

**INTERNATIONAL JOURNAL OF
EDUCATION, PSYCHOLOGY
AND COUNSELLING
(IJEPC)**www.ijepe.com**PROFILING STUDENTS' READINESS FOR THE FUTURE:
CLUSTER ANALYSIS OF 21ST-CENTURY COMPETENCIES
AND CAREER INTERESTS AMONG SECONDARY STUDENTS**Norin Rahayu Shamsuddin^{1,2*}, Fazillah Bosli³, Siti Nur Alwani Salleh⁴¹ Mathematical Sciences Studies, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA (UiTM) Cawangan Kedah, Kampus Sungai Petani, Malaysia² Integrated Simulation & Visualization Research Interest Group, Universiti Teknologi MARA (UiTM) Cawangan Kedah, Kampus Sungai Petani, Malaysia

Email: norinrahayu@uitm.edu.my

³ Mathematical Sciences Studies, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA (UiTM) Cawangan Kedah, Kampus Sungai Petani, Malaysia

Email: fazillah@uitm.edu.my

⁴ Mathematical Sciences Studies, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA (UiTM) Cawangan Kedah, Kampus Sungai Petani, Malaysia

Email: sitinuralwani@uitm.edu.my

* Corresponding Author

Article Info:**Article history:**

Received date: 14.01.2025

Revised date: 23.01.2025

Accepted date: 27.02.2025

Published date: 20.03.2025

To cite this document:

Shamsuddin, N. R., Bosli, F., & Salleh, S. N. A. (2025). Profiling Students' Readiness for The Future: Cluster Analysis of 21st-Century Competencies and Career Interests Among Secondary Students. *International Journal of Education, Psychology and Counseling*, 10 (57), 724-741.

DOI: 10.35631/IJEPC.1057047**Abstract:**

The growing emphasis on 21st-century competencies has highlighted the need to understand how these skills influence students' academic trajectories and career aspirations, particularly within Science, Technology, Engineering, and Mathematics (STEM) education. Despite the recognized importance of competencies such as leadership, collaboration, adaptability, and time management, limited research has explored how these self-perceived skills relate to students' future career interests. This study aims to investigate the relationship between secondary school students' self-assessed 21st-century competencies and their expressed STEM career preferences. A consensus clustering approach with q-fold cross-validation was applied to categorize students into meaningful groups based on their responses to self-assessment surveys measuring leadership, collaboration, adaptability, and career aspirations. The internal cluster validity indices (ICVs) and heatmap visualizations were used to determine the optimal number of clusters and to interpret differences in learning competencies and career interests. The results revealed two distinct student clusters. Cluster 1 exhibited high self-efficacy in leadership, collaboration, and organizational skills, coupled with focused career interests in physics, mathematics, and chemistry. In contrast, Cluster 2 displayed moderate confidence in 21st-century learning skills, particularly in leadership and time management, but showed a broader and more diverse

This work is licensed under [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)

interest in STEM careers, including environmental sciences, medicine, and veterinary sciences. These findings suggest that students with higher self-perceived competencies are more likely to pursue specialized STEM fields, while those with lower self-efficacy may require targeted interventions to enhance their leadership and organizational skills. This study has significant educational implications, emphasizing the need for tailored interventions that foster essential competencies, enhance STEM career motivation, and address disparities in career readiness. Future research should explore longitudinal effects of these profiles on students' academic and career trajectories.

Keywords:

21st-Century Skill, Career Interests, STEM Education, Cluster Analysis, Secondary School.

Introduction

In the rapidly evolving landscape of the 21st century, the development of competencies that equip students to navigate complex global challenges has become paramount (Angeli & Valanides, 2015). These competencies, often referred to as 21st-century skills, encompass a range of cognitive, social, and emotional abilities necessary for success in both personal and professional domains. These include critical thinking, problem-solving, creativity, collaboration, digital literacy, and adaptability, among others (Voogt & Roblin, 2012; van Laar et al., 2017). Alongside these competencies, fostering career awareness and interest, particularly in Science, Technology, Engineering, and Mathematics (STEM) fields, is crucial for addressing the growing demand for skilled professionals in these areas (Brief et al., 2012). The intersection of these two dimensions—21st-century competencies and STEM career interests—forms a vital foundation for preparing secondary school students for their future endeavours.

The 21st century has transformed expectations for education and workforce preparation, making the acquisition of competencies such as critical thinking, problem-solving, digital literacy, and adaptability crucial for students. These skills form the foundation for thriving in a globalized and technology-driven economy, enabling individuals to navigate complex work environments and respond to rapidly changing job markets (Magpantay & Pasia, 2022; Songer et al., 2019). Furthermore, education systems are now expected to integrate these competencies into curricula to better prepare students for interdisciplinary roles that demand creativity and collaboration. Aligning educational practices with these demands is essential for equipping students to meet the challenges and opportunities of future careers (Guaman-Quintanilla et al., 2022; Songer et al., 2019; Cebrián & Pubill, 2015).

Despite the recognized importance of these elements, there remains a need to better understand how they interact within individual students and across diverse student populations. This study aims to address this gap by profiling students' students' for the future through an analysis of their self-perceived 21st-century competencies and career interests. Specifically, we employ cluster analysis to identify distinct groups of students based on their competencies and interests. This methodological approach allows for a nuanced understanding of the various ways in which students are prepared—or not—for the demands of the modern world.

Understanding the profiles of students' competencies and career interests can significantly inform educational strategies and interventions aimed at enhancing their preparedness for the future. By identifying distinct clusters of students with similar competency and interest profiles, educators and policymakers can tailor their efforts more effectively. For instance, targeted programs can be developed to bolster specific competencies or to encourage interest in underrepresented career fields. Moreover, recognizing the diversity in students' profiles can help in creating inclusive educational environments that cater to a wide range of needs and aspirations (Chen et al., 2024).

Despite these advancements, a significant gap remains between the development of 21st-century competencies and students' ability to translate them into meaningful career pathways, particularly within STEM fields. Many students lack clarity about how their acquired skills align with real-world applications or specific career trajectories (Johnson, 2000; Financial Industry Collective Outreach, 2023, Karim et al., 2024). This disconnect often results in mismatched career aspirations, underutilization of developed competencies, and unmet industry demands, ultimately hindering both individual success and societal progress. Addressing this issue requires innovative approaches to better understand and profile students for the future.

A critical challenge in preparing students for future success lies in fostering a clear understanding of the connection between essential 21st-century learning competencies and the diverse landscape of STEM careers (Songer et al., 2019; Nuangchalem et al., 2020; Karimi & Piña, 2021; Roehrig et al., 2021). This disconnect can lead to students developing valuable skills without a clear vision of their practical application within their chosen field. As a result, students may find it challenging to make informed decisions about their educational and career paths, which could hinder their ability to succeed in the rapidly evolving STEM workforce. This misalignment also poses a risk of creating a gap between student preparedness and actual industry demands, impacting both individual career success and the overall competitiveness of the STEM workforce (Hirudayaraj et al., 2021; López et al., 2023).

Motivated by these challenges, this research aims to identify and analyse distinct groups of secondary school students based on their self-perceived 21st-century learning competencies and their expressed career interests within STEM fields. By examining these two key dimensions, the study seeks to understand how students' self-perceptions of their own skills relate to their career aspirations. This clustering approach allows for a more nuanced understanding of student readiness for the future, providing insights into the various combinations of competencies and interests that characterize different student profiles. The findings can inform the development of targeted educational strategies and interventions that enhance students' preparedness for their future careers, particularly in STEM fields. Additionally, the study's results guide educators in creating personalized learning experiences that cater to the diverse needs and aspirations of secondary school students.

The remainder of this paper is structured as follows: Section 2 presents the literature review. Section 3 discussed on the methodology used in the study, including the data collection process and the application of cluster analysis. Section 4 discusses the results of the cluster analysis, highlighting the distinct groups of students identified based on their competencies and interests. Section 5 offers a discussion of the findings, their implications for educational practice, and potential directions for future research. Finally, Section 6 concludes the paper with a summary of the key points and the overall contribution of the study.

Literature Review

21st-Century Competencies and Pedagogical Approaches

Advances in science and technology, along with the need for future job skills, highlight the importance of lifelong learning in helping students develop 21st century skills. These skills have been classified into three main areas: the cognitive domain, the intrapersonal domain, and the interpersonal domain (Haug & Mork, 2021). Frameworks for the 21st century, including technological, pedagogical, contextual, and humanistic aspects, provide strategies to identify the skills students need for the future career (González-Pérez & Ramírez, 2022). These competencies are not only essential for personal development but also for preparing students to meet the challenges of the future workforce.

Pedagogical approaches play a crucial role in fostering these competencies. Traditional teaching methods often fall short in equipping students with the necessary skills for the modern era. Therefore, educators need to adopt innovative strategies that promote active learning, student engagement, and real-world application of knowledge. For instance, project-based learning (PBL) has been shown to effectively integrate various competencies by immersing students in authentic tasks that require them to apply their knowledge in practical scenarios (Barth et al., 2007).

Moreover, the integration of technology in education has opened new avenues for enhancing pedagogical practices. Digital tools and platforms can facilitate collaborative learning, provide instant feedback, and create personalized learning experiences tailored to individual student needs. However, it is important to note that the effectiveness of these technologies depends on how they are implemented and whether they align with the intended learning outcomes (Volman et al., 2020).

Trends in Career Interests

Choosing a career is a major challenge for students because it depends on many interconnected factors such as personal traits, socioeconomic status, career interest, and institutional factors (Abdu-Raheem, 2015; Rafiq et al., 2013; Hadiyati & Astuti, 2023; Ojukwu & Ali, 2020). As the global economy increasingly relies on STEM-related occupations, understanding the factors that shape students' career interests and self-efficacy in these fields is critical. Research indicates that students' confidence in their abilities, motivation to succeed, and self-regulated learning behaviours are central to fostering career interest in STEM disciplines. Self-perceived 21st-century learning competencies, such as leadership, teamwork, decision-making, adaptability, and goal setting, are closely linked to students' self-efficacy and success in STEM pathways. These skills help students navigate complex problem-solving tasks, which are essential for STEM careers.

A key determinant of students' engagement in STEM careers is self-efficacy, which refers to their confidence in performing academic and problem-solving tasks in these domains. Chan (2022) highlighted that gender disparities in STEM self-efficacy arise due to cultural and societal norms, with male students exhibiting greater confidence in technical and engineering fields than their female counterparts. This is further reinforced by Charlesworth and Banaji (2019), who found that implicit gender biases contribute to differences in career preferences, leading to underrepresentation of women in engineering and technology disciplines. These findings align with research on student perceptions of their learning abilities, as confidence in

producing high-quality work and working well with diverse peers are crucial for overcoming social and structural barriers in STEM education.

Another significant factor in STEM career interest is students' ability to manage their learning independently. Self-regulated learning skills, such as time management and prioritization of assignments, are associated with higher engagement in STEM-related activities. Research indicates that students who develop effective self-regulation strategies exhibit higher motivation and persistence in STEM subjects, as these skills enable them to navigate complex problem-solving tasks and adapt to challenging learning environments (Blackmore et al., 2021). Studies suggest that students with strong self-management skills are more likely to pursue careers in physics, mathematics, and engineering, where problem-solving and critical thinking are integral (Hsu et al., 2021). Research also indicates that career interest varies by specific STEM fields, with biology and medical sciences attracting a greater proportion of female students, while fields such as computer science and electrical engineering remain male-dominated (Fang et al, 2021).

Nowadays, many occupations require a strong foundation in science, technology, engineering, and mathematics (STEM). Koyunlu Ünlü and Dökme (2020) conducted a multivariate assessment of middle school students' interest in STEM careers in Turkey, considering factors such as gender, location, grade levels, end-of-semester grades, and parents' educational status and income levels. They found that students' interest in STEM careers varies significantly based on gender, geographical location, and grade level, while no significant differences were observed regarding parents' educational status or family income levels. Other studies reveal that career interests differ between men and women. Men tend to prefer more realistic, artistic, and enterprising fields, whereas women prefer social and conventional fields (Mudhar et al., 2020). Additionally, gender differences play a significant role, with males showing more interest in engineering and technology, and females being more inclined toward health and science-related careers (Ribeirinha et al., 2023).

Profiling and Clustering Analysis in STEM

Profiling and cluster analysis play a significant role in STEM education research by helping educators and researchers understand student behavior, academic performance, and engagement patterns. Clustering techniques enable the identification of distinct groups within student populations, thus aiding in the development of targeted interventions and teaching strategies. Various studies have utilized clustering algorithms to explore different aspects of STEM education, ranging from student academic performance to engagement in active learning environments.

A study by Denaro (2021) investigated the use of clustering algorithms in STEM education through the Classroom Observation Protocol for Undergraduate STEM (COPUS). The study employed cluster analysis to profile student and instructor behaviours, offering valuable insights into teaching practices and their impact on student engagement. In another study, educational data clustering was applied to secondary school sensor-based engineering courses, focusing on the impact of active learning approaches on students' general engineering knowledge. The authors demonstrated that active learning strategies significantly enhanced student engagement and understanding in STEM subjects (Panskyi et al., 2024).

Clustering has also been used to classify students based on academic performance. A study by Ahmad et al. (2022) utilized clustering algorithms to categorize B40 (bottom 40% income group) students in higher education institutions, identifying patterns that could inform academic support strategies. Similarly, optimized clustering techniques were applied to understand how students conceptualize science ideas, providing insights into the varied pathways students take to integrate scientific concepts (Obaid et al., 2023).

Despite the breadth of research, gaps remain in profiling secondary school students' self-perceived 21st-century learning competencies and their expressed career interests within STEM fields. Most studies focus on higher education or specific aspects such as academic performance or engagement in active learning but overlook the comprehensive profiling of secondary school students. This study aims to fill that gap by identifying and analyzing distinct groups of secondary school students based on their self-perceived competencies and career interests, using advanced clustering algorithms.

Table 1: Comparison of Clustering Methods, Findings, and Research Gaps in STEM Education (2020–2024)

Authors	Year	Title	Clustering Method	Main Findings	Limitations
Denaro	2021	Comparison of Cluster Analysis Methodologies for STEM Education	Cluster Analysis (COPUS)	Profiled student and instructor behaviors to improve teaching strategies.	Focused on undergraduate students and instructors.
Ahmad et al.	2022	Clustering Analysis for Classifying Student Academic Performance	K-Means, Hierarchical Clustering	Classified B40 students based on academic performance.	Focused on higher education; lacks competency profiling.
Panskyi et al.	2023	Educational Data Clustering in Secondary School Engineering	K-Means, DBSCAN	Explored impact of active learning on engineering knowledge.	Focused on specific STEM subjects.
Obaid et al.	2023	Using Optimized Clustering to Identify Students' Science Ideas	Optimized Clustering	Identified diverse pathways in students' understanding of science.	Lacks focus on competencies and career interests.

Methodology

This study employed a quantitative approach using cluster analysis to profile students for the future by examining their 21st-century learning competencies and career interests. Data were collected through a survey administered to 120 secondary school students from four different schools in Kuala Muda, Kedah, during a two-day MATHWIZ CHALLENGES held at UiTM

Kedah in collaboration with Jabatan Pendidikan Negeri Kedah (JPN Kedah). The questionnaire was distributed during the program.

Notably, the selection of schools was solely determined by Jabatan Pendidikan Negeri Kedah (JPN Kedah). These participants responded to questions assessing their attitudes toward these competencies and their interest in 12 broad STEM career categories. Table 2 display the information of the students from four schools.

Table 2: Respondents information

	Number of students	Form
SMK Bandar Sungai Petani	30	3
SMK Amanjaya	30	
SMK Bedong	30	
SMK Gurun	30	

The survey instrument included the following key components adopted from the *Student Attitudes toward STEM* survey, developed by the Friday Institute for Educational Innovation (Friday Institute for Educational Innovation, 2012):

- three section on student attitudes towards mathematics, science, engineering, and technology respectively.
- students' students' perceived toward 21st century learning – consists of items measuring students confidence in communication, collaboration, and self-directed learning.
- interests in STEM career categories – measured on a four-point Likert scale (1 = Not at all interested to 4 = Very interested), covering 12 broad STEM-related fields which are physic, environmental work, biology and zoology, veterinary work, mathematics, medicine, earth science, medical science, chemistry, energy, and engineering .

The first four sections employed five-point Likert scale ranging from "strongly disagree" to "strongly agree". However" this study only focuses on the students' students' perceived toward 21st century learning and student interest in STEM careers. The items under students' students' perceived toward 21st century learning are listed as in Table 3:

Table 3: The 21st century learning

Code	Description
A1	I am confident I can lead others to accomplish a goal.
A2	I am confident I can encourage others to do their best.
A3	I am confident I can produce high quality work.
A4	I am confident I can respect the differences of my peers.
A5	I am confident I can help my peers.
A6	I am confident I can include others' peothers'ves when making decisions.
A7	I am confident I can make changes when things do not go as planned.
A8	I am confident I can set my own learning goals.
A9	I am confident I can manage my time wisely when working on my own
A10	When I have many assignments, I can choose which ones need to be done first
A11	I am confident I can work well with students from different backgrounds.

Data Analysis**Cluster Analysis**

To identify the optimal number of clusters for profiling students based on their learning competencies and career interests, a consensus clustering approach was utilized. This method enhances stability and robustness by aggregating results from multiple clustering iterations.

(i) Data Partitioning and Clustering Procedure

The analysis considered a range of potential cluster numbers, $k = 2$ to $k = 10$. The dataset was partitioned using a q-fold cross-validation strategy to ensure a comprehensive exploration of the clustering structure. For each fold, clustering was performed using the k -medoids algorithm with Gower's distance the similarity measure. Gower's distance selected for its capability to handle mixed data types (continuous and categorical), which were characteristic of the dataset. The medoids were randomly initialized in each run to account for variability and avoid local optima.

(ii) Determination of Optimal k

The determination of the optimal number of clusters (k) was based on the evaluation of internal cluster validity indices (ICVIs). These indices included metrics such as the Generalized Dunn's IndDunn's I_{33}), Silhouette Index (Sil), Pakhira-Bandyopadhyay-Maulik Index (PBM), Point-Biserial Index (PB) and Wemmert_Gańcarski Index (WG) which collectively assess the compactness, separation, and overall quality of the clusters (Arbelaitz et al., 2013; Liu et al., 2010; Saitta et al., 2008, 2007). The formulation of the ICVs is presented in Table 4.

Table 4: Internal clustering validation indexes.

ICVs	Equation
Generalized Dunn's Index (GDI ₃₃)	$GDI_{33} = \frac{\min_{y \neq y^*} \left(\frac{1}{ C C' } \sum_{y \in C} d(y_i, y_i^*) \right)}{\max_c \left(2 \left(\frac{\sum_{y \in C} d(y_i, \bar{v})}{ C } \right) \right)}$ $\bar{v} = \frac{1}{ C } \sum_{y \in C} y$
Silhouette Index (Sil)	$Sil = \frac{1}{K_{\exists}} \sum_{k=1}^{K_{\exists}} \left(\frac{1}{ C } \sum_{y \in C} \frac{\min_{c \neq c'} \frac{1}{ C' } \sum_{y^* \in C'} d(y_i, y_i^*) - \frac{1}{ C - 1} \sum_{y \in C} d(y_i, y_i')}{\max \left(\frac{1}{ C - 1} \sum_{y \in C} d(y_i, y_i'), \min_{c \neq c'} \frac{1}{ C' } \sum_{y^* \in C'} d(y_i, y_i^*) \right)} \right)$
Pakhira- Bandyopadh yay-Maulik Index (PBM)	$PBM = \left(\frac{1}{K_{\exists}} \times \frac{E_1}{\sum_{k=1}^{K_{\exists}} \sum_{y \in C} d(y_i, C_{\partial})} \times \left[\max_{\partial, \partial'} C_{\partial}, C'_{\partial'} \right] \right)^2$

Point-
Biserial
Index
(PB)

$$PB = \left(\frac{\sum_{k=1}^{K_3} \sum_{y_i, y_{i'} \in C} d(y_i, y_{i'})}{N_W} - \frac{\sum_{C, C'} \sum_{y \in C, y^* \in C'} d(y_i, y_{i'})}{N_B} \right) \times \frac{\sqrt{N_W \times N_B}}{N_T}$$

Wemmert-
Gańczarski
Index (WG)

$$WB = \frac{1}{N} \sum_{k=1}^{K_3} \max \left\{ 0, n_c - \sum_{y \in C} \left(\frac{d(y_i, C_\partial)}{\min_{C' \neq C} [d(y_i, C')] } \right) \right\}$$

C_∂ – cluster centre; N_W the number of within cluster pairs of points; N_B – the number of between cluster pairs of points; N_T – total number of pairs of points; K_3 – number of clusters in base partition of reference point.

A voting-based aggregation method was employed, where each ICV contributed a weighted vote to identify the most suitable k across all folds. This consensus approach ensured a robust and replicable selection process.

Overall, the rule for determining the best number of clusters in each fold for the five ICVs is summarised in Table 5, where the max refers to maximum score value obtained by each index. The PB index can take either a positive or negative value. However, this index identified the number of clusters based on maximum value regardless of the magnitude, which indicates the magnitude of the index curve. Therefore, in this study, the maximum decrease in the curve is of interest with regards to the value of the PB index. The selected k was then used to perform the final clustering on the full dataset, revealing distinct student profiles based on their 21st-century learning competencies and career interests.

Table 5: Rule to determine the best number of clusters.

Index	Rule
GDI ₃₃	Maximum
Sil	Maximum
PBM	Maximum
PB	Maximum
WG	Maximum

Results and Discussion

Optimal k

The results of the internal cluster validity indices (ICVIs) derived from consensus clustering using a q-fold cross-validation strategy are presented in Figure 1. The indices – Generalized Dunn's IndDunn'sI33), Silhouette Index (Sil), Pakhira-Bandyopadhyay-Maulik Index (PBM), Point-Biserial Index (PB), and Wemmert-Gańczarski Index (WG) exhibited unique trends and variations, reflecting their differing sensitivity to cluster compactness, separation, and overall quality.

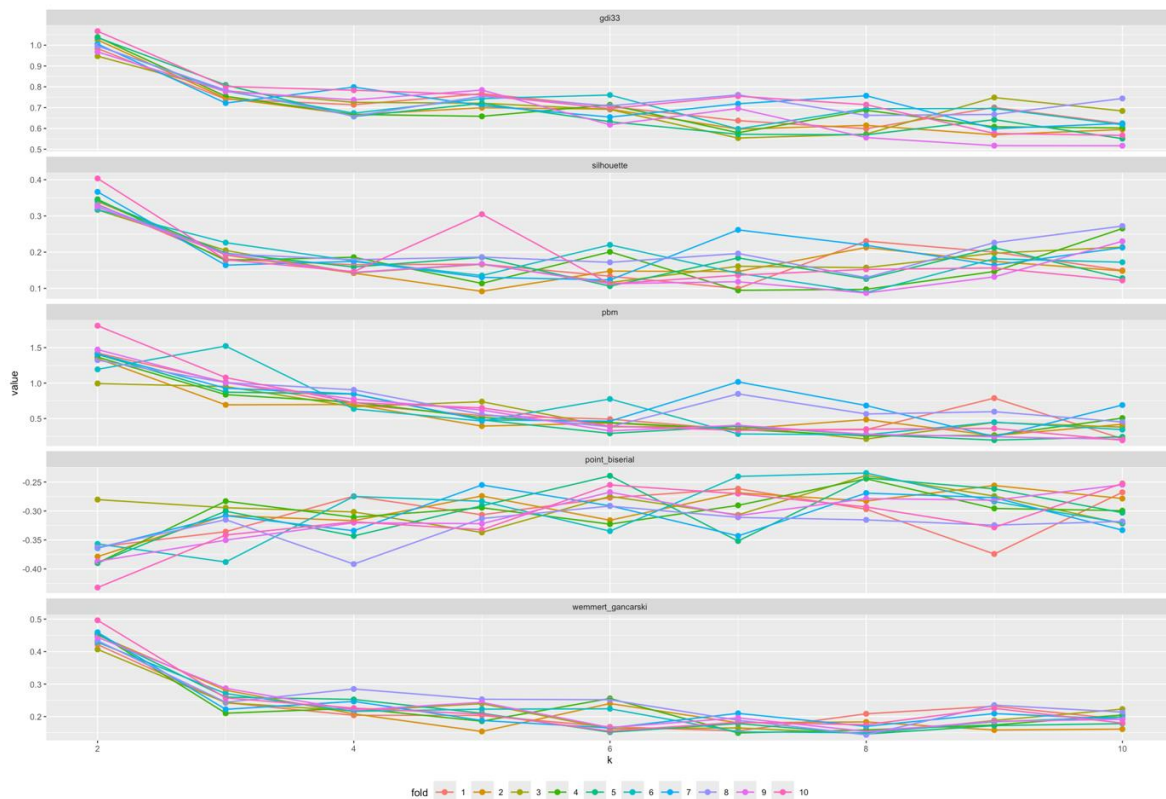


Figure 1: Trends of internal cluster validity indices (ICVIs) across different numbers of clusters (k) and folds using consensus clustering with q -fold cross-validation

GDI₃₃ consistently showed the highest values at $k = 2$ across folds, followed by a gradual decline as k increased. This trend indicates that smaller numbers of clusters yield more compact and well-separated groupings, with $k = 2$ consistently emerging as the most optimal cluster configuration.

The Sil Index also demonstrated a peak at $k = 2$ across all folds, reinforcing the finding that smaller cluster numbers provide the best balance between cluster compactness and separation. A declining trend in Sil values was observed as k increased, reflecting diminishing clustering quality.

The PBM index exhibited a decreasing trend with increasing k , but with some variability across folds. Intermediate values of k displayed slight fluctuations, particularly in some folds, indicating variability in cluster quality. The peak at $k = 2$ remained consistent, signifying its suitability as the optimal cluster number.

The PB index demonstrated higher variability compared to other indices, with both positive and negative values observed across folds. Despite these variations, peak values were often observed at lower k , with $k = 2$ frequently identified as optimal.

The WG index showed a consistently decreasing trend with higher k with $k = 2$ yielding the highest values across most folds. This trend suggests that the quality of clusters deteriorated as the number of clusters increased.

A consensus-based voting mechanism was applied to integrate the results of all ICVs across the folds in determining the optimal k . The results consistently identified $k = 2$ as the most appropriate cluster configuration, as it provided the highest indices for compactness and separation metrics across all indices and folds. This approach ensured robust and reproducible determination of the optimal cluster number, accounting for variability across folds.

Cluster Analysis of 21st Century Learning and Future Career Preferences

Figure 2 presents a heatmap depicting the clustering profiles, illustrating students' learning competencies (A1–A11) and future career preferences. The color gradient, ranging from blue (lower scores) to red (higher scores), reveals clear distinctions between the two identified clusters.

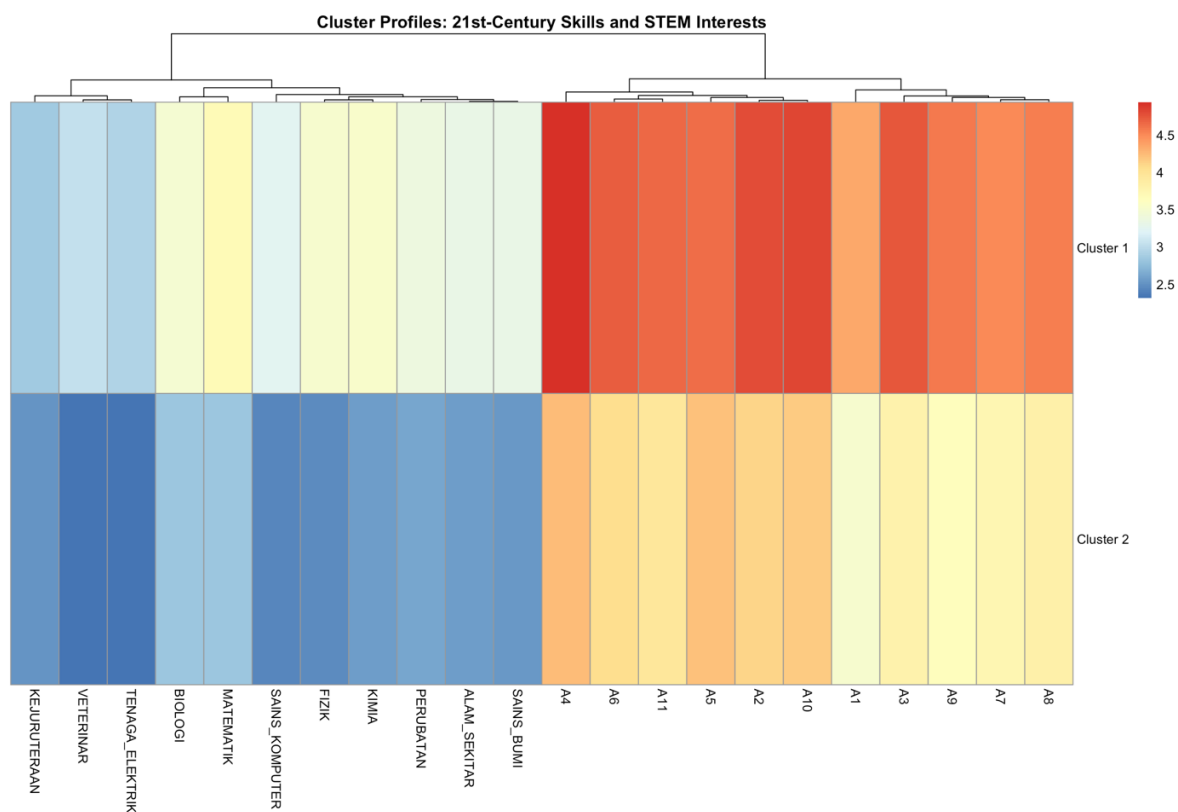


Figure 2: Heatmap of clustering profiles between 21st-century learning competencies and future career preferences

Cluster 1 displayed significantly higher confidence in 21st-century learning competencies, evident in the intense red shades across A1–A11. These students reported the highest confidence levels in leadership and collaboration (A1, A2, and A5), demonstrating exceptional self-assurance in leading, encouraging peers, and collaborating effectively. Additionally, they showed strong adaptability and an ability to incorporate diverse perspectives (A7 and A8), along with strong time management and task prioritization skills (A9 and A10), suggesting well-developed organizational abilities. This profile suggests a preference for fields requiring strong analytical and problem-solving skills, and indicates these students are well-equipped for future academic and professional pursuits demanding independent learning, teamwork, and leadership.

Cluster 2 exhibited moderate confidence in 21st-century learning competencies, as shown by the lighter shades for A1–A11 in the heatmap. Compared to Cluster 1, these students displayed lower confidence in leadership and collaboration (A1, A2, and A5), though their overall confidence remained relatively positive. While they expressed positive attitudes towards helping others and respecting diversity, Cluster 2 participants scored moderately lower than Cluster 1 in adaptability and time management (A7 and A9), indicating potential areas for development. These students may benefit from targeted interventions to enhance leadership, goal setting, and time management skills.

The heatmap also revealed distinct career preferences. Cluster 1 showed a strong preference for careers in Physics, Mathematics, and Chemistry, indicated by darker blue shades. In contrast, Cluster 2's interest were more diverse, encompassing a broader range of STEM fields, including Veterinary Science, Medicine, Environmental Sciences, and Earth Sciences.

Discussion

The analysis of ICVs through consensus clustering with q -fold cross-validation has provided insightful observations regarding the optimal number of clusters within the dataset. As shown in Figure 1, the GDI_{33} consistently reached its peak at $k = 2$ across all folds, indicating a preference for more compact and well-separated clusters when the dataset is partitioned into two groups. The steady decline in GDI_{33} values as k increased aligns with prior studies that emphasize the index's sensitivity to cluster compactness and separation, suggesting that increasing the number of clusters can lead to fragmentation and reduced cohesion within groups (Bezdek & Pal, 1998).

The Sil index also reinforced the findings of the GDI_{33} , with peak values consistently observed at $k = 2$ across all folds. This index, which evaluates the degree of separation between clusters, confirmed that a lower number of clusters provides the best balance between cohesion and separation (Rousseeuw, 1987). The diminishing Sil values observed as k increased reflect the common challenge in clustering, where overly granular partitions lead to overlapping and poorly defined clusters (Kaufman & Rousseeuw, 2009). This consistency across both GDI_{33} and Sil suggests a robust underlying structure favoring two clusters.

The PBM index exhibited a similar decreasing trend with increasing k , although with notable variability across different folds. While $k = 2$ consistently yielded the highest PBM scores, intermediate values displayed fluctuations, indicating fold-specific variability in cluster compactness and separation. Such behavior is characteristic of the PBM index, which balances intra-cluster compactness and inter-cluster separation but can be sensitive to outliers and noise in the data (Pakhira et al., 2004). This variability underscores the importance of using multiple ICVs to ensure a comprehensive understanding of clustering performance.

The PB index revealed the highest degree of variability among all indices, with both positive and negative values across folds. Despite this inconsistency, the PB Index often peaked at $k = 2$, supporting the consensus from other indices. The variability in PB scores may be attributed to its sensitivity to data distribution and the presence of overlapping clusters, a limitation previously noted in clustering literature (Milligan & Cooper, 1985). These results highlight the necessity of considering multiple validation indices to account for different data characteristics when determining the optimal cluster configuration.

The WG index consistently decreased as k increased, with $k = 2$ emerging as the most favourable configuration in terms of cluster quality. This index penalizes instances where data points are closer to clusters other than their own, thereby emphasizing the reliability of cluster assignments (Wemmert & Gançarski, 2002). The consensus-based voting mechanism applied to integrate the results of all ICVIs across folds further validated $k = 2$ as the optimal cluster configuration. This approach enhanced the robustness and reproducibility of the findings, aligning with best practices in cluster validation that advocate for multi-index evaluation to mitigate the biases of individual indices (Arbelaitz et al., 2013).

Building upon these clustering results, the study explored the relationship between 21st-century learning competencies and career interests among secondary school students, revealing two distinct student profiles. Cluster 1, characterized by high self-perceived competencies—particularly in leadership, collaboration, and goal setting—aligns with previous research highlighting the importance of these skills for academic and career success (Heckman & Kautz, 2013; Kuhn & Weinberger, 2005). Their strong organizational skills and ability to incorporate diverse perspectives suggest readiness for the demands of higher education and future careers. Moreover, their pronounced interest in core STEM fields such as physics, mathematics, and chemistry indicate a potential talent pipeline for these disciplines. Prior studies have shown that early interest and high self-efficacy in STEM subjects can significantly influence students' pursuit of STEM careers (Boaler et al., 2022; Nite et al., 2014), suggesting that these students are well-positioned for success in these fields.

In contrast, Cluster 2 exhibited moderate confidence in 21st-century learning competencies, with lighter shades across A1–A11 in the heatmap. While students in this cluster showed positive attitudes towards helping others and respecting diversity, they demonstrated lower confidence in leadership and time management compared to Cluster 1. These findings highlight potential areas for targeted interventions, particularly in fostering leadership and organizational skills, which are critical for academic success and professional readiness (Hensley et al., 2018). Interestingly, Cluster 2 students exhibited broader and more diverse career interests, ranging from veterinary science and medicine to environmental and earth sciences. This openness to multiple fields suggests a phase of exploration, which aligns with developmental theories on career choice, where students explore various options before narrowing their focus (Wang & Degol, 2017).

The contrasting profiles between the two clusters emphasize the complex interplay between self-perceived competencies and career interests. Students in Cluster 1, with higher self-efficacy in key skills, appear more focused in their career aspirations, particularly towards core STEM fields. This focused interest could be influenced by factors such as early exposure to STEM subjects, positive reinforcement from role models, and participation in STEM enrichment programs (Nite et al., 2014). In contrast, the broader interests in Cluster 2 may reflect external influences such as societal expectations, family input, and perceived career opportunities (Yang et al., 2017; Sheldrake & Mujtaba, 2020), highlighting the multifaceted nature of career decision-making among adolescents.

Tailored interventions that enhance leadership, goal setting, and time management skills could particularly benefit students in Cluster 2, helping them build the confidence needed to pursue focused career pathways. Simultaneously, educators should continue to nurture the strengths of students in Cluster 1, providing opportunities for deeper engagement in STEM subjects and

fostering advanced skills in critical thinking and problem-solving. Moreover, career guidance programs should be designed to accommodate the diverse needs of students, offering both broad exposure to various fields and targeted support for those with clear career aspirations.

Profiling students based on 21st-century learning competencies and career interests offers a thorough understanding of their readiness for future academic and professional challenges. Recognizing the distinct needs and strengths within student populations enables educators to develop targeted interventions that promote skill development, foster career exploration, and ultimately prepare students for success in an increasingly complex and dynamic world. Future research should explore the longitudinal impact of these profiles on students' academic trajectories and career outcomes, providing further insights into how educational practices can best support diverse learner needs.

Conclusion

This study applied consensus clustering with q-fold cross-validation to analyze internal cluster validity indices (ICVs), successfully identifying two distinct student profiles based on self-perceived 21st-century competencies and expressed career interests. The analysis revealed that the optimal number of clusters was consistently $k = 2$, as indicated by peak values across multiple indices, including GDI₃₃, Sil index, PBM, PB, and WG. Cluster 1 comprised students with high self-perceived competencies, particularly in leadership, collaboration, and goal setting, and demonstrated strong interest in core STEM fields such as physics, mathematics, and chemistry. In contrast, Cluster 2 consisted of students who, while displaying positive social attitudes, exhibited lower confidence in leadership and time management and showed broader, less focused career interests across various STEM domains. These findings underscore the diversity in students' preparedness for future academic and career pursuits, highlighting specific strengths and areas for intervention within each profile.

Profiling students' readiness for the future is essential in tailoring educational approaches that foster the development of 21st-century competencies and align with students' evolving career interests. Understanding these profiles allows educators and policymakers to design targeted interventions that support skill development, nurture career aspirations, and bridge existing gaps in students' preparedness for the demands of higher education and the workforce. Emphasizing competencies such as leadership, collaboration, adaptability, and time management, alongside nurturing clear career interests, equips students with the tools needed to thrive in an increasingly complex and dynamic world. Ultimately, recognizing and addressing the diverse needs of students ensures a more inclusive and effective educational experience that prepares all learners for success in the future.

Acknowledgement

The authors would like to extend their heartfelt gratitude to all who contributed to this research. We are particularly thankful to the staff at Jabatan Pendidikan Negeri Kedah for their enthusiastic cooperation and support. We also express our sincere appreciation to the dedicated teachers from SMK Bandar Sungai Petani, SMK Amanjaya, SMK Bedong, and SMK Gurun. Their active participation and commitment have been vital in facilitating our initiatives.

References

- Abdu-Raheem, B. O. (2015). Parents' socioeconomic status as a predictor of secondary school students' performance in Ekiti State, Nigeria. *Journal of Education and Practice*, 6(1), 123–128
- Angeli, C., & Valanides, N. (2020). Developing young children's computational thinking with educational robotics: An interaction effect between gender and scaffolding strategy. *Computers in Human Behavior*, 105, 105954. <https://doi.org/10.1016/j.chb.2019.03.018>
- Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M., & Perona, I. (2013). An extensive comparative study of cluster validity indices. *Pattern Recognition*, 46(1), 243–256. <https://doi.org/10.1016/j.patcog.2012.07.021>
- Barth, M., Godemann, J., Rieckmann, M., & Stoltenberg, U. (2007). Developing key competencies for sustainable development in higher education. *International Journal of Sustainability in Higher Education*, 8(4), 416–430. <https://doi.org/10.1108/14676370710823582>
- Bezdek, J. C., & Pal, N. R. (1998). Some new indexes of cluster validity. *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, 28(3), 301–315. <https://doi.org/10.1109/3477.678624>
- Blackmore, C., Vitali, J., Ainscough, L., Langfield, T., & Colthorpe, K. (2021). A review of self-regulated learning and self-efficacy: The key to tertiary transition in science, technology, engineering and mathematics (STEM). *International Journal of Higher Education*, 10(3), 169–182. <https://doi.org/10.5430/ijhe.v10n3p169>
- Boaler, J., Brown, K., LaMar, T., Leshin, M., & Selbach-Allen, M. (2022). Infusing mindset through mathematical problem solving and collaboration: Studying the impact of a short college intervention. *Education Sciences*, 12(10), 694. <https://doi.org/10.3390/educsci12100694>
- Brief, R. R., Ly, J., & Ion, B. A. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. https://www.bowdoin.edu/childrens-center/pdf/A_Framework_for_K12_science.pdf
- Cebrián, G., & Pubill, M. J. I. (2015). Competencies in education for sustainable development: Exploring the student teachers' views. *Sustainability*, 7(3), 2768–2786. <https://doi.org/10.3390/su7032768>
- Chan, R. C. H. (2022). A social cognitive perspective on gender disparities in self-efficacy, interest, and aspirations in science, technology, engineering, and mathematics (STEM): The influence of cultural and gender norms. *International Journal of STEM Education* 9, Article 37. <https://doi.org/10.1186/s40594-022-00352-0>
- Charlesworth, T. E. S., & Banaji, M. R. (2019). Gender in science, technology, engineering, and mathematics: Issues, causes, solutions. *The Journal of Neuroscience*, 39(37), 7228–7243. <https://doi.org/10.1523/JNEUROSCI.0475-18.2019>
- Chen, Y., So, W. W. M., Zhu, J., & Chiu, S. W. K. (2024). STEM learning opportunities and career aspirations: The interactive effect of students' self-concept and perceptions of STEM professionals. *International Journal of STEM Education*, 11, Article 1. <https://doi.org/10.1186/s40594-024-00466-7>
- Denaro, K., Sato, B., Harlow, A., Aebersold, A., & Verma, M. (2021). Comparison of cluster analysis methodologies for characterization of classroom observation protocol for undergraduate STEM (COPUS) data. *CBE—Life Sciences Education*, 20(1), Article ar3. <https://doi.org/10.1187/cbe.20-04-0077>

- Fang, J., He, L., Hwang, G., Zhu, X., Bian, C., & Fu, Q. (2022). A concept mapping-based self-regulated learning approach to promoting students' achievement and self-regulation in STEM activities. *Interactive Learning Environments*, 31, 7159–7181. <https://doi.org/10.1080/10494820.2022.2061013>
- Financial Industry Collective Outreach. (2023). *From classroom to career: Students' decision making—A survey on Malaysian students' transition from Form 5*. https://www.finco.my/wp-content/uploads/2023/07/From-Classroom-to-Careers_-Students-Transition-from-Form-5_FINCOs-Report_2023.pdf
- Friday Institute for Educational Innovation. (2012). *Middle and high school STEM-students survey*. Raleigh, NC: Author.
- González-Pérez, L. I., & Ramírez-Montoya, M. S. (2022). Components of Education 4.0 in 21st century skills frameworks: Systematic review. *Sustainability*, 14(3), 1493. <https://doi.org/10.3390/su14031493>
- Guaman-Quintanilla, S., Everaert, P., Chiluita, K., & Valcke, M. (2022). Impact of design thinking in higher education: A multi-actor perspective on problem solving and creativity. *International Journal of Technology and Design Education*, 33(1), 217. <https://doi.org/10.1007/s10798-021-09724-z>
- Hadiyati, M. A., & Astuti, B. (2023). Student careers: What factors influence career choice? *Journal of Education Research and Evaluation*, 7(4). <https://doi.org/10.23887/jere.v7i4.61686>
- Haug, B. S., & Mork, S. M. (2021). Taking 21st century skills from vision to classroom: What teachers highlight as supportive professional development in the light of new demands from educational reforms. *Teaching and Teacher Education*, 100, 103286. <https://doi.org/10.1016/j.tate.2021.103286>
- Heckman, J. J., & Kautz, T. (2013). *Fostering and measuring skills: Interventions that improve character and cognition*. https://www.nber.org/system/files/working_papers/w19656/w19656.pdf
- Hensley, L. C., Wolters, C. A., Won, S., & Brady, A. C. (2018). Academic probation, time management, and time use in a college success course. *Journal of College Reading and Learning*, 48(2), 105–123. <https://doi.org/10.1080/10790195.2017.1411214>
- Hirudayaraj, M., Baker, R. M., Baker, F., & Eastman, M. (2021). Soft skills for entry-level engineers: What employers want. *Education Sciences*, 11(10), 641. <https://doi.org/10.3390/educsci11100641>
- Hsu, A., Chen, M., & Shin, N. (2021). From academic achievement to career development: Does self-regulated learning matter? *International Journal for Educational and Vocational Guidance*, 22, 285–305. <https://doi.org/10.1007/s10775-021-09486-z>
- Johnson, L. S. (2000). The relevance of school to career: A study in student awareness. *Journal of Career Development*, 26(4), 263–276
- Karim, N. B. A., Abdul Rahman, N. I. B., & Kaur, H. (2024). The impact of outcome expectation towards career choices among secondary school students in Petaling District, Malaysia. *International Journal of Academic Research in Business and Social Sciences*, 14(5), 1489–1507.
- Karimi, H. S., & Piña, A. A. (2021). Strategically addressing the soft skills gap among STEM undergraduates. *Journal of Research in STEM Education*, 7(1), 21. <https://doi.org/10.51355/jstem.2021.99>
- Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: An introduction to cluster analysis*. John Wiley & Sons.

- Koyunlu Ünlü, Z., & Dökme, İ. (2020). Multivariate assessment of middle school students' interest in STEM careers: A profile from Turkey. *Research in Science Education*, 50(3), 1217–1231.
- Kuhn, P., & Weinberger, C. J. (2005). Leadership skills and wages. *Journal of Labor Economics*, 23(3), 395. <https://doi.org/10.1086/430282>
- Liu, Y., Li, Z., Xiong, H., Gao, X., & Wu, J. (2010). Understanding of internal clustering validation measures. In *Proceedings of the 2010 IEEE International Conference on Data Mining* (pp. 911–916). IEEE Computer Society. <https://doi.org/10.1109/ICDM.2010.35>
- López, P., Simó, P., & Marco, J. (2023). Understanding STEM career choices: A systematic mapping. *Heliyon*, 9(6), e16676. <https://doi.org/10.1016/j.heliyon.2023.e16676>
- Magpantay, I. C. D., & Pasia, A. E. (2022). Problem-based learning materials in upskilling mathematics critical thinking skills. *International Journal of Science Technology Engineering and Mathematics*, 2(4), 74. <https://doi.org/10.53378/352940>
- Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50(2), 159–179. <https://doi.org/10.1007/BF02294245>
- Mudhar, Murwani, F. D., Hitipeuw, I., & Rahmawati, H. (2020). Career interest data trends in the era of information technology of high school students at Surabaya, Indonesia. *Data in Brief*, 30, 105480. <https://doi.org/10.1016/j.dib.2020.105480>
- Nite, S. B., Margaret, M., Capraro, R. M., Morgan, J., & Peterson, C. A. (2014). Science, technology, engineering, and mathematics (STEM) education: A longitudinal examination of secondary school intervention. In *2014 IEEE Frontiers in Education Conference (FIE) Proceedings* (pp. 1–7). IEEE. <https://doi.org/10.1109/FIE.2014.7044214>
- Nuangchalerm, P., Prachagool, V., Islami, R. A. Z. E., & Abdurrahman, A. (2020). Contribution of integrated learning through STEM education in ASEAN countries. *Jurnal Pendidikan Progresif*, 10(1), 11. <https://doi.org/10.23960/jpp.v10.i1.202002>
- Obaid, T., Aghajani, H., & Linn, M. C. (2023). Using optimized clustering to identify students' science learning paths to knowledge integration. *STEM Education Review*, 1. <https://doi.org/10.54844/stemer.2023.0354>
- Ojukwu, M. O., & Ali, E. N. (2020). The demographic variables of parents as predictors of career preference of in-school adolescent students in Abia State. *GPH-International Journal of Educational Research*, 3(08), 20–35.
- Pakhira, M. K., Bandyopadhyay, S., & Maulik, U. (2004). Validity index for crisp and fuzzy clusters. *Pattern Recognition*, 37(3), 487–501. <https://doi.org/10.1016/j.patcog.2003.06.005>
- Panskyi, T., Korzeniewska, E., & Firych-Nowacka, A. (2024). Educational data clustering in secondary school sensor-based engineering courses using active learning approaches. *Applied Sciences*, 14(12), 5071. <https://doi.org/10.3390/app14125071>
- Rafiq, H. M. W., Fatima, T., Sohail, M. M., Saleem, M., & Khan, M. A. (2013). Parental involvement and academic achievement: A study on secondary school students of Lahore, Pakistan. *International Journal of Humanities and Social Science*, 3(8), 209–223.
- Ribeirinha, T., Correia, M., & Baptista, M. (2023, November). STEM career aspirations among Portuguese secondary school students. In *2023 International Symposium on Computers in Education (SIIE)* (pp. 1–5). IEEE.

- Roehrig, G. H., Dare, E. A., Ellis, J. A., & Ring-Whalen, E. (2021). Beyond the basics: A detailed conceptual framework of integrated STEM. *Disciplinary and Interdisciplinary Science Education Research*, 3(1), 1. <https://doi.org/10.1186/s43031-021-00041-y>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Saitta, S., Raphael, B., & Smith, I. (2008). A comprehensive validity index for clustering. *Intelligent Data Analysis*, 12(6), 529–548. <http://iospress.metapress.com/index/h657285100831154.pdf>
- Saitta, S., Raphael, B., & Smith, I. F. C. (2007). A bounded index for cluster validity. In *Machine Learning and Data Mining in Pattern Recognition* (pp. 174–187). https://doi.org/10.1007/978-3-540-73499-4_14
- Sheldrake, R., & Mujtaba, T. (2020). Children's aspirations towards science-related careers. *Canadian Journal of Science, Mathematics and Technology Education*, 20, 7–26. <https://doi.org/10.1007/s42330-019-00070-w>
- Songer, N. B., Newstadt, M. R., Lucchesi, K., & Ram, P. (2019). Navigated learning: An approach for differentiated classroom instruction built on learning science and data science foundations. *Human Behavior and Emerging Technologies*, 2(1), 93. <https://doi.org/10.1002/hbe2.169>
- Voogt, J., & Roblin, N. P. (2012). A comparative analysis of international frameworks for 21st century competences: Implications for national curriculum policies. *Journal of Curriculum Studies*, 44(3), 299–321. <https://doi.org/10.1080/00220272.2012.668938>
- Volman, M., Karssen, M., Emmelot, Y., & Heemskerk, I. (2020). The focus of schools on twenty-first-century competencies and students' views of these competencies. *The Curriculum Journal*, 31(4), 648–665
- Wang, M. T., & Degol, J. L. (2017). Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. *Educational Psychology Review*, 29, 119–140. <https://doi.org/10.1007/s10648-015-9355-x>
- Wemmert, C., & Gañarski, P. (2002). A multi-viewpoint approach to cluster validity. In *Proceedings of the International Conference on Pattern Recognition* (pp. 880–883). <https://doi.org/10.1109/ICPR.2002.1048490>
- Yang, G., Badri, M., Al-Mazroui, K., Al-Rashedi, A., & Nai, P. (2017). Science as interests but not for career: Understanding high school students' engagement in science in Abu Dhabi. *EURASIA Journal of Mathematics Science and Technology Education*, 13(7), 3621–3639. <https://doi.org/10.12973/eurasia.2017.00749a>
- Zastudil, C., Rogalska, M., Kapp, C., Vaughn, J., & MacNeil, S. (2023). Generative AI in computing education: Perspectives of students and instructors. *arXiv*. <https://doi.org/10.48550/arXiv.2308>