Development and Validation of a Survey Instrument on Knowledge, Attitude, and Practices (KAP) Regarding the Educational Use of ChatGPT among Preservice Teachers in the Philippines

Dave Arthur R. Robledo*, Celso G. Zara, Sherryl M. Montalbo, Norrie E. Gayeta, Abegail L. Gonzales, Mary Grace A. Escarez, and Erma D. Maalihan

***Abstract*—ChatGPT has gained popularity among Philippine educational institutions due to its versatility and usefulness in academic activities. However, educators have raised concerns about the ethical issues, academic dishonesty, and technology dependence associated with its use. This study aimed to develop and validate an instrument called KAP-CQ39 that assesses the knowledge, attitudes, and practices of preservice teachers regarding the use of ChatGPT. The study involved 4 experts, 12 independent evaluators, and 206 preservice teachers. The instrument underwent a series of assessments to establish quality, validity, and reliability, resulting in the final version of KAP-CQ39. Results showed that KAP-CQ39 has acceptable content validity (CVR and CVI values>0.78), face validity (IS>1.5), construct validity (chi-squared value of 897.564, *df*=577, p<0.001), and internal consistency (overall Cronbach coefficient=0.91). This proved that KAP-CQ39 can be a valuable tool for assessing the impact of ChatGPT on the Philippine educational system and can provide insights for educational policy makers regarding the use of artificial intelligence in education.**

***Index Terms*—Artificial intelligence-based education, attitude, ChatGPT, knowledge, practices**

I. Introduction

Artificial intelligence (AI) is the process when machines, particularly computer systems, simulate human intelligence processes. Expert systems, natural language processing, speech recognition, and machine vision are a few examples of specific AI applications [1]. Because of its adaptability and usefulness, AI has the potential to transform education by improving student learning, instructor productivity, and institutional efficiency [2, 3]. In the past years, several AI tools and software were developed to revolutionize the educational system and improve teaching and learning experiences. The applications of AI in education are wide-ranging, which include the personalization of learning platforms that adjust to specific student needs and the development of smart tutoring systems that provide real-time feedback and support [4]. In addition, AI can also help with administrative activities like scheduling and grading, freeing teachers' time to focus on instruction and student engagement [5]. AI tools can analyze large amounts of data to find patterns and trends in student performance. This can assist teachers in identifying areas where learners require additional assistance and improve teaching tactics [6]. Since AI tools can analyze and understand natural language, these tools can be used as virtual assistants that can answer students' questions and provide academic support [7].

ChatGPT, short for "*Chat Generative Pretrained Transformer*" (https://chat.openai.com/chat), is an artificial intelligence Chatbot developed by OpenAI and launched in November 2022 [8]. ChatGPT is versatile even though its primary job is to resemble human conversationalists, it can create and debug computer programs, produce literary works like poems and song lyrics, compose music, play scripts, fictional stories, essays, and answer test questions [9, 10]. As shown in Fig. 1, ChatGPT has an array of potential applications in education and is expected to advance continuously in the next few years [10]. It has made information and support available to students and educators in an easily accessible and convenient manner. ChatGPT can provide tailored feedback and help for students. It can also answer questions on diverse topics, from introductory algebra to advanced scientific concepts [10, 11]. As a result, ChatGPT immediately became well-known for its thorough responses and articulate comments in a variety of subject areas.

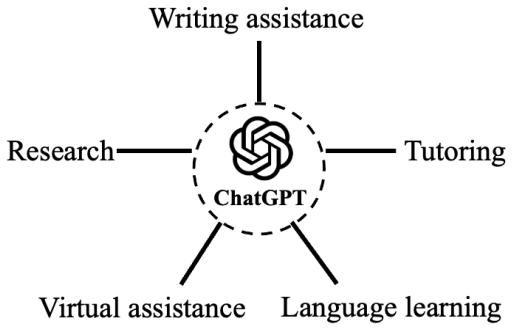

Fig. 1. ChatGPT applications in education.

Although ChatGPT can be a powerful learning tool, it can also present several risks and challenges. Educational researchers emphasized the growing concern about the negative impact of ChatGPT in the educational system, such

Manuscript received March 22, 2023; revised April 23, 2023; accepted June 15, 2023.

Dave Arthur R. Robledo is with the Graduate School of Science and Engineering and the Center for Marine and Environmental Studies of Ehime University, Matsuyama City, Japan.

Celso G. Zara, Sherryl M. Montalbo, Norrie E. Gayeta, Abegail L. Gonzales, Mary Grace A. Escarez, and Erma D. Maalihan are with the College of Teacher Education of the Batangas State University, The National Engineering University (BatStateU-TNEU), Batangas City, Philippines.

*Correspondence: j865001u@mails.cc.ehime-u.ac.jp (D.A.R).

as ethical concerns, promotion of academic dishonesty, depreciation of students' creativity, and dependence on technology [11, 12]. Tlili and Shehata *et al.* [13] revealed that various issues, such as cheating and privacy misleading, might arise when ChatGPT is used in educational settings. Furthermore, Kasneci and Seßler *et al.* [14] emphasized that integrating and fully utilizing large language models like ChatGPT in learning settings and teaching curricula require a clear strategy within educational systems and a clear pedagogical approach with a strong focus on critical thinking and fact-checking strategy. Thus, school policies and regulations play a crucial role in ensuring that the use of ChatGPT is safe, secure, and effective for academic purposes.

For these reasons, schools should conceptualize strategic options to mitigate risks. This is where the Knowledge, Attitude, and Practices (KAP) survey as a tool is essential to determine how educational stakeholders perceive and engage with a specific issue or topic for improved policymaking [15]. In the knowledge section of a KAP survey, participants may be required to provide information or to identify key terminologies or concepts related to the problem or topic. The attitude component asks respondents to indicate how much they agree or disagree with a list of items related to their feelings or attitude toward the topic. In addition, questions concerning respondents' present actions or practices about the issue or topic and their opinions of these behaviors are asked as part of the practice component [16]. In the context of ChatGPT, a KAP survey might offer essential data regarding how students view the tool and how it could be incorporated into the learning process, which is necessary for policymaking.

Innovative options like this is essential especially in the context of the COVID-19 pandemic, which has accelerated the need for technology integration in education [17, 18]. While the Philippines has progressed in integrating technology in some areas, teachers and students still need more training and support to effectively use technology in their classrooms [19, 20]. By presenting innovations for teacher training, professional development, and classroom assistance, AI has the potential to play a key role in teacher education [21–23]. Preservice teachers are college students enrolled in a teacher education program but have not yet completed the criteria to obtain a teaching license [24, 25]. This status makes them suitable respondents for the KAP survey on ChatGPT since they are on track to becoming teachers. This means that they will be responsible for using and integrating technology like ChatGPT in their future classrooms. In addition, understanding pre-service teachers' KAP towards ChatGPT can help identify potential barriers and challenges that may hinder its effective implementation in the school, allowing for developing strategies to address these issues [26].

To the researchers' knowledge, no reliable and validated KAP questionnaire is currently available regarding the use of ChatGPT in educational settings. Hence, this study was conducted to develop, validate, and evaluate a KAP questionnaire regarding face validity, content validity, construct validity, and reliability in using ChatGPT among pre-service teachers.

II. Methodology

A. Study Design and Participants

The study employed a cross-sectional survey design, which involves gathering data from a single group of participants at a single point in time [27]. A purposive sampling technique was used to recruit participants, where researchers selected individuals who meet specific criteria related to the objective of the study. In this case, the researchers recruited 206 preservice teachers enrolled in a public university's teacher education program in Batangas City, Philippines. For ethical considerations, the researchers obtained permission from the university administration and distributed online informed consent and data privacy forms to the respondents. After the authorization, data collection was carried out using Google Forms in February 2023.

B. KAP-C Questionnaire Development

Following the protocol of Koo and Poh *et al.* [28] the KAP questionnaire regarding the use of ChatGPT (KAP-CQ39) was developed based on five phases: item generation, content validation, face validation, construct validation, and reliability assessment.

1) Phase I: Item generation and questionnaire design

The items comprising the KAP-CQ39 instrument were derived from a comprehensive literature review and a semi-structured interview with four experts in educational technology, information technology, computer science, and curriculum design. The software *Harzing: Publish or Perish* was used for the literature search (https://harzing.com/resources/publish-or-perish). This open-source software program allows academics and researchers to conduct citation analysis on academic publications from Google Scholar, Web of Science, Microsoft Academic, and Scopus database [29]. Keywords such as "ChatGPT", "Knowledge", "Attitudes", "Practices", "Artificial Intelligence", "Educational Technology", "Behavior", and "KAP" were used to narrow down the literature search. The identified documents included peer-reviewed research articles, conference proceedings, and short communications. Additional sources were retrieved from websites and forums about artificial intelligence and educational technology.

A virtual, semi-structured interview was conducted with the four experts who agreed to participate in the study to gather more information about the educational use of ChatGPT. Open-ended questions were used to probe the interview responses, such as "*What are some key concepts related to the use of ChatGPT in education that you believe should be included in the KAP survey?", "What attitudes or practices do you think students may have towards the use of ChatGPT in education?"* and *"Based on your experience and knowledge, what factors should be considered when developing survey items to assess knowledge, attitudes, and practices related to the use of ChatGPT in education*?". All responses were coded twice for domain identification and thematic analysis using the protocol of Braun and Clark [30]. These thematic patterns were used to generate survey items that capture the relevant themes or concepts.

For the questionnaire design, the knowledge domain of

KAP-CQ39 consists of true or false items about the features of ChatGPT. According to Krumpal [31], true or false questions are objective, which means that there is no ambiguity in the response options. This can help reduce bias in responses and increase the reliability of the survey. The researchers delved on the users' belief, perceptions, and experiences on ChatGPT to generate Likert scale items that assessed the respondents' attitudes towards ChatGPT. For the practice domain, this study used the activities, exercises, and actions related to the applications of ChatGPT in educational settings. Additional questions were generated to assess the advantages, and disadvantages of the educational use of ChatGPT and their perceptions of the banning or regulating of ChatGPT in schools and academic institutions.

It is important to note that KAP-CQ39 items focused primarily on the applications of ChatGPT in educational settings. Other features and functions of ChatGPT that are not related to education, such as marketing content creation, customer service assistance, banking and financial services, and entertainment purposes, were excluded from the study.

The initial set of KAP-CQ39 comprised 55 items generated from the literature review and the experts' responses; and classified into three domains—knowledge, attitude, and practices on using ChatGPT in educational settings. Items on the knowledge domain (Items 1 to 20) were presented in a true or false question format. An additional "I don't know" option was added to determine if the respondents had no prior knowledge about the presented items. Items on attitude (Items 21 to 35) were measured using a four-point Likert scale with the responses of "1 = Strongly Disagree, 2 = Disagree, 3 = Agree, 4 = Strongly Agree". A neutral or midpoint option was not included because it dilutes the responses' overall strength, resulting in ambiguity and less reliable results. Items on practices (Items 36 to 55) were measured using a three-point scale – "1 = Yes, 2 = No, and 3 = I don't know". In addition, an item was added to assess the awareness and frequency of utilization of ChatGPT, and another item to determine the gender of the preservice teachers as their sociodemographic profile.

2) Phase II: Content validity

To assess the content validity of KAP-CQ39, the researchers invited six panelists from different relevant fields to form a multidisciplinary panel of evaluators. Initially, four experts (two educational technology experts, one curriculum designer, and one computer scientist) reviewed and evaluated the initial set of items of KAP-CQ39 to assess its content validity (Fig. 2). Using the Content Validity Ratio (CVR) calculation, all panel members evaluated each item based on relevance, clarity, simplicity, and necessity. CVR is a numerical measure that indicates the instrument's validity based on expert evaluations of its content validity [32]. The formula for CVR is presented below.

$$\text{CVR} = \frac{Ne - \frac{N}{2}}{\frac{N}{2}} \quad (1)$$

where *Ne* is the number of panelists who classify an item as "essential", and *N* refers to the total number of experts involved in the evaluation. Typically, a CVR value of 0.78 or higher is considered necessary to establish the validity of an item or scale [33]. Moreover, each item was evaluated according to relevance, clarity, simplicity, and necessity using Content Validity Index (CVI). CVI is computed as the number of experts giving a "relevant to highly relevant," "clear to very clear", "simple to very simple", and "necessary to highly necessary" ratings of 3 or 4 for each item, divided by the total number of experts [26]. A CVI>0.79 indicates the relevance, clarity, simplicity, and necessity of the item. While 0.70 to 0.79 means that items should be revised or improved, CVI<0.70 shows that items should be eliminated based on experts' judgment [27].

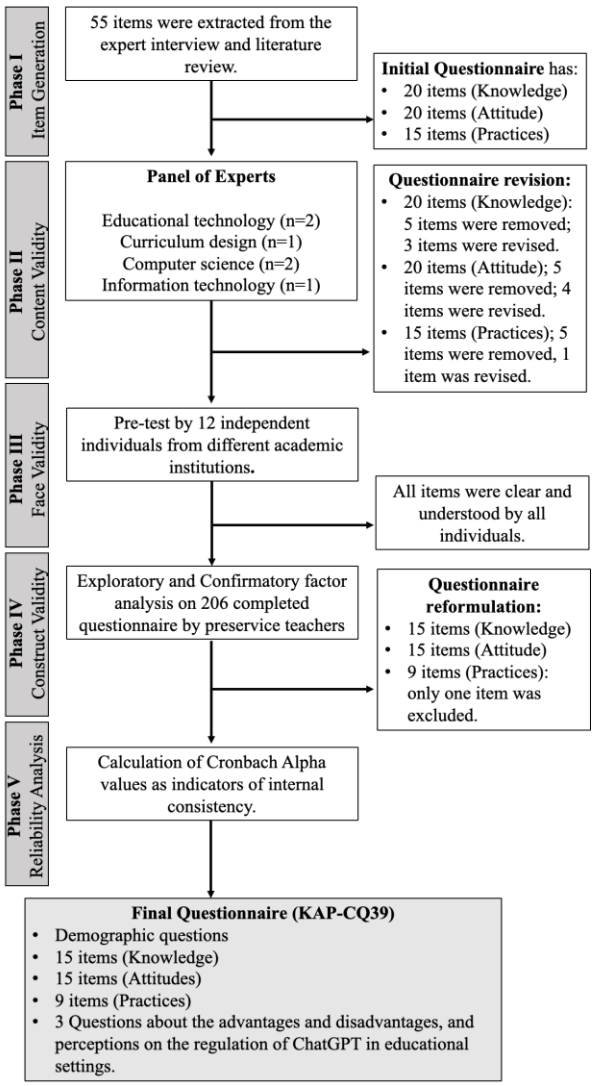

Fig. 2. Flow chart of the KAP-CQ39 development process.

3) Phase III: Face validity

Quantitative and qualitative methods were used to evaluate the face validity of KAP-CQ39. Impact Score (IS) was used to assess the quantitative face validity, while a three-point checklist about the items' organization and clarity ("good, moderate, or bad) was used to evaluate the qualitative face validity [34]. The impact score for each item in the KAP-CQ39 was calculated using a five-point Likert scale ranging from "1 = not essential" to "5 = most essential". This was done by multiplying the *frequency* by the *importance* score, where the frequency was the number of individuals

who rated the item as "3" or "4", and importance was the mean score of the item on the 1–5 rating scale [35]. Any item with an impact score of 1.5 or higher was included in the KAP-CQ. Twelve independent individuals from different institutions were requested to evaluate each item's face validity.

4) Phase IV: Construct validity

The multidimensional structure of the domains of KAP-CQ39 was evaluated through Exploratory Factor Analysis (EFA) using the principal components method and varimax rotation. To determine whether the data were suitable for factor analysis, the Kaiser-Meyer-Olkin (KMO) measure was used. The KMO test calculates the degree of correlation between the variables in the dataset, with higher values indicating better suitability for factor analysis [36]. A value of 0 to 1 is produced, with values of 0.6 or higher considered acceptable and 0.8 or higher considered good. Sampling adequacy was confirmed by conducting Bartlett's Test of Sphericity, which assesses whether the correlation matrix of the variables in the dataset is significantly different from an identity matrix [37]. If the correlation matrix is not significantly different, it would suggest that the variables are unrelated and thus unsuitable for factor analysis.

Factor loading was performed to assess the strength of the relationship between a given variable and a factor in factor analysis. Factor loadings ranged from –1 to 1, with larger values indicating stronger associations. Items with>0.3 loadings were considered "essential" and therefore retained, while items with loadings below 0.3 were removed [38]. A Confirmatory Factor Analysis (CFA) was also conducted to evaluate whether the extracted structure model from the Exploratory Factor Analysis (EFA) fit the data well. The acceptable criteria for model fit indices were a chi-square value of less than three and a Root Mean Square Error of Approximation (RMSEA) of less than 0.08. These indices were used as cut-off points to determine the suitability of the model in the CFA analysis [34, 39, 40].

5) Phase V: Assessment of reliability

The study utilized Cronbach's alpha coefficient to determine the Internal Consistency (IC) of the KAP-CQ39 items. A total Cronbach's alpha value exceeding 0.7 was used as the criterion for good internal consistency among the items [41]. All statistical analysis was performed using IBM SPSS version 16.5 and Microsoft Excel Version 2016.

III. Results

A total of 206 preservice teachers (87.4% female) aged 19 to 22 completed the KAP-CQ39, as shown in Table I. Results showed that 78.6% of the preservice teachers were aware of or have used ChatGPT. Since there was no significant difference between the male and female groups regarding their awareness level, this indicates that the sample of respondents was representative of the target population and had greater confidence in the generalizability and reliability of the survey results (p=0.087). Since the respondents were on the same year level, age was not considered a variable of interest.

TABLE I: Sociodemographic of Preservice Teachers (n = 206)

Variables		Frequency (%)	p-value*
Gender	Male	26 (12.6)	0.087
	Female	180 (87.4)	
Awareness and Use of ChatGPT	Yes	162 (78.6)	
	No	44 (21.4)	

* Fisher exact test was used to determine the association between variables.

A. Item Generation, Content Validity, and Face Validity

The present study conducted a thorough literature review and expert interviews to identify a total of 55 items for the KAP-CQ39. Following a content validation process, 15 items were eliminated from the pool due to concerns of confusion, repetitiveness, or irrelevance, as suggested by the expert panel (Fig. 1). The remaining 40 items underwent a process of revision to enhance their clarity and ensure their relevance to each of the three domains (knowledge = 15, attitude = 15, and practices = 10). The content validity ratio (CVR) and content validity index (CVI) values for the revised items were higher than 0.78, signifying a high level of content validity for the questionnaire. The face validation process revealed that all the retained items were clear and easy to understand, and the impact scores of all items were above 1.5, indicating their significance in the KAP-CQ39 (Table II).

TABLE II: Results of Content Validity and Face Validity

			CVI			Impact
Item	CVR	Simplicity	Relevance	Clarity	Total	Score
1	0.93	1	1	1	1	3.2
2	1	0.87	0.97	0.99	0.94	3.5
3	0.88	0.98	0.98	1	0.99	2.9
4	1	1	1	1	1	1.9
5	0.98	0.93	0.93	0.93	0.93	2.4
6	0.82	0.88	0.98	0.87	0.91	3.4
7	0.81	0.81	0.99	1	0.93	3.1
8	1	1	1	1	1	3.5
9	1	1	1	1	1	2.8
10	0.93	1	0.87	0.97	0.95	1.9
11	0.88	0.88	0.98	0.98	0.95	2.0
12	0.94	1	1	1	1	3.1
13	1	0.98	0.93	0.93	0.95	3.2
14	0.98	1	0.87	0.97	0.95	3.3
15	0.82	0.88	0.98	0.98	0.95	3.2
16	0.81	1	1	1	1	2.6
17	1	0.93	1	0.87	0.93	2.5
18	1	0.88	0.88	0.98	0.91	2.5
19	1	0.94	1	1	0.98	2.6
20	1	1	0.98	0.93	0.97	3.6
21	0.99	0.98	1	0.87	0.95	3.2
22	0.86	0.82	0.88	0.98	0.89	3.1
23	0.79	0.81	1	1	0.94	3.2
24	0.89	0.93	1	0.87	0.93	3.6
25	0.88	0.88	0.88	0.98	0.91	3.8
26	1	0.94	1	1	0.98	3.5
27	1	1	0.98	0.93	0.97	2.7
28	1	1	0.87	0.97	0.95	2.8
29	0.98	0.88	0.98	0.98	0.95	2.7
30	0.84	1	1	1	1	2.6
31	0.88	0.98	0.93	0.93	0.95	1.9
32	1	0.98	0.93	0.93	0.95	2.8
33	0.89	0.82	0.88	0.98	0.89	2.6
34	0.78	0.79	0.80	0.79	0.81	1.5
35	1	1	1	1	1	2.3
36	0.79	1	1	1	1	2.5
37	0.81	0.93	1	0.87	0.93	1.8
38	0.98	0.88	0.88	0.98	0.91	2.5
39	0.89	0.94	1	1	0.98	2.5
40	1	1	0.98	0.93	0.97	2.7

TABLE III: RESULTS OF FACTOR LOADING OF KAP-CQ39

Items	Knowledge	Attitude	Practice
1. ChatGPT uses artificial intelligence to generate human-like responses.	**0.433**	0.043	–0.122
2. ChatGPT can only provide answers in English.	**0.355**	0.122	0.110
3. ChatGPT responses are 100% accurate.	**0.433**	–0.44	0.099
4. ChatGPT is designed to provide human-like conversations.	**0.655**	0.211	0.022
5. ChatGPT is trained on a diverse range of topics.	**0.677**	0.122	0.122
6. ChatGPT is a commercial product and is not for free.	**0.504**	0.003	0.188
7. ChatGPT can only provide text-based responses	**0.400**	0.102	–0.221
8. ChatGPT responses are generated by a pre-programmed algorithm	**0.322**	0.203	0.199
9. ChatGPT can check and grade student assignments.	**0.310**	0.122	0.187
10. ChatGPT can help teachers with lesson planning.	**0.446**	0.089	–0.122
11. ChatGPT can be used to assist students with their coursework	**0.533**	0.044	0.221
12. ChatGPT can be integrated with virtual learning environments	**0.587**	–0.133	0.199
13. ChatGPT can create essays and articles about a specific topic.	**0.322**	0.021	0.187
14. ChatGPT can provide additional teaching resources and learning materials for students.	**0.443**	0.199	–0.221
15. ChatGPT can provide instant feedback on pronunciation.	**0.377**	0.122	0.166
16. I find ChatGPT helpful in answering questions.	–0.221	**0.388**	–0.122
17. I trust the responses provided by ChatGPT.	0.199	**0.474**	0.193
18. I find ChatGPT responses to be accurate.	0.187	**0.433**	0.099
19. I find ChatGPT to be a useful tool for learning.	–0.122	**0.544**	0.021
20. I believe ChatGPT has the potential to revolutionize the way we access information.	0.221	**0.655**	0.189
21. I am concerned about the ethical implications of using ChatGPT.	0.199	**0.388**	–0.177
22. I believe that using ChatGPT to complete academic assignments is unethical.	0.003	**0.399**	0.088
23. I believe that students and teachers should be allowed to use ChatGPT in the classroom.	0.102	**0.410**	0.032
24. I believe that using ChatGPT for academic purposes should be discouraged.	0.203	**0.405**	0.011
25. I believe that ChatGPT should be banned in all schools and academic institutions.	0.122	**0.499**	0.002
26. I believe that the use of ChatGPT for academic purposes undermines the learning process	0.089	**0.556**	0.033
27. I think that the use of ChatGPT for academic purposes should be monitored and regulated	0.044	**0.488**	0.102
28. I think people who use ChatGPT for academic purposes are cheating	0.003	**0.522**	0.032
29. I think that schools and educators should educate students on the dangers of relying on ChatGPT for academic purposes.	0.102	**0.467**	0.022
30. I think ChatGPT is useful when used correctly and monitored accordingly.	0.203	**0.544**	00
31. I can use ChatGPT for remote or online education.	0.099	–0.221	**0.402**
32. I usually review and revise the outputs of ChatGPT before using or submitting them.	0.021	0.199	**0.466**
33. I use third-party paraphrasing tools (e.g., Quill Bot) to avoid plagiarism detection.	0.189	0.187	**0.576**
34. I use ChatGPT to learn or teach a foreign language.	–0.177	–0.122	0.122
35. I use ChatGPT to get an initial idea about specific topics.	0.088	0.221	**0.553**
36. I use ChatGPT more often than Google Search.	0.032	0.199	**0.399**
37. I use ChatGPT to confirm my ideas or arguments.	0.011	0.187	**0.433**
38. I use ChatGPT to explain complicated concepts or topics.	0.002	–0.221	**0.555**
39. I use ChatGPT for educational purposes only.	0.033	0.166	**0.343**
40. I use ChatGPT to make my work easier and faster.	0.102	–0.122	**0.598**

B. Construct Validity

The study employed the Kaiser-Mayer-Olkin test and Bartlett's sphericity test to assess the dataset's suitability for factor analysis. The outcomes indicated that all items satisfied the requirements for factor analysis, with a KMO value of 0.885 and a Chi-squared value of 897.564 (df=577, $p<0.001$). The screen plot identified three constructs with eigenvalues greater than 1, which accounted for 74.6% of the total variance. Finally, the researchers opted for a three-factor solution using varimax rotation for the final analysis.

Table III displays all items of the KAP-CQ39 and their corresponding factor loadings in the knowledge, attitude, and practices domains. The findings revealed that only 39 out of 40 items were retained in the final questionnaire, as Item No. 34 had a factor loading of less than 0.3. Additionally, the total percentage of variance for the KAP-CQ39 was 38.7%, with the knowledge, attitude, and practices domains accounting for 15.4%, 10.6%, and 12.7%, respectively.

The hypothetical model of the Confirmatory Factor Analysis (CFA) based on the 39 remaining items in the three domains is presented in Fig. 3. The model exhibited acceptable fit indices with a Root Mean Square Error of Approximation (RMSEA) of 0.059, Comparative Fit Index (CFI) of 0.88, the goodness of Fit Index (GFI) of 0.87, Normal Fit Index (NFI) of 0.88, and Incremental Fit Index (IFI) of 0.8.

The factor loading, which indicates the standardized measure of the relationship between each item and its underlying construct, was presented at the top of the paths. T-values determined the significance of the correlation between domains. In terms of the "knowledge" domain, the highest and lowest loadings were observed in Item 5 (λ=0.677, T=12.45) and Item 9 (λ=0.310, T=6.88), respectively. The "attitude" domain exhibited the maximum loading in Item 20 (λ=0.655, T=11.76) and the minimum in Item 21 (λ=0.388, T=6.19). For the "practice" domain, the maximum loading was observed in Item 40 (λ=0.598, T=9.88), while the minimum was found in Item 39 (λ=0.343, T=5.46).

C. Reliability Assessment

The KAP-CQ39's reliability assessment showed strong internal consistency and reliability in measuring knowledge, attitude, and practices domains, as indicated by high Cronbach's alpha coefficients. Specifically, the knowledge domain had a Cronbach's alpha coefficient of 0.89, the attitude domain had a coefficient of 0.91, and the practices domain had a coefficient of 0.93. Moreover, the overall Cronbach's alpha coefficient for KAP-CQ39 of 0.91 demonstrates a high level of internal consistency.

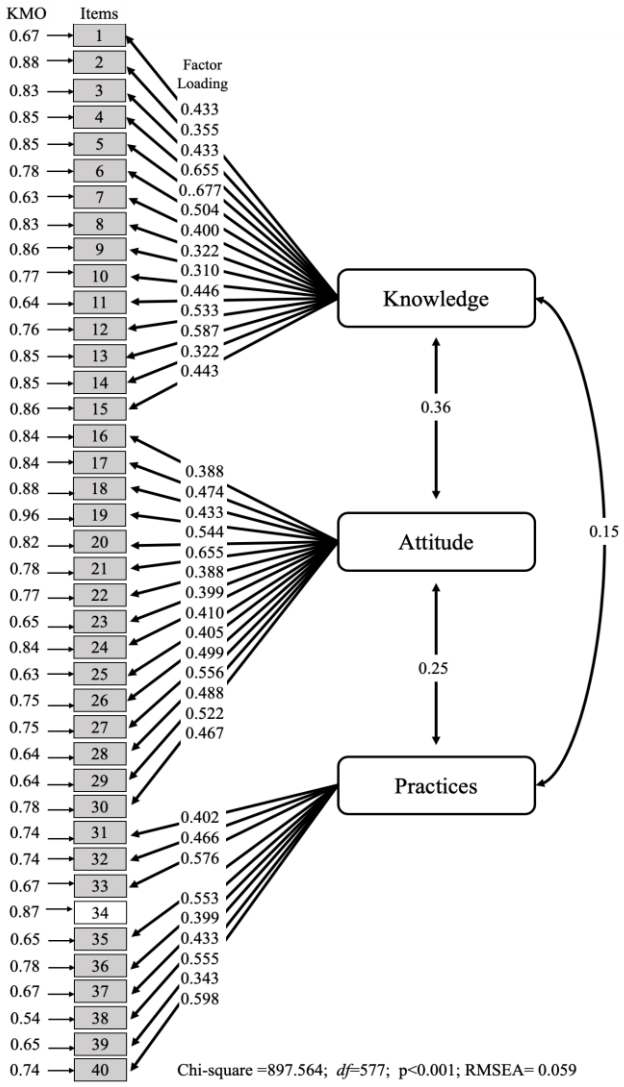

Fig. 3 Standardized factor loadings of the KAP-CQ39 model. Three constructs and 39 observed items were included in the Confirmatory Factor Analysis (CFA).

D. Final Questionnaire—KAP-CQ39

After undergoing five phases of development, the KAP-CQ39 questionnaire was finalized with five distinct sections. The first section consists of six items to determine the respondents' awareness of ChatGPT and to gather their sociodemographic information. The second section consists of 15 items presented in a true or false format and assessed the respondents' knowledge of ChatGPT, which were deemed reliable and valid. The third section is comprised of 15 reliable items that aimed to determine the attitudes of the respondents towards the use of ChatGPT in educational settings. The fourth section contained 9 items about the educational practices and activities involving ChatGPT. The final section is comprised of two additional questions that aimed to assess the respondents' perceptions regarding the advantages and disadvantages of using ChatGPT, as well as their views on regulating its use in academic institutions. Using KAP-CQ39, researchers and educational stakeholders assess the respondents' awareness of ChatGPT, their knowledge about it, their attitudes towards its use, and their practices and activities involving it. Additionally, KAP-CQ39 can help identify potential advantages and disadvantages of using ChatGPT and inform the development of regulations around its use in academic institutions.

IV. Discussion

Because of the rapid increase in usage and popularity of ChatGPT in educational settings, this study was conducted to develop an instrument to assess students' knowledge, attitudes, and practices regarding the use of ChatGPT. Following a thorough validation and assessment process, the final questionnaire (KAP-CQ39), consisting of 39 valid and reliable items related to the use of ChatGPT, was developed. Parmenter and Wardle [42] proposed that it is crucial to determine if a scale or questionnaire is valid and reliable, even for knowledge and behavior questions, to ensure it accurately measures what it intends to measure. Moreover, by assessing the validity and reliability of survey instruments, researchers can establish the extent to which the instrument is appropriate for their study and ensure that the data collected is valid and reliable. This helps ensure that the findings from the study are trustworthy and can be used to make informed decisions and recommendations [43−45]. Hence, this study evaluated the psychometric properties of KAP-CQ39, which include content validity, face validity, construct validity, and reliability to ensure its quality and internal consistency.

Content validity refers to the degree to which the questions included in the instrument reflect the full range of the concept or construct being measured [46]. To determine the content validity of KAP-CQ39, a group of experts was invited to evaluate and review each item of the instrument. From the original pool of 55 items, 15 were removed due to consistency, clarity, and appropriateness issues. Consequently, some items were modified, revised, and reformulated based on the experts' suggestions. The CVR and CVI values (>0.78) of the remaining 40 items indicate that all items were essential and appropriate for KAP-CQ39. According to Zamanzadeh [47], instruments with strong content validity are useful for the intended purposes of any study, and the data collected may provide a valid and accurate representation of the measured construct.

Although face validity is not a conclusive test of an instrument's validity, it is still a crucial aspect to consider as it may assist in spotting any apparent flaws with the instrument, such as irrelevant or poorly phrased items [48]. Twelve independent evaluators assessed the face validity of KAP-CQ39 using a three-point scale. Results showed that all items were straightforward and easy to understand, with Impact Scores (IS) of more than 1.5. Einola and Alvesson [49] emphasized that an instrument with high and acceptable face validity increases the likelihood that the respondents will answer the questions truthfully and accurately. It reduces the possibility of response bias or misinterpretation.

Construct validity was evaluated using Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) to confirm whether KAP-CQ39 items accurately measure each domain of knowledge, attitude, and practices regarding ChatGPT. Findings generated the three KAP domains and the corresponding factor loadings for each item. Item 34, "*I use ChatGPT to learn or teach a foreign language*," was removed

due to factor loadings of less than 0.3. This finding implies that this item was not deemed relevant or essential factor in understanding preservice teachers' practices related to ChatGPT. This result contradicts the claim of OpenAI, the developer of ChatGPT, who stated that the AI program could assist in foreign language learning [50]. The possible factors for this are the variation in the specialization of preservice teachers, and the absence of foreign language courses in the current teacher education curriculum. Still, the KAP-CQ39 demonstrated acceptable construct validity, as confirmed in the CFA with chi squared value of 897.564 (df=577, $p<0.001$). A KAP survey with good construct validity can be a valuable tool for identifying knowledge gaps, changing attitudes, and improving practices related to a particular issue [51]. It can also provide evidence-based policy and program development recommendations, making it an essential tool for researchers, policymakers, and practitioners [52].

Since the researchers intend to use KAP-CQ39 in future educational research, the reliability assessment was performed to ensure that the survey consistently measures the intended constructs over time and across different groups of respondents [53]. Results revealed that all domains (knowledge, attitude, and practices) had high Cronbach alpha coefficients of 0.81, 0.91, and 0.93, respectively. These findings denote that KAP-CQ39 items within each domain were highly correlated, indicating a high degree of internal consistency and reliability of the survey results. Zamanzadeh and Ghahramanian *et al.* [53] explained that a KAP survey with high internal consistency could be used to compare different populations or time points.

The study had several limitations that should be taken into account by both the readers and future researchers. Firstly, the research was confined to preservice teachers from a single public higher education institution in the Philippines. Secondly, while the KAP-CQ39 demonstrated reliability, various factors, such as digital literacy skills, internet accessibility, and implementation of educational technology programs in the university, could cause differences and fluctuations in the survey results. Thirdly, the survey was conducted online, which reduced the interaction with the respondents and made it more challenging to monitor their behavior while completing the KAP-CQ39. Lastly, because the responses were self-reported, the possibility of biased and untruthful responses cannot be ignored.

To the researchers' knowledge, this is the first study to develop and validate an instrument that aims to assess the educational use of ChatGPT. Hence, the development of KAP-CQ39 provides a systematic approach to evaluate the educational use of ChatGPT, which can help educators and researchers to identify the strengths and weaknesses of using this technology for educational purposes. The instrument can also be used to compare the educational use of ChatGPT with other AI-based educational tools, which can provide valuable insights into the effectiveness and efficiency of different AI-based educational tools.

V. Conclusion

In conclusion, the development and validation of KAP-CQ39 as an instrument to assess the educational use of ChatGPT may significantly contribute to the fields of artificial intelligence and education. This research may provide a systematic approach to evaluate the educational use of ChatGPT and enhance the credibility of the results obtained from the instrument. These findings may be substantial for educators and researchers, as they can provide valuable insights into how AI-based educational tools can be better designed and implemented. Therefore, the development of the KAP-CQ39 instrument is a significant step toward understanding and leveraging the academic potential of artificial intelligence. Additionally, it can inspire further research and innovation in the field of AI and education. To further advance the field of artificial intelligence and education, a potential future research recommendation is to utilize the KAP-CQ39 tool to assess other AI-based educational tools and compare their effectiveness. Additionally, conducting a longitudinal study to evaluate the lasting impact of ChatGPT on student learning outcomes could provide valuable insights.

Data Availability

The final version of KAP-CQ39 is available at https://doi.org/10.6084/m9.figshare.22285573.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

All authors were involved in conceptualizing the main research topic, data gathering, and evaluation. DAR wrote the initial draft, and CZ, SM, NG, AG, MGE, and EM revised the manuscript and supervised the writing process.

Acknowledgment

The authors would like to acknowledge the assistance and professional English editing services of Dr. Kristoffer Conrad Tejada, the participation of the preservice teachers from Batangas State University and the assistance in the validation of the experts from Ehime University, Japan, and De La Salle University- Manila, Philippines. This research is supported by the Department of Science and Technology – Science Education Institute (DOST-SEI).

References

[1] L. Chen, P. Chen, and Z. Lin, "Artificial intelligence in education: A review," *IEEE Access*, vol. 8, pp. 75264–75278, 2020, https://10.1109/ACCESS.2020.2988510

[2] M. M. Civaner, Y. Uncu, F. Bulut, E.G. Chalil, and A. Tatli, "Artificial intelligence in medical education: a cross-sectional needs assessment," *BMC Medical Education*, vol. 22, no. 1, p. 772, 2022, https://doi.org/10.1186/s12909-022-03852-3

[3] J. Lee, A. S. Wu, D. Li, and K. M. Kulasegaram, "Artificial intelligence in undergraduate medical education: A scoping review," *Academic Medicine: Journal of the Association of American Medical Colleges*, vol. 96, no. 11S, pp. S62–S70, 2022, https://doi.org/10.1097/ACM.0000000000004291

[4] J. Dahmen, M. E. Kayaalp, M. Ollivier *et al.*, "Artificial intelligence bot ChatGPT in medical research: The potential game changer as a double-edged sword," *Knee Surg Sports Traumatol Arthrosc*, 2023, https://doi.org/10.1007/s00167-023-07355-6

[5] E. Bilgic, A. Gorgy, A. Yang, M. Cwintal, H. Ranjbar, K. Kahla, D. Reddy, K. Li, H. Ozturk, E. Zimmermann, A. Quaiattini, A. S. Abbasgholizadeh-Rahimi, D. Poenaru, and J. M. Harley, "Exploring the roles of artificial intelligence in surgical education: A scoping review," *American journal of surgery*, vol. 224(1 Pt A), pp. 205–216, 2022, https://doi.org/10.1016/j.amjsurg.2021.11.023

[6] I. Khan, A. R. Ahmad, N. Jabeur *et al.*, "An artificial intelligence approach to monitor student performance and devise preventive measures," *Smart Learning Environments*, vol. 8, no. 17, pp. 1–14, 2021, https://doi.org/10.1186/s40561-021-00161-y.

[7] A. H. Verhoef, M. Fourie, Z. J. Rensburg *et al*., "The enhancement of academic integrity through a community of practice at the North-West University, South Africa," *International Journal for Educational Integrity*, vol. 18, no. 18, pp. 1–14, 2022, https://doi.org/10.1007/s40979-022-00115-y.

[8] What is ChatGPT: Definition, How to Use It. (2023). Jagranjosh.com. [Online]. Available: https://www.jagranjosh.com/general-knowledge/what-is-chat-gpt-1676870553-1

[9] L. Tung. (January 26, 2023). ChatGPT can write code. Now researchers say it's good at fixing bugs, too. ZDNET. [Online]. Available: https://www.zdnet.com/article/chatgpt-can-write-code-now-researchers-say-its-good-at-fixing-bugs-too/

[10] R. Gozalo-Brizuela and E. Garrido-Merch án. (2023). ChatGPT is not all you need. *A State-of-the-Art Review of large Generative AI Models*. [Online]. Available: https://arxiv.org/abs/2301.04655.

[11] A. Subaveerapandiyan, A. Vinoth, and N. Tiwary. (2023). Netizens, Academicians, and Information Professionals' Opinions about AI with Special Reference to ChatGPT. [Online]. Available: https://arxiv.org/abs/2302.07136.

[12] S. M. Ravizza, D. Z. Hambrick, and K. M. Fenn, "Non-academic internet use in the classroom is negatively related to classroom learning regardless of intellectual ability," *Computers & Education*, vol. 78, pp. 109–114, 2014. https://doi.org/10.1016/j.compedu.2014.05.007.

[13] A. Tlili, B. Shehata, M. A. Adarkwah *et al*., "What if the devil is my guardian angel: ChatGPT as a case study of using Chatbots in education," *Smart Learning Environments*, vol. 10, no. 15, pp. 1-13, 2023. http://doi.org/10.1186/s40561-023-00237-x.

[14] E. Kasneci, K. Seßler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, and G. Kasneci, *ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education*, 2023, https://doi.org/10.35542/osf.io/5er8f.

[15] E. Vandamme, "Concepts and challenges in the use of Knowledge-Attitude-Practice surveys: Literature review," *Strateg. Netw. Negl. Dis. Zoonoses*, Department of Animal Health, Institute of Tropical Medicine, vol. 1, 2009.

[16] H. C. Koo, B. K. Poh, and A. T. Ruzita, "Development, validity and reliability of a questionnaire on knowledge, attitude and practice (KAP) towards whole grain among primary school children in Kuala Lumpur, Malaysia," *International Food Research Journal*, vol. 23, pp. 797-805, 2023.

[17] D. A. Robledo, "Biology at Home: The Six Attributes of Home-based Biology Experiments (HBEs) for Remote Authentic Learning," *Psychology and Education*, 2021, vol. 58, no. 3, pp. 4319–43123, 2021, https://ssrn.com/abstract=3836530.

[18] D. A. Robledo, A. Lapada, F. Miguel, and Z. Alam, "COVID-19 vaccine confidence and hesitancy among schools' stakeholders: A Philippine survey," *Journal of Cardiovascular Disease and Research*, vol. 12, issue 3, pp. 29–35, 2021, https://ssrn.com/abstract=3846821.

[19] A. A. Lapada, F. F. Miguel, D. A. R. Robledo, and Z. F. Alam, "Teachers' Covid-19 awareness, distance learning education experiences and perceptions towards institutional readiness and challenges," *International Journal of Learning, Teaching and Educational Research*, vol. 19, no. 6, pp. 127–144, 2020. https://doi.org/10.26803/ijlter.19.6.8

[20] N. Iivari, S. Sharma, and L. Vent ä-Olkkonen, "Digital transformation of everyday life – How COVID-19 pandemic transformed the basic education of the young generation and why information management research should care?" *International Journal of Information Management*, vol. 55, pp. 102183–102183, 2020.

[21] D. A. Robledo, G. Muhammad, and A. U. Rahman, "The perceptions of pre-service teachers on bachelor of education program in Shaheed Benazir Bhutto University, Pakistan," *Journal* of *Social Sciences Review*, vol. 1, no. 1, pp. 1–12, 2021. https://doi.org/10.54183/jssr.v1i1.11

[22] D. A. Robledo, E. Catapang, A. Motin, and E. Maalihan, "Teaching beyond borders: Effectiveness of heuristic approach in teaching science in public secondary schools in area IV, division of Batangas, Philippines" *International Engineering Journal For Research & Development*, vol. 6, no. 2, May 4, 2021, https://ssrn.com/abstract=3839132

[23] B. du Boulay, "Artificial intelligence as an effective classroom assistant," *IEEE Intelligent Systems*, vol. 31, no. 6, pp. 76–81, Nov.-Dec. 2016, https://doi.org/10.1109/MIS.2016.9

[24] D. Robledo and M. Prudente, "Effectiveness of simulation-based activities vs. traditional interventions in teaching biology: A meta-analysis," *Turkish Online Journal of Qualitative Inquiry*, 2020. https://doi.org/10.5281/zenodo.5824594

[25] D. Robledo and M. Prudente, "A virtual fieldtrip: Effects of Google Earth Learning Activities (GELA) on students' environmental awareness and environmental attitudes," in *Proc. 2022 13th International Conference on E-Education, E-Business, E-Management, and e-Learning (IC4E) (IC4E 2022), Association for Computing Machinery,* New York, NY, USA, 2022, pp. 1–8, https://doi.org/10.1145/3514262.3514293

[26] D. M. Kagan, "Professional growth among preservice and beginning teachers," *Review of Educational Research*, vol. 62, no. 2, pp. 129–169, 1992, https://doi.org/10.3102/00346543062002129

[27] M. S. Setia, "Methodology series module 3: Cross-sectional studies," *Indian Journal of Dermatology*, vol. 61, no. 3, pp. 261–264, 2016. https://www.e-ijd.org/text.asp?2016/61/3/261/182410

[28] H. C. Koo, B. K. Poh, and A. T. Ruzita, "Development, validity and reliability of a questionnaire on knowledge, attitude and practice (KAP) towards whole grain among primary school children in Kuala Lumpur, Malaysia," *International Food Research Journal*, vol. 23, pp. 797–805, 2016.

[29] A. W. Harzing. (2007). Publish or Perish. [Online]. Available: https://harzing.com/resources/publish-or-perish

[30] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, pp. 101–77, 2006.

[31] I. Krumpal, "Determinants of social desirability bias in sensitive surveys: a literature review," *Qualitative and Quantitative Research*, vol. 47, pp. 2025–2047, 2013. https://doi.org/10.1007/s11135-011-9640-9

[32] I. B. Rodrigues, J. D. Adachi, K. A. Beattie *et al.*, "Development and validation of a new tool to measure the facilitators, barriers and preferences to exercise in people with osteoporosis," *BMC Musculoskeletal Disorders*, vol. 18, p. 540, 2017, https://doi.org/10.1186/s12891-017-1914-5

[33] J. Shi, X. Mo, and Z. Sun, "Content validity index in scale development," *Journal of Central South University*, vol. 37, no. 2, pp. 152–155, 2012. https://doi.org/10.3969/j.issn.1672-7347.2012.02.007

[34] F. Koohi, P. Amiri, Y. Mehrabi *et al*., "Development and validation of a knowledge, attitude, and practice questionnaire regarding cardiovascular diseases in an Iranian general population," *BMC Public Health*, vol. 21, p. 2050, 2021, https://doi.org/10.1186/s12889-021-12135-3

[35] R. A. Mohamed, N. A. A. Rahim, S. M. Mohamad *et al*., "Validity and reliability of knowledge, attitude, and practice regarding exercise and exergames experiences questionnaire among high school students," *BMC Public Health*, vol. 22, p. 1743, 2022, https://doi.org/10.1186/s12889-022-14147-z

[36] T. Zhang, C. Yin, Y. Geng, Y. Zhou, S. Sun, and F. Tang, "Development and validation of psychological contract scale for hospital pharmacists," *Journal of Multidisciplinary Healthcare*, vol. 13, pp. 1433–1442, 2020, https://doi.org/10.2147/JMDH.S270030

[37] T. R. Knapp and V. H. Swoyer, "Some empirical results concerning the power of Bartlett'S test of the significance of a correlation matrix," *American Educational Research Journal*, vol. 4, no. 1, pp. 13–17, 1967, https://doi.org/10.3102/00028312004001013

[38] R. A. Peterson, "A meta-analysis of variance accounted for and factor loadings in exploratory factor analysis," *Marketing Letters*, vol. 11, pp. 261–275, 2000, https://doi.org/10.1023/A:1008191211004

[39] A. L. Burton, P. Hay, S. Kleitman, E. Smith, J. Raman, J. Swinbourne, S. W. Touyz, and M. J. Abbott, "Confirmatory factor analysis and examination of the psychometric properties of the eating beliefs questionnaire," *BMC Psychiatry*, vol. 17, no. 1, p. 237, 2017, https://doi.org/10.1186/s12888-017-1394-z

[40] H. W. Marsh, J. Guo, T. Dicke, P. D. Parker, and R. G. Craven, "Confirmatory Factor Analysis (CFA), Exploratory Structural Equation Modeling (ESEM), and Set-ESEM: Optimal balance between goodness of fit and parsimony," *Multivariate Behavioral Research*, vol. 55, no. 1, pp. 102–119, 2020, https://doi.org/10.1080/00273171.2019.1602503

[41] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis*, vol. 7, 2010.
[42] K. Parmenter and J. Wardle, "Evaluation and design of nutrition knowledge measures," *Journal of Nutrition Education*, vol. 32, no. 5, pp. 269–277, 2000, https://doi.org/10.1016/S0022-3182(00)70575-9
[43] K. Kaliyaperumal, "Guideline for conducting a knowledge, attitude, and practice (KAP) study," *AECS Illumination*, vol. 4, no. 1, pp. 7–9, 2004.
[44] C. L. Kimberlin and A. G. Winterstein, "Validity and reliability of measurement instruments used in research," *American Journal of Health-System Pharmacy*, vol. 65, no. 23, pp. 2276–2284, 2008. https://doi.org/10.2146/ajhp070364
[45] H. Taherdoost. (August 10, 2016). Validity and reliability of the research instrument; How to test the validation of a questionnaire/survey in a research. [Online]. Available: http://dx.doi.org/10.2139/ssrn.3205040
[46] J. Connell, J. Carlton, A. Grundy *et al*., "The importance of content and face validity in instrument development: Lessons learnt from service users when developing the Recovering Quality of Life measure (ReQoL)," *Quality of Life Research*, vol. 27, pp. 1893–1902, 2018. https://doi.org/10.1007/s11136-018-1847-y
[47] A. Althubaiti, "Information bias in health research: definition, pitfalls, and adjustment methods," *Journal of Multidisciplinary Healthcare*, vol. 9, pp. 211–217, 2016, https://doi.org/10.2147/JMDH.S104807
[48] V. Zamanzadeh, A. Ghahramanian, M. Rassouli, A. Abbaszadeh, H. Alavi-Majd, and A. R. Nikanfar, "Design and implementation content validity study: Development of an instrument for measuring patient-centered communication," *Journal of Caring Sciences*, vol. 4, no. 2, pp. 165–178, 2015, https://doi.org/10.15171/jcs.2015.017
[49] K. Einola and M. Alvesson, "Behind the numbers: Questioning questionnaires," *Journal of Management Inquiry*, vol. 30, no. 1, pp. 102–114, 2021, https://doi.org/10.1177/1056492620938139
[50] ChatGPT can be powerful tool for language learning. [Online]. Available: University Affairs. https://www.universityaffairs.ca/career-advice/career-advice-article/chatgpt-can-be-powerful-tool-for-language-learning/
[51] C. Andrade, V. Menon, S. Ameen, and S. K. Praharaj, "Designing and conducting knowledge, attitude, and practice surveys in psychiatry: Practical guidance," *Indian Journal of Psychological Medicine*, vol. 42, no. 5, pp. 478–481, 2020, https://doi.org/10.1177/0253717620946111
[52] M. Lee, B. A. Kang, and M. You, "Knowledge, attitudes, and practices (KAP) toward COVID-19: A cross-sectional study in South Korea," *BMC Public Health*, vol. 21, p. 295, 2021, https://doi.org/10.1186/s12889-021-10285-y
[53] V. Zamanzadeh, A. Ghahramanian, M. Rassouli, A. Abbaszadeh, H. Alavi-Majd, and A. R. Nikanfar, "Design and Implementation content validity study: Development of an instrument for measuring patient-centered communication," *Journal of Caring Sciences*, vol. 4, no. 2, pp. 165–178, 2015, https://doi.org/10.15171/jcs.2015.017