item 9, and the most accessible item is item 4.

KEYWORDS

Five-tier, chemical diagnostic, computer-based test, Generalized Partial Credit Model

HOW TO CITE

Suparman A. R., Rohaeti E., Wening S. (2024) 'Development of Computer-Based Chemical Five-Tier Diagnostic Test Instruments: A Generalized Partial Credit Model', *Journal on Efficiency and Responsibility in Education and Science*, vol. 17, no. 1, pp. 91-105. http://dx.doi.org/10.7160/eriesj.2024.170108

Article history
Received
July 5, 2023
Received in revised form
August 19, 2023
Accepted
October 24, 2023
Available on-line
March 31, 2024

Highlights

- *The five-tier chemical diagnostic test is in the form of questions with five interrelated levels.*
- *The Generalized Partial Credit Model (GPCM) is a psychometric model in item response theory used to analyze polytomous data.*
- *The items developed have difficulty parameters ranging from -2 to +2, which indicates that the produced items are excellent and informative about students' abilities.*

INTRODUCTION

One of the factors that can influence learning is students' prior knowledge (Merriënboer and Bruin, 2014). Students' misunderstandings about a material can affect the following learning process, play a role in the formation of new knowledge, and can be an inhibiting factor in constructing actual knowledge for these students (Özmen, 2004). Correcting student conceptual errors must be implemented (Üce and Ceyhan, 2019). Teachers can make contextual errors experienced by students as a basis for starting learning so that the expected goals can direct the learning methods used. According to Barke, Hazari and Sileshi Yitbarek (2009), a good lesson is correcting misunderstandings in students and providing correct knowledge, not just providing knowledge without detecting student misunderstandings.

Chemistry is a compulsory subject in high school that studies matters relating to the matter and its changes. In chemistry, many concepts are macroscopic, microscopic, and symbols. According to Tien and Osman (2017), macroscopic chemical processes can be observed and felt by sensory motors; microscopic describe particles' arrangement, interaction, and movement, while representations in symbols, numbers, formulas, and equations are called chemical symbols. Treagust, Chittleborough and Mamiala (2003) state that the symbolic level is represented in chemical symbols, formulas, and reaction equations. Misunderstandings in schools can be caused by specific scientific terminology and language problems, especially substances, particles, and chemical symbols that must be distinguished (Barke et al., 2009). The symbolic level is a representation of chemistry, so the symbolic level must be understood so that students can broadly realize chemistry concepts. Wang et al. (2017) stated that the representation of chemical symbols is a medium for the transformation between

the actual phenomena of the macroscopic world and the sub-microscopic world. Chi et al. (2018) due to its abstract nature, many students struggle with learning and effectively utilizing these symbolic representations, which can lead to ongoing failure in subsequent chemistry learning. Taking the perspective of learning progressions, this study identifies how students' abilities in chemical symbol representation progress at different grade levels (Grade 10-12) also revealed that the representation of chemical symbols is widely used in chemistry learning. However, because of its abstract nature, many students need help to learn and use these symbolic representations effectively, causing difficulties in understanding chemistry.

Misconceptions resist change and hinder students' ability to understand scientific concepts and form new cognitive structures; therefore, misunderstandings about concepts must be corrected (Potvin, 2023). Some misconceptions students experience in studying chemistry include the following: many students still think that in an equilibrium system, the forward reaction rate differs from the reverse reaction rate (Harza, Wiji and Mulyani, 2021). Students assume that the volume of liquid mixed in liquid Solubility does not affect the density of the mixture (Kiray and Simsek, 2020). Students need help correctly abstracting the concept of acids and bases (Rusmini et al., 2021). Students need clarification about polarity and need to distinguish between covalent bonds and polar and nonpolar bonds (Derkach, 2021). Students write negative reaction equations (Widarti et al., 2021). Misconceptions about reaction rates are in the form of the assumption that activation energy is the amount of energy released during a reaction and that the catalyst does not affect the reaction mechanism (Jusniar et al., 2021). Students experience misconceptions about redox because they need to understand the reduction and oxidation of the term (Murniningsih, Muna and Irawati, 2020). Students must still clearly understand the primary variable's effect on the solution's boiling point (Llanos et al., 2021). Some students need clarification about the concept of rate constants (Lamichhane, Reck and Maltese, 2018).

Equating knowledge or cognitive structures, such as very complex chemistry, is not easy, and it is unsurprising that students from high school to university still need various clarifications (Vladusic, Bucat and Ozic, 2022). A diagnostic test can be used to find out whether students understand a concept correctly or not (Istiyono et al., 2023). In addition to diagnosing student errors in understanding concepts, another function of the diagnostic is to provide input to teachers in making decisions in learning (Wang et al., 2023). Diagnostic tests can be used to discover concepts truly understood by students, ideas only partially understood, and visions that students have misconceptions about. In understanding the level of student misconceptions, especially in the field of chemistry, several types of diagnostic tests can be used, such as a two-tier diagnostic test (Mutlu and Sesen, 2015), three-tier diagnostic test (Prodjosantoso, Hertina and Irwanto, 2019), four-tier diagnostic test (Dewi, Parlan and Suryadharma, 2020) (two and finally the five-tier diagnostic test (Putra, Hamidah and Nahadi, 2020).

The five-tier diagnostic test can be combined with a computer-based test to make it easier for students to take it and for teachers to check students' work. According to Pokorný (2023), teachers must integrate modern technology into teaching. Lowyck (2014) states that the basic principle in the interaction between technology and education is how technology can support individuals and groups to achieve learning goals. Groen and Eggen (2020) said that developing a test using a Computer Based Test is the first choice the developer must make. Currently, facilities in the form of computer technology due to the discovery of computer software for use in the classroom are giving positive results, one of which significantly affects the motivation to use it (Kimmons, Clark and Lim, 2017; Suparman, Rohaeti and Wening, 2023) teacher candidates and K-12 students in a state in the USA (n = 2261). Istiyono et al. (2020) state that using Computer Based Tests can save time, and the results obtained by students come out immediately after students complete the test. Mills and Breithaupt (2016) also argue various benefits of implementing Computer Based Tests in testing, including increased measurement accuracy and efficiency, convenience, speed of reporting results, increased access to information sources and tools, and ability to assess complex skills and experience of examinees.

METHODOLOGY

Research design

Two types of research and development models are used in this study: the design of the test instrument development model and the design of the media development model. The design of the test instrument development model used the Oriondo and Antonio test development model, and the media development model used the rapid prototyping model. The design of the test instrument development model and the creation of the media development model collaborate to make it more effective because there are stages in instrument development and media development that can be carried out simultaneously. According to Oriondo and Antonio (1984), the test development model consists of four stages: instrument design, instrument testing, empirical validity determination, and reliability determination. According to Martin and Betrus (2019), the rapid prototyping model consists of assessing needs and analyzing content, constructing a prototype, utilizing a prototype, and maintaining the system. The collaboration of the two models resulted in four stages, namely: (1) designing the test instrument and CBT media, (2) integrating the instrument into CBT, (3) testing the instrument using CBT, and (4) analyzing the results of the trial.

Analysis and sample

The research analysis uses the Generalized Partial Credit Model (GPCM) so that the research sample meets the minimum requirements for analysis with GPCM. According to Debelak, Stobl and Zeigenfuse (2022), for items 5 to 20, the required sample size is 500 to obtain an accurate estimate of the GPCM model. This research develops 20 items, so the subjects have at least 500 samples. The research subjects used were 580 students from 19 schools consisting of schools with a and B accreditation. The sample selection was based on the sampling area, so the sample consisted of

students from three regions of Indonesia: West Indonesia, Central Indonesia, and East Indonesia.

Data collection technique

The data collection technique was done through a five-tier chemical diagnostic test based on CBT. The five-tier chemical diagnostic test consists of five levels of questions that form a single unit. The first question is the central question; the second question is the level of confidence in answering the main question; the third question is the reason for choosing the answer to the main question; the fourth question is the level of confidence in the cause, and the fifth question is a chemical symbolic question related to the main question. Table 1 shows the categories and scoring of the five-tier diagnostic test resulting from the development and modification of Anam et al. (2019) and Bayuni, Sopandi and Sujana (2018), which consists of 32 answer patterns. The five-tier chemical diagnostic test is integrated with a computer-based test, and students do it online. Figure 1 shows the Computer-based Test flowchart used in this study.

Answer	Confidence Level of Answers	Reason	Reason Confidence Level	Chemical Symbolic Knowledge	Category	Score
Correct	Sure	Correct	Sure	Correct	Understand	10
Correct	Not sure	Correct	Sure	Correct	Understand but lack confidence	9
Correct	Sure	Correct	Not sure	Correct		
Correct	Not sure	Correct	Not sure	Correct		
Correct	Sure	Correct	Sure	Wrong	Type 1 (lack of knowledge)	8
Correct	Not sure	Correct	Not sure	Wrong		
Correct	Sure	Correct	Not sure	Wrong		
Correct	Not sure	Correct	Sure	Wrong		
Correct	Sure	Wrong	Not sure	Correct	Type 2 (lack of knowledge)	7
Correct	Not sure	Wrong	Not sure	Correct		
Correct	Sure	Wrong	Sure	Correct		
Correct	Not sure	Wrong	Sure	Correct		
Correct	Sure	Wrong	Not sure	Wrong	Type 3 (lack of knowledge)	6
Correct	Not sure	Wrong	Sure	Wrong		
Correct	Sure	Wrong	Sure	Wrong		
Correct	Not sure	Wrong	Not sure	Wrong		
Wrong	Not sure	Correct	Sure	Correct	Type 4 (lack of knowledge)	5
Wrong	Not sure	Correct	Not sure	Correct		
Wrong	Sure	Correct	Not sure	Correct		
Wrong	Sure	Correct	Sure	Correct		
Wrong	Not sure	Correct	Sure	Wrong	Type 5 (lack of knowledge)	4
Wrong	Sure	Correct	Not sure	Wrong		
Wrong	Not sure	Correct	Not sure	Wrong		
Wrong	Sure	Correct	Sure	Wrong		
Wrong	Not sure	Wrong	Not sure	Correct	Guess knowledge	3
Wrong	Sure	Wrong	Not sure	Correct	Partial misconception	2
Wrong	Not sure	Wrong	Sure	Correct		
Wrong	Sure	Wrong	Sure	Correct		
Wrong	Sure	Wrong	Not sure	Wrong	Complete misconception	1
Wrong	Not sure	Wrong	Sure	Wrong		
Wrong	Sure	Wrong	Sure	Wrong		
Wrong	Not sure	Wrong	Not sure	Wrong	No knowledge	0

Table 1: Categories And Scoring Of the Five-Tier Diagnostic Test

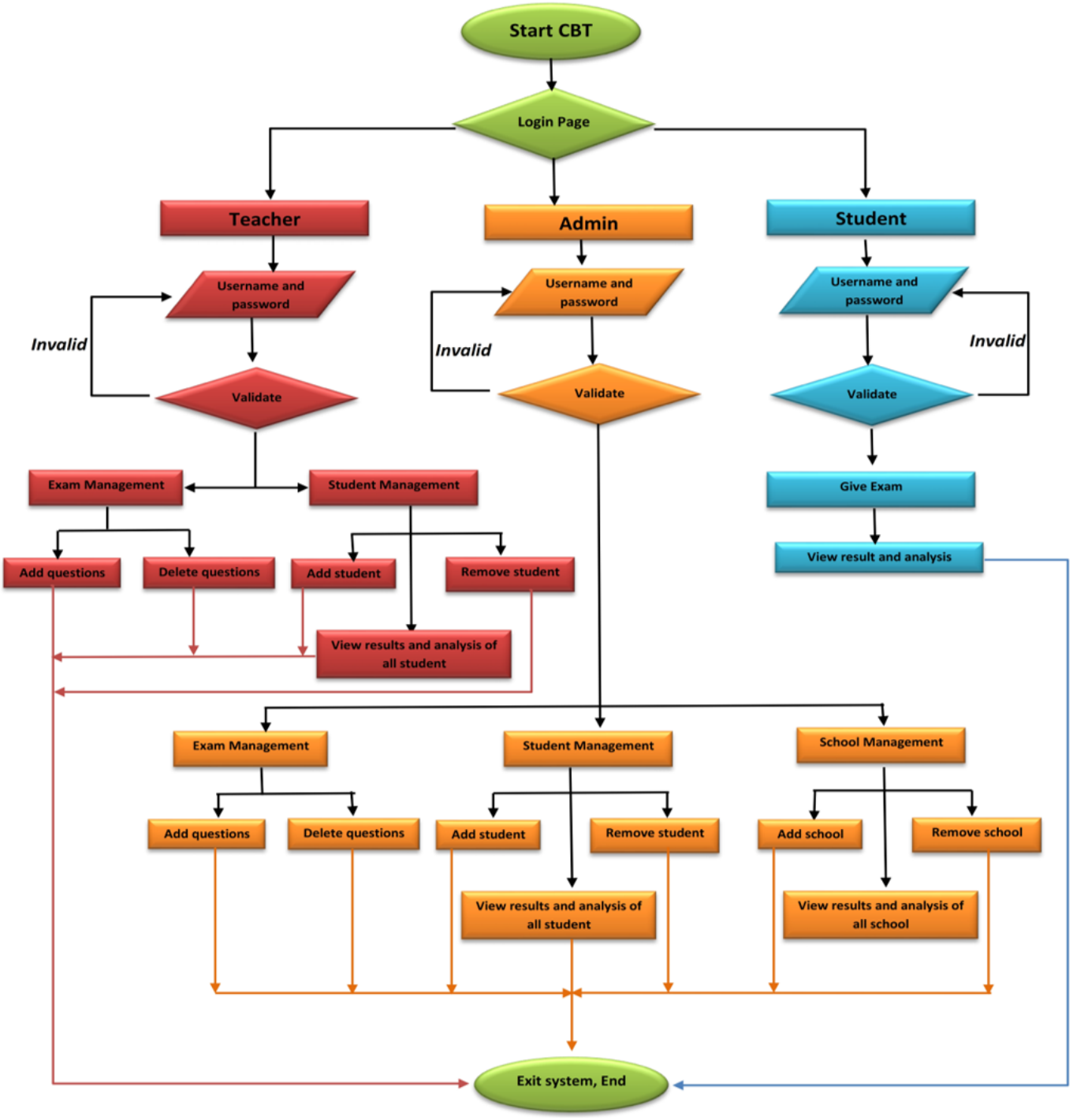

Figure 1: Flowchart Computer Based Test

RESULTS

The five-tier chemical diagnostic test grid contains indicators of a chemical material developed and continued in writing chemical questions. The suitability of the indicators with the items can be seen from the validation results carried out by experts. Content validation performed by experts was calculated using Aiken's V formula according to Table 2. Aiken suggested that valid instruments must have a validity range between 0.7 and 1. A validity range of 0.7 indicates that the set of tools is quite good, while a range of 0.9 means high validity (Aiken, 1980).

Table 2 shows that all item items are declared valid because the material, grid, indicators, and item items are appropriate or essential. The analysis results show that the instruments developed are crucial and by the curriculum, as evidenced by the average value of the Aiken index test > 0.70. The valid instruments were then tested on 580 respondents. The results of the instrument testing were analyzed using Confirmatory Factor Analysis (CFA). Figure 2 shows the Confirmatory Factor Analysis plot.

The feasibility analysis of the instrument obtained from the CFA analysis by Table 3.

Item number	Aiken index	Information	Item number	Aiken index	Information
Q1	0.93	Valid	Q11	0.96	Valid
Q2	0.96	Valid	Q12	1.00	Valid
Q3	0.96	Valid	Q13	0.93	Valid
Q4	1.00	Valid	Q14	1.00	Valid
Q5	1.00	Valid	Q15	0.96	Valid
Q6	0.93	Valid	Q16	0.93	Valid
Q7	0.96	Valid	Q17	0.96	Valid
Q8	1.00	Valid	Q18	0.96	Valid
Q9	0.86	Valid	Q19	1.00	Valid
Q10	1.00	Valid	Q20	0.86	Valid

Table 2: Categories And Scoring Of the Five-Tier Diagnostic Test

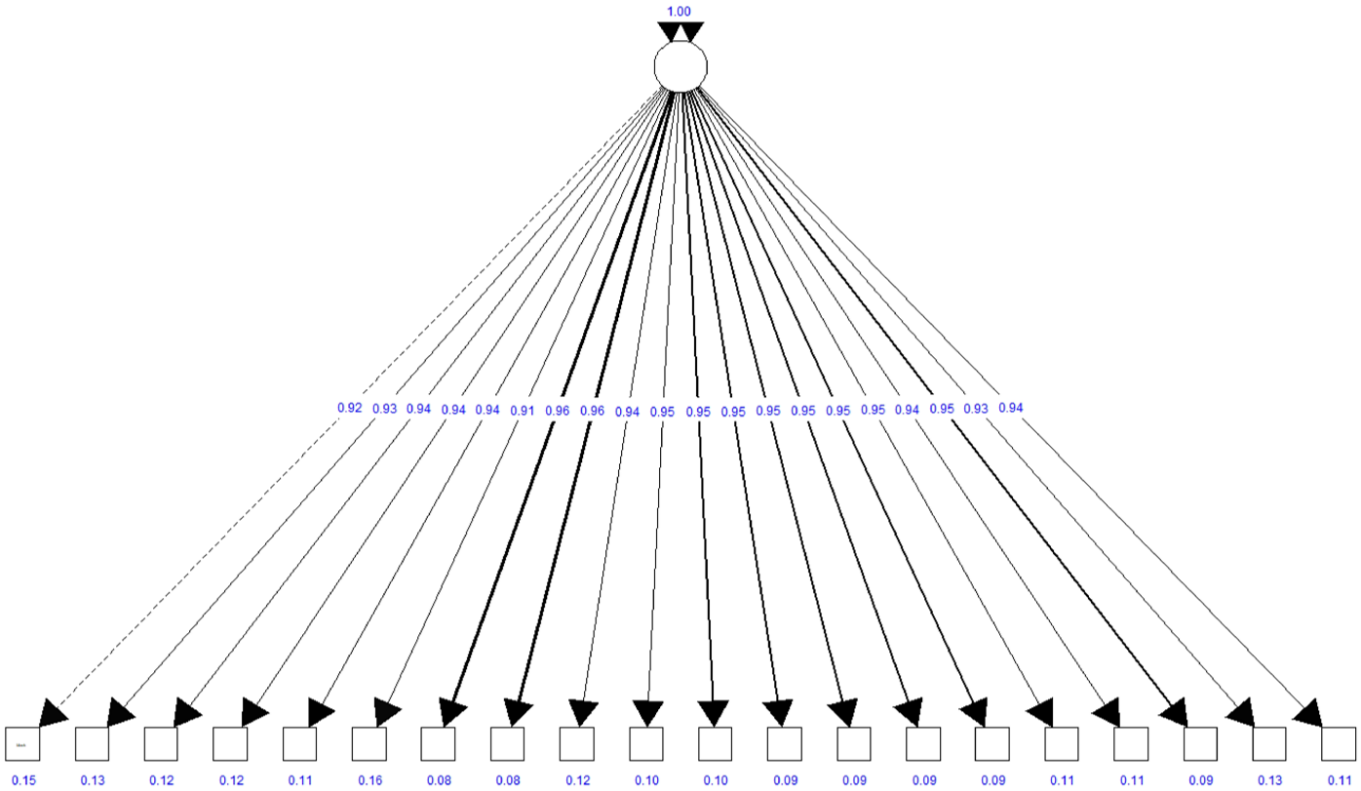

Figure 2: Confirmatory Factor Analysis (CFA)

Number	Category Name	Acceptance Category According to Theory	Analysis Results	Information
1	RMSEA	RMSEA < 0.08 (Cornick, 2015)	0.07	Fit
2	CFI	CFI > 0.87 (Dagnall et al., 2018)	0.99	Fit
3	SRMR	SRMR < 0.10 (Dagnall et al., 2018)	0.01	Fit
4	GFI	GFI > 0.90 (Kwahk and Lee, 2008)	0.95	Fit
5	NFI	NFI > 0.90 (Kwahk and Lee, 2008)	0.98	Fit
6	NNFI	NNFI > 0.90 (Kwahk and Lee, 2008)	0.97	Fit
7	IFI	IFI > 0.90 (Marsh, Balla and Mcdonald, 1988)	0.99	Fit
8	TLI	TLI > 0.90 (Marsh et al., 1988)	0.97	Fit
9	RFI	RFI > 0.90 (Marsh et al., 1988)	0.97	Fit

Table 3: Instrument Feasibility Analysis

Proving the assumptions of item response theory

The proof of the assumptions of the theory response items consists of three: unidimensional tests, local independence, and parameter invariance. Unidimensional is the ability of a question to measure only one ability. The test is unidimensional if the items are statistically dependent on the entire population (Crocker and Algina, 2008). The unidimensional assumption can be seen from the scree plot exploratory factor analysis shown in Figure 3. In Figure 3, it can be seen that there is one

factor that is measured in the chemical five-tier diagnostic test instrument. The steepness of the graph on one element is enough to prove unidimensional assumptions (Linden, 2018; Linden and Hambleton, 1997; Suparman, Rohaeti and Wening, 2022).

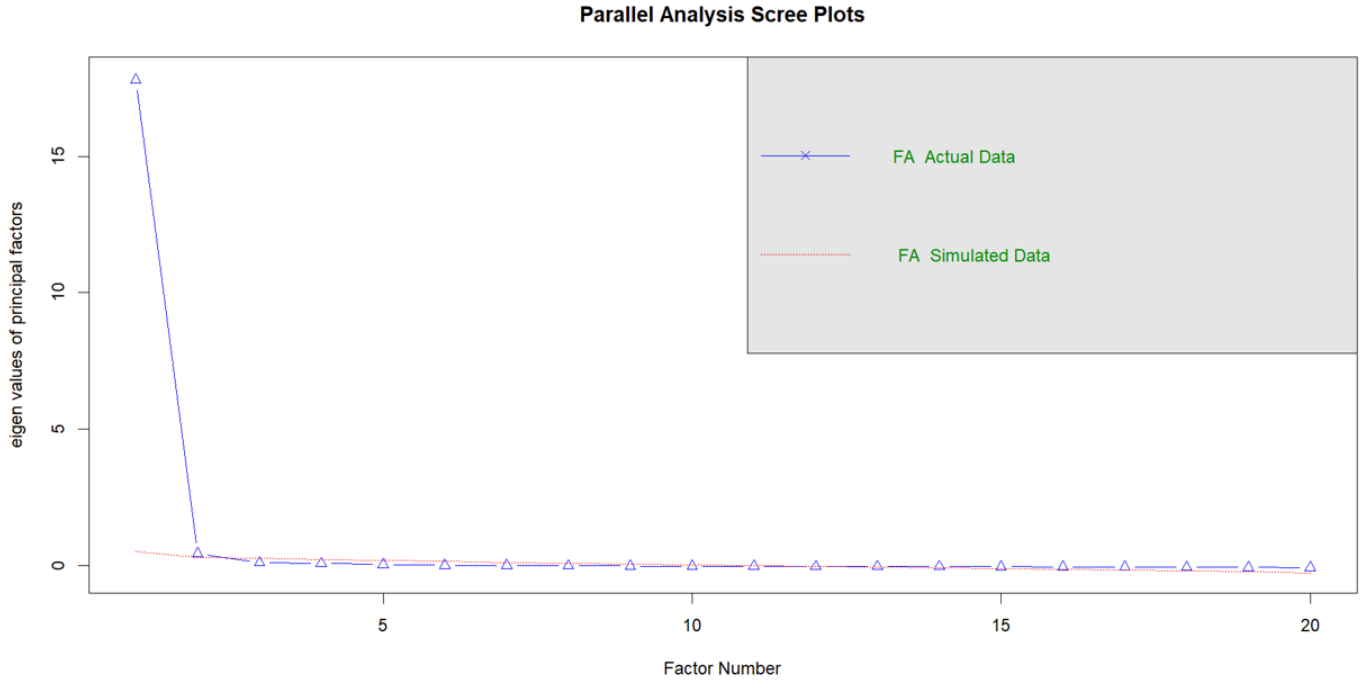

Figure 3: Scree Plot Exploratory Factor Analysis

The second assumption is local independence. Local independence is fulfilled if the students' answers are independent of their answers to other questions. The premise of local independence will automatically be proven if the unidimensional test has been established (Demars, 2010; Hambleton, 2006; Hambleton, Swaminathan and Rogers, 1991).

The third assumption is parameter invariance. Parameter invariance indicates that question parameters do not depend on the sample of examinees (Rupp and Zumbo, 2006) the equality of item and examinee parameters from different examinee populations or measurement conditions. In this article, using the well-known fact that item and examinee parameters are identical only up to a set of linear transformations specific to the functional form of a given IRT model, violations of these transformations for unidimensional IRT models are investigated using analytical, numerical, and visual tools. Because item parameter drift (IPD). In GPCM, there are two parameters, so parameter invariance also consists of item parameter invariance and ability parameter invariance. There are two item invariances, namely item parameter invariance based on differential power, according to Figure 4, and item parameter invariance based on difficulty, according to Figure 5. Ability parameter invariance based on odd and even items is shown in Figure 6.

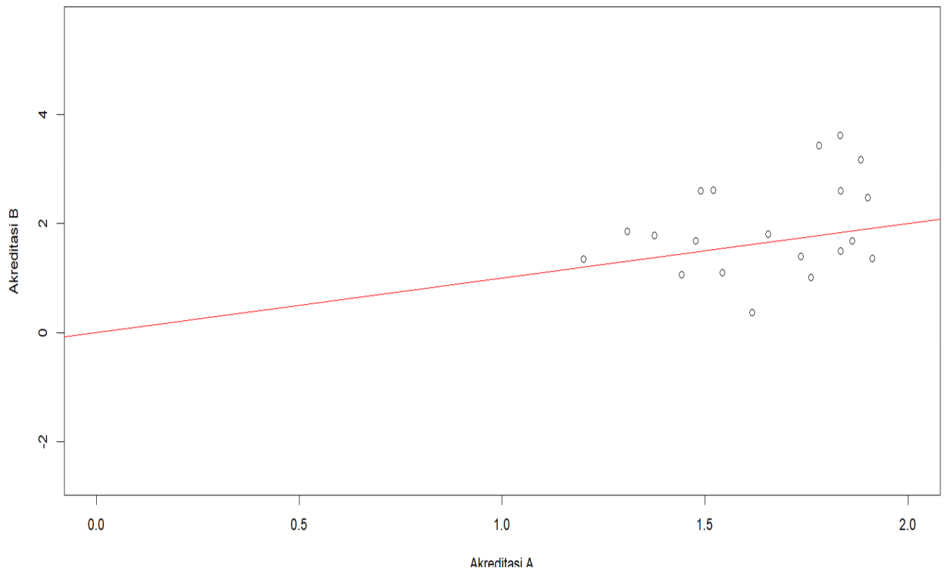

Figure 4: Invariance of Item Parameters Based on Differential Power

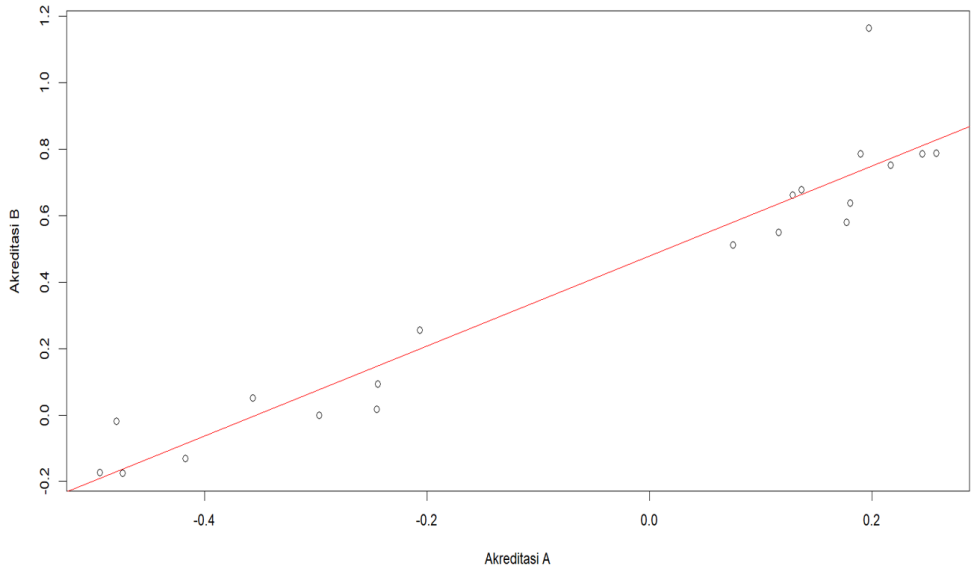

Figure 5: Invariance of Item Parameters by Difficulty

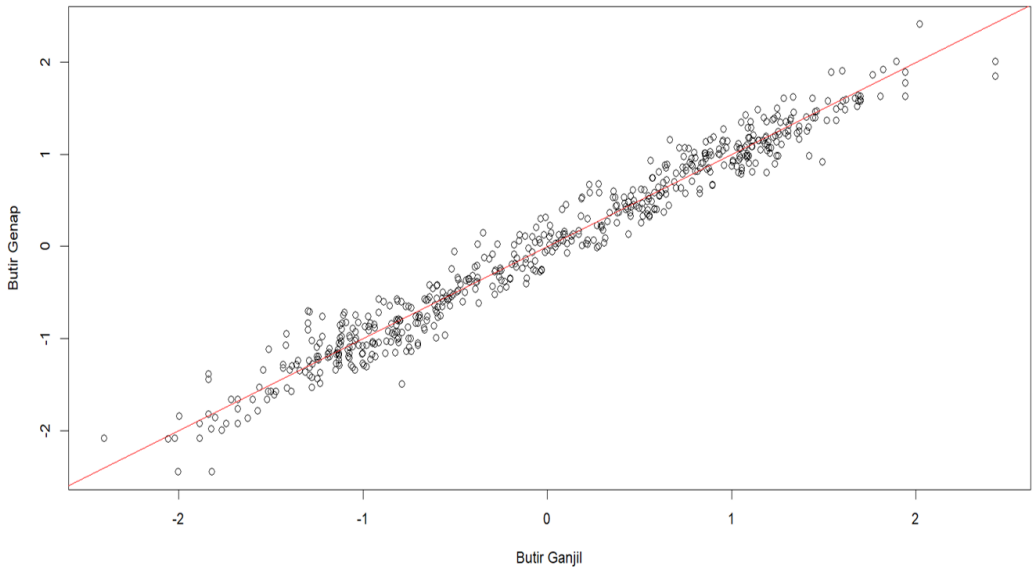

Figure 6: Invariance of Capability Parameters

Item analysis

The GPCM model has discrimination parameters and difficulty parameters. The discrimination parameter measures how well an item differentiates between people with different ability levels, while the difficulty parameter measures how difficult a question (Muraki, 1992). The value of the discrimination parameter in this study is shown in Table 4, and the difficulty value is shown in Table 5.

Item number	Discriminant Value	Item number	Discriminant Value
Q1	1.1	Q11	1.9
Q2	1.3	Q12	2.0
Q3	1.4	Q13	2.0
Q4	1.4	Q14	1.8
Q5	1.6	Q15	1.9
Q6	1.1	Q16	1.7
Q7	2.0	Q17	1.6
Q8	1.9	Q18	1.8
Q9	1.4	Q19	1.4
Q10	1.8	Q20	1.5

Table 4: Discriminant Parameter Values For Each Item

Item Number	Parameter Difficulty										
	b	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10
Q1	-0.347	-2.491	-2.119	-1.263	-1.199	-0.367	-0.116	0.293	1.000	1.263	1.525
Q2	-0.292	-2.637	-1.977	-1.353	-1.179	-0.226	-0.060	0.473	0.987	1.296	1.752
Q3	-0.310	-2.562	-1.981	-1.406	-1.014	-0.412	-0.008	0.235	0.920	1.488	1.643
Q4	-0.362	-2.901	-1.881	-1.425	-1.079	-0.396	-0.131	0.442	0.769	1.321	1.660
Q5	-0.197	-2.831	-1.755	-1.217	-0.954	-0.230	0.010	0.478	1.010	1.461	2.059
Q6	0.399	-1.267	-1.059	-0.962	-0.390	0.261	0.386	0.513	1.756	2.190	2.563
Q7	-0.039	-2.350	-1.573	-1.137	-0.589	-0.141	-0.011	0.688	0.992	1.568	2.161
Q8	-0.104	-2.349	-1.655	-1.209	-0.820	-0.126	-0.019	0.496	1.012	1.507	2.119
Q9	0.431	-1.382	-0.985	-0.940	-0.437	0.253	0.330	0.872	1.609	2.079	2.915
Q10	0.307	-1.848	-1.291	-0.844	-0.538	-0.052	0.328	0.811	1.478	2.132	2.893
Q11	0.277	-1.811	-1.171	-0.730	-0.542	-0.244	0.395	0.843	1.351	2.018	2.664
Q12	0.344	-1.862	-1.203	-0.878	-0.435	0.016	0.445	0.934	1.531	2.179	2.715
Q13	0.333	-1.900	-1.211	-0.819	-0.596	0.107	0.519	0.927	1.465	2.126	2.707
Q14	0.308	-1.854	-1.159	-0.936	-0.305	-0.221	0.410	0.728	1.333	2.047	3.032
Q15	0.233	-1.919	-1.514	-0.920	-0.445	-0.002	0.305	0.795	1.155	1.948	2.923
Q16	0.394	-1.431	-1.076	-0.990	-0.443	0.142	0.424	0.812	1.637	2.081	2.782
Q17	-0.165	-2.491	-1.723	-1.295	-0.920	-0.243	0.149	0.369	1.168	1.233	2.103
Q18	-0.120	-2.248	-1.742	-1.208	-0.742	-0.348	0.062	0.466	1.003	1.405	2.153
Q19	0.420	-1.153	-1.088	-1.060	-0.431	0.286	0.525	0.624	1.741	1.989	2.764
Q20	0.371	-1.416	-0.988	-0.903	-0.659	0.383	0.103	0.881	1.535	2.184	2.592

Table 5: the Value of the Difficulty Parameter For Each Item

Apart from the discrimination parameter and the difficulty parameter, the subsequent analysis tests the item fit of the test items that have been developed. Item fit is an essential consideration in developing and using IRT-based tests. An item is called fit if it has a p.S_X^2 value > 0.05 (Dewanti, Hadi and Nu'man, 2021). The qualified items for the instruments developed using the GPCM model are according to Table 6.

Item	S_X^2	df.S_X^2	RMSEA.S_X^2	p.S_X^2
Q1	189.061	166	0.015	0.106
Q2	171.598	148	0.017	0.090
Q3	163.302	147	0.014	0.169
Q4	149.883	150	0.000	0.487
Q5	142.553	143	0.000	0.495
Q6	181.653	166	0.013	0.192
Q7	125.160	127	0.000	0.530
Q8	140.389	130	0.012	0.252
Q9	161.297	149	0.012	0.232
Q10	148.672	137	0.012	0.234
Q11	158.512	136	0.017	0.091
Q12	157.265	132	0.018	0.066
Q13	151.752	132	0.016	0.115
Q14	164.438	138	0.018	0.062
Q15	143.844	131	0.013	0.209
Q16	166.823	142	0.017	0.076
Q17	140.685	141	0.000	0.492
Q18	142.472	134	0.010	0.292
Q19	179.288	157	0.016	0.107
Q20	140.152	141	0.000	0.504

Table 6: Test Instrument Fit Items

DISCUSSION

The nine categories in Table 3 show that the instrument construct meets the fit category so that it can be concluded that the instrument construct is proven valid (construct validity is fulfilled). According to Figure 2, the CFA model's output shows that all loading factors are positive and significant; each item measures the relevant factor (Kwahk and Lee, 2008). The standardized loading factor obtained also ranges from 0.90 to 0.96, which indicates that these items significantly contribute to factor construction or purification of construct validity; according to Igbaria et al. (1997), loading factor > 0.3 is significant, loading factor > 0.4 is more important, and loading factor ≥ 0.5 is very significant.

Calculating the estimated value of construct reliability uses construct reliability (CR) using factor loading values and unique error indexes obtained from the CFA results. The construct is reliable if the CR obtained is more significant than 0.70 (Ghozali and Fuad, 2008). The calculation results show that the CR is 0.99; this indicates that the construct is proven reliable.

Before being analyzed with GPCM, it must first test the assumptions of response theory consisting of unidimensional tests, local independence, and parameter invariance. The purpose of the unidimensional assumption test is to count the number of items or questions designed to measure a test construct that genuinely represents one dimension or the construct that is structured and does not take advantage of other unrelated dimensions. Unidimensional assumptions are essential to ensure that the scores generated from items are meaningful and reliable and that measured construct representations are valid (Hambleton and Swaminathan, 1985; Hambleton et al., 1991; Linden and Hambleton, 1997) "Fundamentals of Item Response Theory" introduces the basics of item response theory (IRT). Scree plot exploratory factor analysis in Figure 3 shows that one factor is measured in the chemical five-tier diagnostic test instrument. This shows that one dominant factor is obtained to fulfill the unidimensional assumption.

The second assumption of IRT is local independence. Local independence is an illustration if the correlation between pairs of items is only caused by the main trait or ability that is measured by a series of test items and is not influenced by some traits or abilities that are not modeled that affect the two items (Demars, 2010). According to Hambleton et al. (1991) "Fundamentals of Item Response Theory" introduces the basics of item response theory (IRT) the assumption of local independence will be automatically proven if the unidimensional test has been proven. This means that this chemical five-tier diagnostic test instrument meets the assumption of local independence.

The third assumption of IRT is the invariance of item and capability parameters. Parameter invariance is a concept in the measurement field that refers to a parameter model of consistency or similarity across groups or subpopulations (Millsap and Kwok, 2004). The invariance of the item parameters is seen from the level of difficulty and differential power because, in the GPCM IRT model, there are two parameters, namely differential power and difficulty level, so it is necessary to look at the invariance of item parameters from difficulty level and differential power. The item invariance parameters for the difficulty level and differential power are based on the distribution of students in schools with a and B accreditation. The invariance of item parameters based on the questions' differential power and difficulty level are shown in Figures 4 and 5. In addition to the item invariance parameters, the invariance parameters are also determined by dividing even and odd questions. The invariance of the ability parameters is shown in Figure 6. If each point is close to slope line 1, then this indicates that there is no parameter variation (Drasgow and Mattern, 2006; Hambleton, 2006; Hambleton and Swaminathan, 1985; Hambleton et al., 1991; Linden and Hambleton, 1997) "Fundamentals of Item Response Theory" introduces the basics of item response theory (IRT). Figures 4, 5, and 6 show that the close points are with the red line, which is slope 1. This indicates that there is no variance in the estimation result parameters.

The IRT model used in this study is the GPCM model. GPCM is an IRT polytomous model used to estimate the probability of a person responding to a test item at a certain difficulty level. The GPCM is more flexible than the Rasch model, as it allows for different levels of difficulty between items and different ability levels between people. The GPCM model has parameters different from the Rasch model; the GPCM model has discriminant parameters and parameter difficulties (Muraki, 1992).

The discrimination parameter measures how well an item discriminates between people with different ability levels. Hambleton et al. (1991) "Fundamentals of Item Response Theory" introduces the basics of item response theory (IRT) state that good grains have discrimination parameters greater than zero and less than or equal to 2. Table 4 shows that all discrimination parameters have a value of $1.1 \leq a \leq 2$, meaning that all of these items are good because they can distinguish between students with high ability and those with low ability.

The difficulty parameter in GPCM measures how difficult a question is (Muraki, 1992). Hambleton et al. (1991) "Fundamentals of Item Response Theory" introduces the basics of item response theory (IRT state that a good item difficulty index is -2 to +2. Based on Table 5, it was found that all items with difficulty parameters b1 to b10 had values from most minor to most prominent, and the average b was in the range -2 to +2; this shows that the items developed were excellent and informative about students' abilities.

The subsequent analysis is to test the fit items of the test items that have been developed. Item fit is an essential consideration in developing and using IRT-based tests. There are various statistics to determine item fit that can be used to assess item fits, such as infit statistics, outfit statistics, standardized residuals, S-X^2 fit index, and many other references. This study used the S-X^2 fit index to determine fit items because they correspond to polytomous items in educational and psychological research (Kang and Chen, 2011). An item is called fit if it has a p.S_X^2 value > 0.05 (Dewanti et al., 2021). Table 6 shows that all items from Q1 to Q20 have a p.S_X^2 value greater than 0.05; this indicates that all items developed in the five-tier chemical diagnostic test instrument fit and are by the GPCM model. Based on Table 6, all items are appropriate, but if you wish to change the number of items to be used,

removing items with a p.S_X^2 value close to 0.05 is better. For example, in this instrument, two items can be omitted, namely Q14 and Q16. Item Q14 has a p.S_X^2 value of 0.062, and Q16 has a p.S_X^2 value of 0.076. If you only want to use 18 of the 20 available items, you should delete items with a p.S_X^2 value close to 0.05.

Apart from item analysis, another essential thing to note in GPCM is test information and standard measurement errors. Test information refers to the ability of test items to distinguish between individuals with different levels of ability. In contrast, the standard error measurement refers to the amount of uncertainty associated with the estimation of individual ability, where the standard error measurement is the inverse of the square root of the test information so that the greater the information, the smaller the standard error and the greater the reliability (Demars, 2010). The information function of the test and standard error is presented in Figure 7.

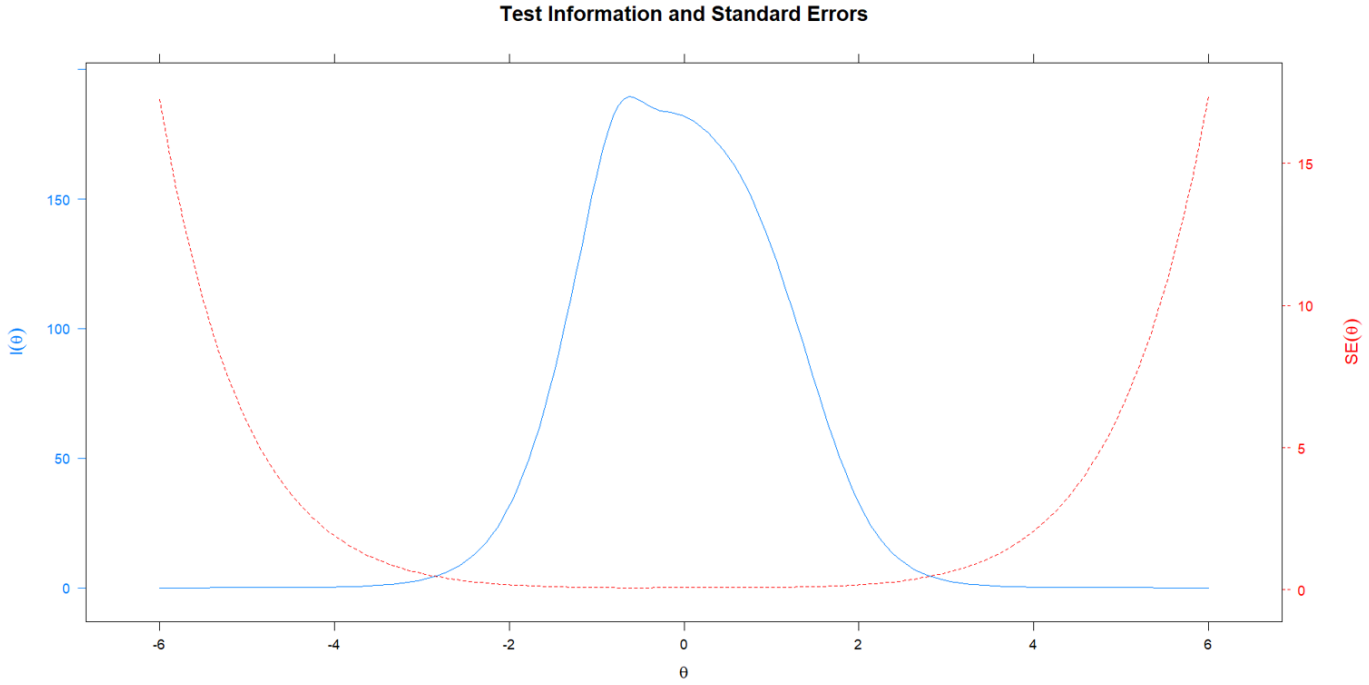

Figure 7: Information Function Test And Standard Error

Figure 7 shows that the theta obtained from the information function and standard error measurement on the CBT-based five-tier diagnostic test instrument has the intersection of the lines at the low limit with a theta of -2.85 and at the upper limit with a theta of +2.80. These results indicate that the CBT-based five-tier diagnostic test is suitable for students who have abilities between -2.85 and +2.80.

Table 5 shows that the most challenging item is item 9, and the easiest item is item 4. Item 4 contains questions about chemical problems related to electron configurations when electrons are released. This problem was made based on the consideration that there is a misconception among students who think that the 2 electrons released in $26Fe^{2+}$ come from the 3d orbital rather than the 4s orbital because they are farther from the nucleus (Kay et al., 2010). Item 9 contains questions about the analysis of the relationship between atomic number and the periodicity of elements (atomic radii) based on data on the periodicity of elements. This problem is made because all students believe that the size of the atomic radius increases down and to the right on the periodic table (Nicoll, 2001). Display of item 4 and item 9 on the computer-based test according to Figures 8 and 9.

The items' difficulty level analysis results can be interpreted into the Item Characteristic Curve (ICC). ICC is a graph that shows the relationship between the probability of correct answers from participants and their level of ability in a particular domain in the Item Response Theory model, which serves to determine the level of difficulty of a test item, determine the differentiability of a test item, and assess the quality of a test item. ICC shows the characteristics of the difficulty level of the items in the form of a curve of the relationship between the probability of answering correctly 50% and the level of student ability. The ICC item with the lowest difficulty level is item 4, which has a value of -0.362, and the greatest difficulty index is item 9, which is 0.431. Display ICC item 4 and item 9 according to Figures 10 and 11.

The ICC shown in Figures 10 and 11 shows that item 4 can be answered by students with a minimum θ ability of -2.901. In contrast, item 9 can be answered if they have a minimum θ ability of -1.382, meaning that to be able to work on item 9 a student must have higher abilities than when working on item 4.

Comparison of test results based on gender is analyzed using differential item functioning (DIF) according to Figure 12. DIF shows the probability that supports certain items between males and females (Tie, Chen and He, 2022).

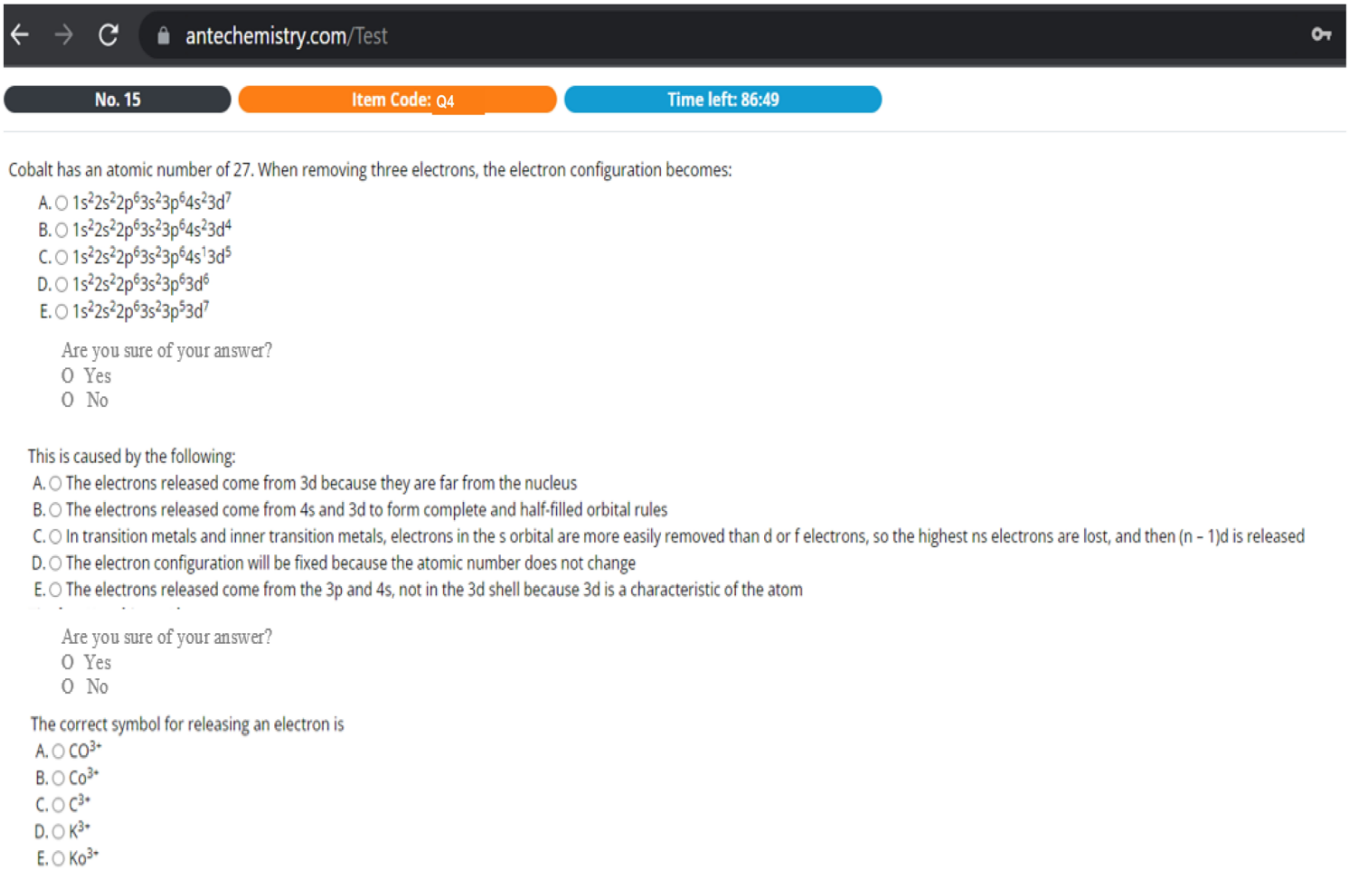

Figure 8: Display items 4

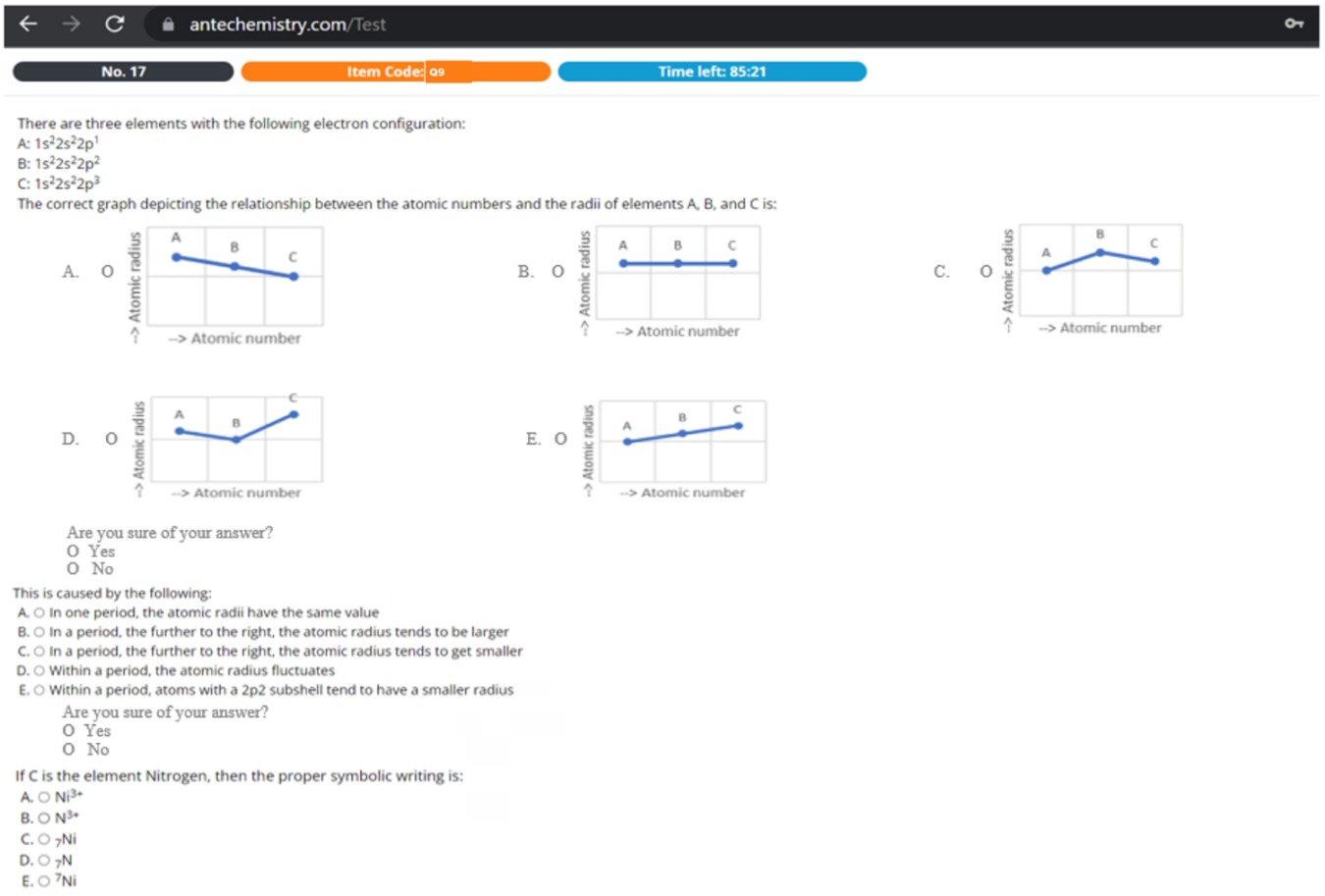

Figure 9: Display items 9

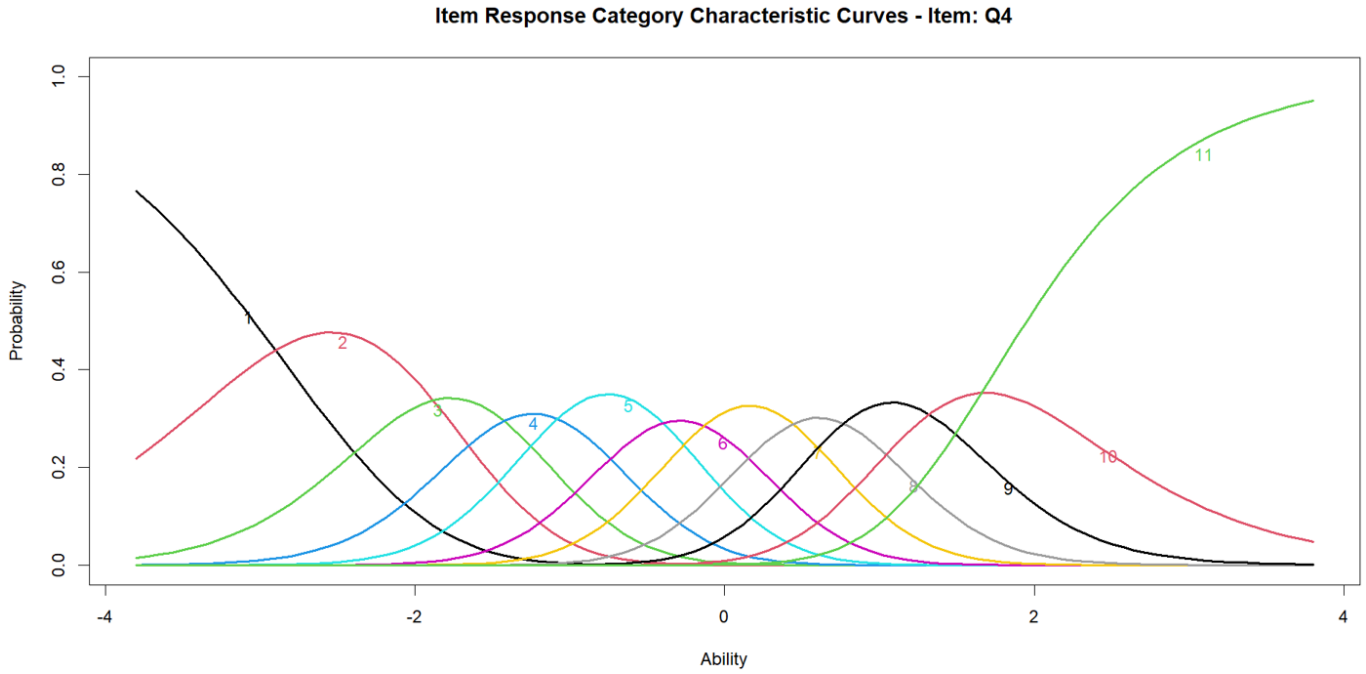

Figure 10: Item Characteristic Curve (ICC) item 4

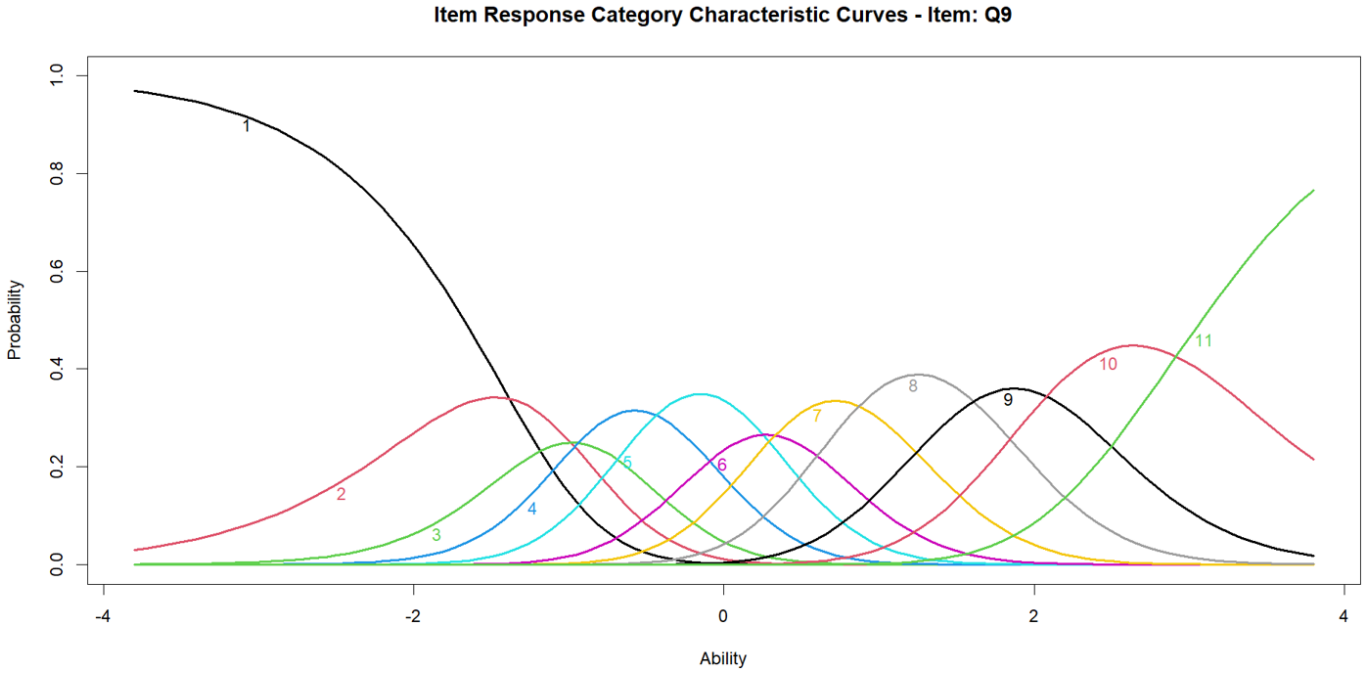

Figure 11: Item Characteristic Curve (ICC) item 9

Based on Figure 12, it can be seen that Q9 items are close to the upper limit, and Q4 items are close to the lower limit. Items close to the upper limit mean items have great difficulty, and items close to the lower limit indicate easy items. Figure 12 also shows that females find it easier to work on items in Q1, Q5, Q7, Q9, Q10, Q13, Q16, Q17, and Q20, while males find it easier to work on items in Q2, Q3, Q4, Q6, Q8, Q11, Q12, Q14, Q15, Q18, and Q19.

CONCLUSION

This research succeeded in developing a five-tier chemical diagnostic test instrument based on a computer-based test. The test instrument consists of five levels of questions: the first is the central question, the second is the confidence level, the third is the reason for the main question, the fourth is the confidence level for a reason, and the fifth is symbolic in chemistry related to the main question. The developed chemical test instruments cover three main chemistry sections: macroscopic, microscopic, and symbolic chemistry.

The test results of the test instruments on a sample of 580 students showed that the test instruments developed had good validity and reliability and could distinguish between different student abilities based on each student's ability θ. Instrument development uses the IRT approach with the GPCM model for

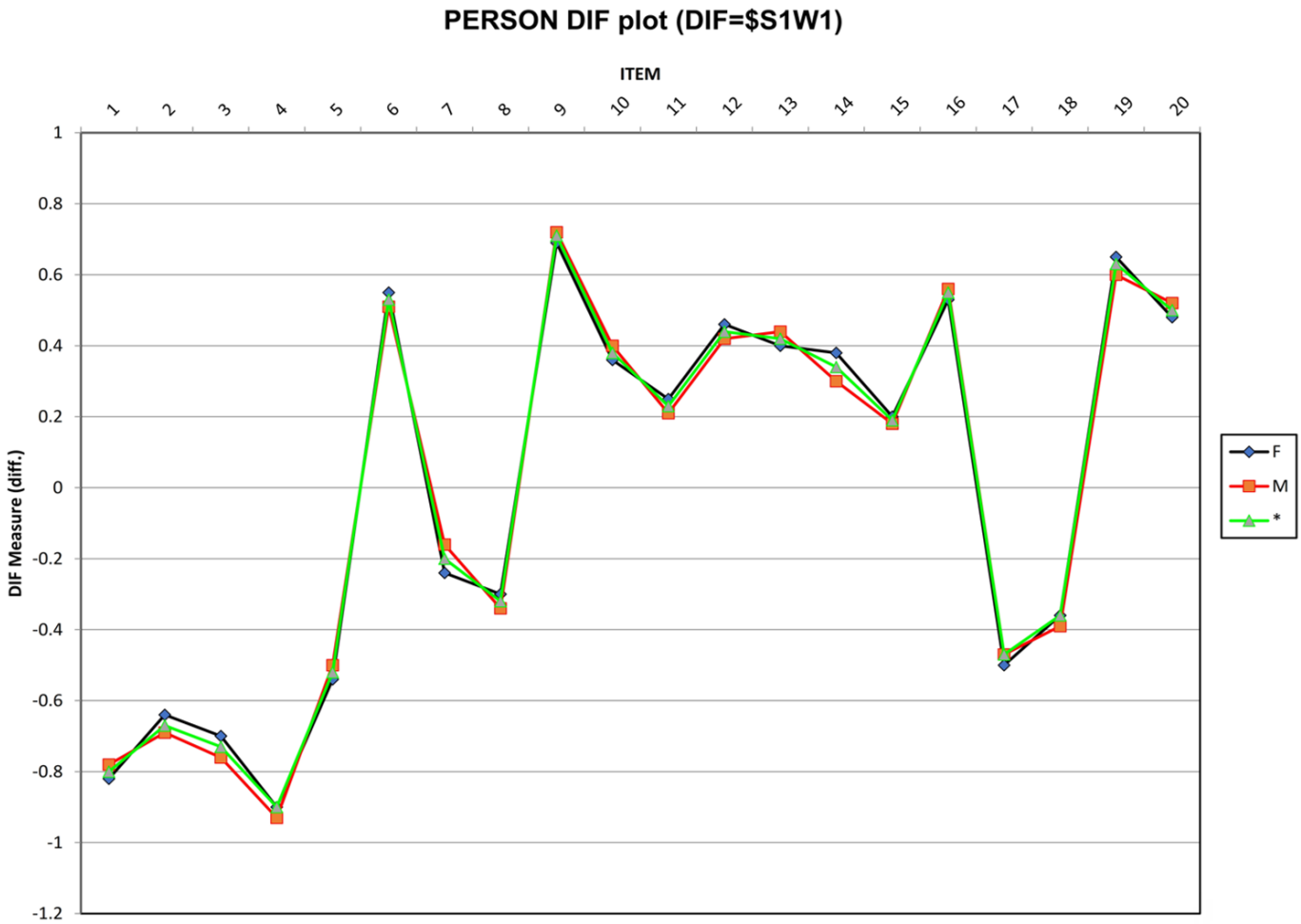

Figure 12: Differential Item Functioning (DIF) male and females

polytomous. This research makes an essential contribution to developing test instruments to detect students' misconceptions in chemistry and can be used as a reference in developing further misconception test instruments in the future.

This instrument consists of 20 items, and in one item, there are five levels of questions that can measure students' knowledge abilities which are divided into 11 ability categories consisting of no knowledge, complete misconception, partial misconception, guess knowledge, type 5 (lack of knowledge), type 5 (lack of knowledge), type 4 (lack of knowledge), type 3 (lack of knowledge), type 2 (lack of knowledge), type 1 (lack of knowledge), understand but lack confidence, and understand. This instrument has gone through content validity tests, constructs, criteria, and reliability tests. The reliability test results show a construct reliability value of 0.99, indicating that this instrument can be relied upon in measuring students' abilities, and the Aiken test score > 0.70 suggests that the instrument developed is essential and follows the curriculum.

The results of this study can be used as a reference for chemistry teachers to find out their students' abilities and find out where their students' misconceptions are. Future research can use this research as a reference in developing misconception instruments in other fields.

This study provides information about valid and reliable misconception instruments. The limitation of this research is that the research subjects are still in one country. This research can be expanded by using samples from various countries.

REFERENCES

Aiken, L. R. (1980) 'Content validity and reliability of single items or questionnaires', *Educational and Psychological Measurement*, Vol. 40, No. 4, pp. 955-959. https://doi.org/10.1177/001316448004000419

Anam, R. S., Widodo, A., Sopandi, W. and Wu, H. K. (2019) 'Developing a five-tier diagnostic test to identify students' misconceptions in science: an example of the heat transfer concepts', *Elementary Education Online*, Vol. 18, No. 3, pp. 1014-1029. https://doi.org/10.17051/ilkonline.2019.609690

Barke, H. D., Hazari, A. and Sileshi Yitbarek (2009) *Misconception in Chemistry*, Berlin: Springer.

Bayuni, T. C., Sopandi, W. and Sujana, A. (2018) 'Identification misconception of primary school teacher education students in changes of matters using a five-tier diagnostic test', *Journal of Physics: Conference Series*, Bandung, Vol. 1013, p. 012086 https://doi.org/10.1088/1742-6596/1013/1/012086

Chi, S., Wang, Z., Luo, M., Yang, Y. and Huang, M. (2018) 'Student progression on chemical symbol representation abilities at different grade levels (Grades 10-12) across gender', *Chemistry Education Research and Practice*, Vol. 19, No. 4, pp. 1055-1064. https://doi.org/10.1039/c8rp00010g

Cornick, J. E. (2015) 'Factor structure of the exercise self-efficacy scale', *Measurement in Physical Education and Exercise Science*, Vol. 19, No. 4, pp. 208-218. https://doi.org/10.1080/1091367X.2015.1074579

Crocker, L. and Algina, J. (2008) *Introduction to Classical and Modern Test Theory*, 2nd edition, Mason: Cengage Learning.

Dagnall, N., Denovan, A., Parker, A., Drinkwater, K. and Stephen Walsh, R. (2018) 'Confirmatory factor analysis of the inventory of personality organization-reality testing subscale', *Frontiers in Psychology*, Vol. 9, 1116, pp. 1-12. https://doi.org/10.3389/fpsyg.2018.01116

Debelak, R., Stobl, C. and Zeigenfuse, M. D. (2022) *An Introduction to the Rasch Model with Examples in R*, New York: CRC Press. https://doi.org/10.1201/9781315200620

Demars, C. (2010) *Item Response Theory*, New York: Oxford University Press.

Derkach, T. M. (2021) 'The origin of misconceptions in inorganic chemistry and their correction by computer modelling', *Journal of Physics: Conference Series*, Kryvyi Rih, Vol. 1840, pp. 1-13. https://doi.org/10.1088/1742-6596/1840/1/012012

Dewanti, S. S., Hadi, S. and Nu'man, M. (2021) 'The application of item response theory in analysis of characteristics of mathematical literacy test items', *İlköğretim Online*, Vol. 20, No. 1, pp. 1226-1237. https://doi.org/10.17051/ilkonline.2021.01.119

Dewi, F. C., Parlan, P. and Suryadharma, I. B. (2020) 'Development of four-tier diagnostic test for identifying misconception in chemical equilibrium', *AIP Conference Proceedings*, Vol. 2215, pp. 1-6. https://doi.org/10.1063/5.0000510

Drasgow, F. and Mattern, K. (2006) 'New tests and new items: Opportunities and issues', in Bartram, D. and Hambleton, R. K. (ed.) *Computer-Based Testing and the Internet*, Chichester: John Wiley & Sons Ltd.

Ghozali, I. and Fuad (2008) *Structural equation modeling: teori, konsep, dan aplikasi dengan Program Lisrel 8.80*, Semarang: Badan Penerbit Universitas Diponegoro.

Groen, M. M. van and Eggen, T. J. H. M. (2020) 'Educational Test Approaches: The Suitability of Computer-Based Test Types for Assessment and Evaluation in Formative and Summative Contexts', *Journal of Applied Testing Technology*, Vol. 21, No. 1, pp. 12-24.

Hambleton, R. K. (2006) 'Psychometric Models, Test Designs and Item Types for the Next Generation of Educational and Psychological Tests', in Bartram, D. and Hambleton, R. K. (ed.) *Computer Based Testing and the Internet*, Chichester: John Wiley & Sons Ltd.

Hambleton, R. K. and Swaminathan, H. (1985) *Item Response Theory: Principles and Applications*, New York: Springer Science+Business Media. https://doi.org/10.1007/978-94-017-1988-9

Hambleton, R. K., Swaminathan, H. and Rogers, H. J. (1991) *Fundamentals of Item Response Theory Library*, London: Sage Publications.

Harza, A. E. K. P., Wiji, W. and Mulyani, S. (2021) 'Potency to overcome misconceptions by using multiple representations on the concept of chemical equilibrium', *Journal of Physics: Conference Series*, Jawa Barat, Vol. 1806, pp. 1-6. https://doi.org/10.1088/1742-6596/1806/1/012197

Igbaria, M., Zinatelli, N., Cragg, P. and Cavaye, A. L. M. (1997) 'Personal computing acceptance factors in small firms: A structural equation model', *MIS Quarterly*, Vol. 21, No. 3, pp. 279-305. https://doi.org/10.2307/249498

Istiyono, E., Dwandaru, W. S. B., Erfianti, L. and Astuti, W. (2020) 'Applying CBT in physics learning to measure students' higher order thinking skills', *Journal of Physics: Conference Series*, Yogyakarta, Vol. 1440, p. 012061. https://doi.org/10.1088/1742-6596/1440/1/012061

Istiyono, Edi, Dwandaru, W. S. B., Fenditasari, K., Ayub, M. R. S. S. N. and Saepuzaman, D. (2023) 'The development of a four-tier diagnostic test based on modern Test theory in physics education', *European Journal of Educational Research*, Vol. 12, No. 1, pp. 371-385. https://doi.org/10.12973/eu-jer.12.1.371

Jusniar, J., Effendy, E., Budiasih, E. and Sutrisno, S. (2021) 'Eliminating misconceptions on reaction rate to enhance conceptual understanding of chemical equilibrium using EMBE-R strategy', *International Journal of Instruction*, Vol. 14, No. 1, pp. 85-104. https://doi.org/10.29333/IJI.2021.1416A

Kang, T. and Chen, T. T. (2011) 'Performance of the generalized S-X2 item fit index for the graded response model', *Asia Pacific Education Review*, Vol. 12, No. 1, pp. 89-96. https://doi.org/10.1007/s12564-010-9082-4

Kay, C. C., Yiin, H. K., Chu, C. K. and Hong, K. Y. (2010) 'Misconceptions in the teaching of chemistry in secondary schools in Singapore & Malaysia', *Proceedings of the Sunway Academic Conference*, Petaling Jaya, No. 1996, pp. 1-10.

Kimmons, R., Clark, B. and Lim, M. (2017) 'Understanding web activity patterns among teachers, students and teacher candidates', *Journal of Computer Assisted Learning*, Vol. 33, No. 6, pp. 588-596. https://doi.org/10.1111/jcal.12202

Kiray, S. A. and Simsek, S. (2020) 'Determination and evaluation of the science teacher candidates' misconceptions about density by using four-tier diagnostic test', *International Journal of Science and Mathematics Education*, Vol. 19, No. 5, pp. 935-955. https://doi.org/10.1007/s10763-020-10087-5

Kwahk, K. Y. and Lee, J. N. (2008) 'The role of readiness for change in ERP implementation: Theoretical bases and empirical validation', *Information and Management*, Vol. 45, No. 7, pp. 474-481. https://doi.org/10.1016/j.im.2008.07.002

Lamichhane, R., Reck, C. and Maltese, A. V. (2018) 'Undergraduate chemistry students' misconceptions about reaction coordinate diagrams', *Chemistry Education Research and Practice*, Vol. 19, No. 3, pp. 834-845. https://doi.org/10.1039/c8rp00045j

Linden, W. J. van der (2018) *Handbook of Item Response Theory Volume Three: Applications*, Boca Raton: CRC Press.

Linden, W. J. and Hambleton, R. K. (1997) *Handbook of Modern Item Response Theory*, New York: Springer.

Llanos, J., Fernández-Marchante, C. M., García-Vargas, J. M., Lacasa, E., De La Osa, A. R., Sanchez-Silva, M. L., De Lucas-Concuegra, A., Garcia, M. T. and Borreguero, A. M. (2021) 'Game-based learning and just-in-time teaching to address misconceptions and improve safety and learning in laboratory activities', *Journal of Chemical Education*, Vol. 98, No. 10, pp. 3118-3130. https://doi.org/10.1021/acs.jchemed.0c00878

Lowyck, J. (2014) *Handbook of Research on Educational Communications and Technology* New York: Springer. https://doi.org/10.1007/978-1-4614-3185-5

Marsh, H. W., Balla, J. R. and Mcdonald, R. P. (1988) 'Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size', *Psychological Bulletin*, Vol. 103, No. 3, pp. 391-410. https://doi.org/10.1037/0033-2909.103.3.391

Martin, F. and Betrus, A. K. K. (2019) *Digital media for learning: Theories, processes, and solutions digital media for learning: Theories, processes, and solutions*, Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-030-33120-7

Merriënboer, J. J. G. van and Bruin, A. B. H. de (2014) 'Research Paradigms and Perspectives on Learning', in Spector, J. M., Merrill, M. D., Elen, J. and Bishop M. J. (ed.) *Handbook of Research on Educational Communications and Technology*, New York: Springer.

Mills, C. n. and Breithaupt, K. J. (2016) *Educational Measurement From Foundations to Future*, New York: The Guilford Press.

Millsap, R. E. and Kwok, O. M. (2004) 'Evaluating the impact of partial factorial invariance on selection in two populations', *Psychological Methods*, Vol. 9, No. 1, pp. 93-115. https://doi.org/10.1037/1082-989X.9.1.93

Muraki, E. (1992) 'A generalized partial credit model: Application of an EM algorithm', *Applied Psychological Measurement*, Vol. 16, No. 2, pp. 159-176. https://doi.org/10.1177/014662169201600206

Murniningsih, Muna, K. and Irawati, R. K. (2020) 'Analysis of misconceptions by four tier tests in electrochemistry, case study on students of the chemistry education study program UIN Antasari Banjarmasin', *Journal of Physics: Conference Series*, Vol. 1440, pp. 1-6. https://doi.org/10.1088/1742-6596/1440/1/012008

Mutlu, A. and Sesen, B. A. (2015) 'Development of a two-tier diagnostic test to assess undergraduates' understanding of some chemistry concepts', *Procedia - Social and Behavioral Sciences*, Vol. 174, pp. 629-635. https://doi.org/10.1016/j.sbspro.2015.01.593

Nicoll, G. (2001) 'A report of undergraduates' bonding misconceptions', *International Journal of Science Education*, Vol. 23, No. 7, pp. 707-730. https://doi.org/10.1080/09500690010025012

Oriondo, L. L. and Antonio, E. M. D. (1984) *Evaluating Educational Outcomes (Tests, Measurement and Evaluation)*, Manila: Rex Book Store.

Özmen, H. (2004) ‹Some student misconceptions in chemistry: A literature review of chemical bonding', *Journal of Science Education and Technology*, Vol. 13, No. 2, pp. 147-159. https://doi.org/10.1023/b:jost.0000031255.92943.6d

Pokorný, M. (2023) 'Experience with online learning of mathematics in primary education', *International Journal of Emerging Technologies in Learning (IJET)*, Vol. 18, No. 02, pp. 203-213. https://doi.org/10.3991/ijet.v18i02.35401

Potvin, P. (2023) 'Response of science learners to contradicting information: A review of research', *Studies in Science Education*, Vol. 59, No. 1, pp. 67-108. https://doi.org/10.1080/03057267.2021.2004006

Prodjosantoso, A. K., Hertina, A. M. and Irwanto (2019) 'The misconception diagnosis on ionic and covalent bonds concepts with three tier diagnostic test', *International Journal of Instruction*, Vol. 12, No. 1, pp. 1477-1488. https://doi.org/10.29333/iji.2019.12194a

Putra, A. S. U., Hamidah, I. and Nahadi (2020) 'The development of five-tier diagnostic test to identify misconceptions and causes of students' misconceptions in waves and optics materials', *Journal of Physics: Conference Series*, Vol. 1521, No. 2. https://doi.org/10.1088/1742-6596/1521/2/022020

Rupp, A. A. and Zumbo, B. D. (2006) 'Understanding parameter invariance in unidimensional IRT models', *Educational and Psychological Measurement*, Vol. 66, No. 1, pp. 63-84. https://doi.org/10.1177/0013164404273942

Rusmini, Suyono, Jatmiko, B. and Yonata, B. (2021) 'The diagnosis of misconception on the concept of acid-base theory in prospective teacher students used a three-tier test', *Journal of Physics: Conference Series*, Vol. 1899, No. 1, pp. 1-7. https://doi.org/10.1088/1742-6596/1899/1/012061

Suparman, A. R., Rohaeti, E. and Wening, S. (2022) 'Development of attitude assessment instruments towards socioscientific issues in chemistry learning', *European Journal of Educational Research*, Vol. 11, No. 4, pp. 1947-1958. https://doi.org/10.12973/eujer.11.4.1947

Suparman, A. R., Rohaeti, E. and Wening, S. (2023) 'Effect of computer based test on motivation: A meta-analysis', *European Journal of Educational Research*, Vol. 12, No. 4, pp. 1583-1599. https://doi.org/10.12973/eu-jer.12.4.1583

Tie, B., Chen, G. and He, J. (2022) 'Validation of the inflexible eating questionnaire in a large sample of Chinese adolescents: psychometric properties and gender-related differential item functioning', *Eating and Weight Disorders*, Vol. 27, No. 3, pp. 1029-1037. https://doi.org/10.1007/s40519-021-01239-9

Tien, L. T. and Osman, K. (2017) *Overcoming Students' Misconceptions in Science*, Singapore: Springer Nature. https://doi.org/10.1007/978-981-10-3437-4

Treagust, D., Chittleborough, G. and Mamiala, T. (2003) 'The role of submicroscopic and symbolic representations in chemical explanations', *International Journal of Science Education*, Vol. 25, No. 11, pp. 1353-1368. https://doi.org/10.1080/0950069032000070306

Üce, M. and Ceyhan, İ. (2019) ‹Misconception in chemistry education and practices to eliminate them: Literature analysis', *Journal of Education and Training Studies*, Vol. 7, No. 3, pp. 202. https://doi.org/10.11114/jets.v7i3.3990

Vladusic, R., Bucat, R. B. and Ozic, M. (2022) 'Understanding covalent bonding - a scan across the Croatian education system', *Chemistry Education Research and Practice*, Vol. 23, No. 4. https://doi.org/10.1039/D2RP00039C

Wang, F., Franco-Penya, H.-H., Kelleher, J. D., Pugh, J. and Ross, R. (2017) 'An analysis of the application of simplified silhouette to the evaluation of k-means clustering validity', *International Conference Machine Learning and Data Mining in Pattern Recognition*, Vol. 13, pp. 291-305. https://doi.org/10.1007/978-3-319-62416-7_21

Wang, Z., Yan, W., Zeng, C., Tian, Y. and Dong, S. (2023) 'A unified interpretable intelligent learning diagnosis framework for learning performance prediction in intelligent tutoring systems', *International Journal of Intelligent Systems*, Vol. 2023, pp. 1-20. https://doi.org/10.1155/2023/4468025

Widarti, H. R., Permanasari, A., Mulyani, S., Rokhim, D. A. and Habiddin (2021) 'Multiple representation-based learning through cognitive dissonance strategy to reduce student's misconceptions in volumetric analysis', *TEM Journal*, Vol. 10, No. 3, pp. 1263-1273. https://doi.org/10.18421/TEM103-33