

Enhancing Dementia Care Education in Nursing: A Pilot Study on Virtual Teaching and Concept Mapping with ChatGPT

Pamela Diaz*
Helene Fuld College of Nursing, New York City University
24 East 120th Street, NY, New York 10035
E-mail: pamela.diazulloa@helenefuld.edu

Erica Germain
Helene Fuld College of Nursing, New York City University
24 East 120th Street, NY, New York 10035
E-mail: erica.germain@helenefuld.edu

Yejoo Hwang
Helene Fuld College of Nursing, New York City University
24 East 120th Street, NY, New York 10035
E-mail: yejoo.hwang@helenefuld.edu

Juliet Kadriu (Corresponding author)
Helene Fuld College of Nursing, New York City University
24 East 120th Street, NY, New York 10035
E-mail: juliet.kadru@helenefuld.edu

Abstract
The study aims to demonstrate the effectiveness of a 30-minute virtual teaching session via Microsoft Teams, supplemented by ChatGPT-aided concept mapping under a quasi-experimental design. Sixteen senior nursing students demonstrated improvements in both knowledge and caring behavior scores. Statistical analysis revealed a significant enhancement in knowledge retention from posttest 1 to posttest 2, with substantial evidence supporting this result: $p = .014$ for the average posttest 1 (Day 1), average posttest 2 (Day 1), and $p < .001$ for the average posttest 2 (Day 1) and average posttest 2 (Day 2). Moreover, the effect sizes for these paired samples indicate medium to high effects, with Cohen's d values of 0.698 and 1.415, respectively. This study offers valuable insights into student-led virtual teaching sessions, especially demonstrating how ChatGPT can effectively clarify concepts related to dementia care.
Keywords: virtual teaching, concept map, ChatGPT, dementia, nursing education
DOI: 10.7176/JHMN/115-03
Publication date: April 30th 2024

1. Background

The use of virtual teaching combined with the use of artificial intelligence (AI) has become a very modern day and common practice in our society. There were many benefits to using AI apps such as ChatGPT, but also there are various drawbacks and bias'. "ChatGPT is an AI trained interactive conversational model chatbot capable of responding to prompts in various text formats" (Achibald & Clark, 2023).

Using AI allows for immediate feedback of work, takes the workload off of the professor, as well as allowing for many different learning styles that students have. "ChatGPT offers personalized learning, immediate feedback and simulation scenarios that can increase student engagement and knowledge acquisition" (Liu 2023). Among the drawbacks to virtual teaching and AI use are the lack of human interaction which may affect the absorption of materials. In addition, the accuracy of information is at risk when using AI and virtual learning. A major worry in nursing education is that students who rely too much on ChatGPT and other AI programs may acquire inaccurate knowledge about a concept or subject, which would then make it more difficult to administer safe and effective care.

Within the nursing field, it is important to ensure adequate and accurate information and often AI is used unethically, violating the college's policy. The use of AI in any and all parts of this research project will be in full compliance with all of Helene Fuld College of Nursing policies.

2. Methods

2.1. Research Design

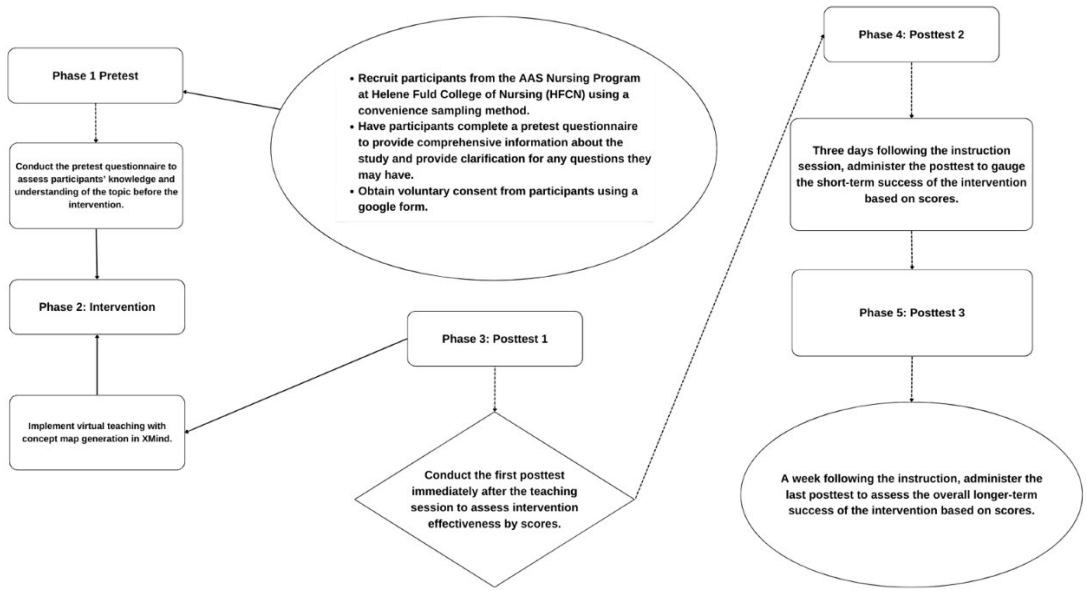
A one-group, pretest and multiple posttest (quasi-experimental) design was adopted.

Phase 0 (Study Preparation). This study was constructed and performed to answer the research question: "How

does virtual teaching about older persons' needs with concept mapping using ChatGPT (E) influence the expression of caring knowledge and skills (O) among senior nursing students (P)?". We performed a comprehensive literature review from 2021-2023 targeting educational strategies within nursing programs covering the use of virtual education and webinars (AKA web seminars) for student instruction. The study design was subjected to a thorough review and approval process by Helene Fuld College's research ethics committee.

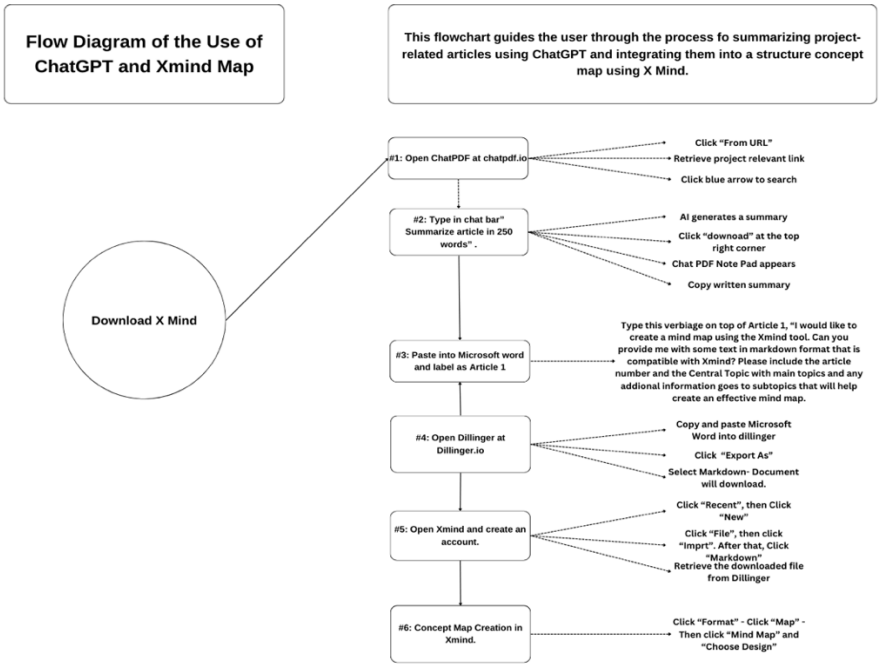
Figure 1 *Flow Diagram of Research Design and Procedures*

Hypothesis: Third-year BSN students would have increased knowledge of caring for the elderly with dementia after receiving an educational intervention using virtual teaching, from pre-test to post-test 2.



This flowchart guides the presentation of our study to nursing students at Helene Fuld College, illustrating each step in a sequential manner. It provides a clear visual representation of the study's unfolding process.

Figure 2 *Flow Diagram of the Use of ChatGPT*



The above flow chart will guide the user through the process of summarizing our project-related articles using ChatGPT and integrating them into a structured concept map using Xmind.

2.2. Participants/Setting

Phase 1 (Student Participants Recruitment and Pretest). Invitations to potential participants were sent via email and text, which provided comprehensive information, instructions, and clarification.
This directed them to an online pretest questionnaire structured as a Google Form survey. A total of 16 responses were returned.

2.3. Data-Gathering & Teaching Procedures

Phase 2 (Intervention). A week after the pretest, we implemented virtual teaching with a concept map generated in Xmind for the participants which was limited to 5 participants per session via Microsoft Teams meetings. In addition, pre-recorded educational videos about general wellness for the geriatric client population were included in the virtual teaching. There was 1.5 hours or 90 minutes of non-credit instruction provided at 30 minutes per session over 3 days. Study participants were asked to complete three posttests consisting of the caring behavior questionnaire: immediately, three days after, and a week after the virtual teaching sessions.

2.4. Questionnaires

This research utilized ten questionnaires (one pretest and nine posttest). Phase 3 (Posttest 1): We conducted the first posttest immediately after each of the 3 teaching sessions to assess intervention effectiveness by scores and, again, received 16 responses for each of the 3 days. Phase 4 (Posttest 2): Three days after the instruction session, we administered the 2nd posttest to gauge the intervention's short-term success based on the scores. The questions for the 2nd posttest were slightly more difficult and required more interpretive thought than the questions on posttest 1. Again, we received 16 responses for each of the 3 teaching days. Phase 5 (Posttest 3): A week following the instruction, we administered the last posttest to assess the student volunteer's knowledge retention and thus the overall longer-term success of the intervention based on scores. As in Posttest 2, the questions asked in Posttest 3 were somewhat more complicated thus requiring a degree of critical thinking to answer. As in the previous tests, we received 16 responses for each of the 3 teaching days.

2.4.1. Scoring & Interpretation Using the Likert Scale

The 5-item Likert scale measures opinions, or behaviors. In both the pre and posttests, there are five possible responses ranging from 'Never' to 'Always' for each question. The Likert scale has symmetrical sides, which means they have an equal number of positive and negative spots within both poles.

2.5. Ethical Considerations

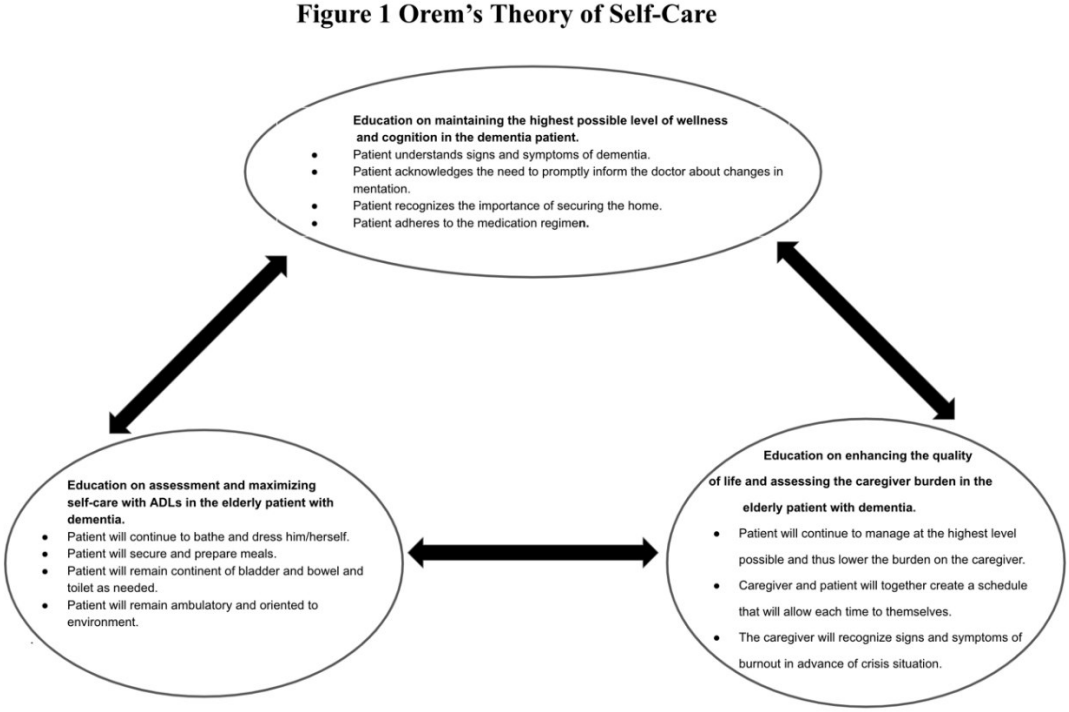
This research was cleared by the Provost Dr. Sandy Carollo of Helene Fuld College of Nursing.

2.6. Data Analysis

Jamovi version 2.3 was used to analyze the data.

Figure 3

Conceptual Diagram in the Study based on Orem's Theory of Self-Care



This program educates dementia patients and caregivers about recognizing symptoms, seeking medical help, home safety, and medication adherence to optimize well-being and cognitive function. This educational session focuses on helping dementia patients with self-care tasks like bathing, dressing, and toileting, while aiming to maintain their independence and mobility at home. Educating caregivers of dementia patients on improving quality of life and managing their own burden empowers them to create a sustainable care plan that benefits both themselves and the patient.

3. Results

Table 1 presents a comparative analysis of the mean results obtained by students in posttest 1 (PT1) and posttest 2 (PT2), with reference to two different teaching days (TD), TD1 and TD2. The statistical analysis conducted on these results yielded notable findings, particularly regarding the *p*-values and effect sizes.

Firstly, the *p*-values associated with the comparison between the average scores of PT1 and PT2 under TD1 ($p = .014$) and TD2 ($p < .001$) indicate statistical significance. Specifically, the *p*-value of .014 for PT1 vs. PT2 under TD1 suggests that there is a significant difference between the mean scores of these two tests for students exposed to teaching design 1. Similarly, the *p*-value of less than .001 for PT1 vs. PT2 under TD2 implies a highly significant disparity in the mean scores between the two posttests for students following teaching design 2.

Furthermore, the effect sizes, measured using Cohen's d statistic, provide additional insights into the practical significance of the observed differences. For the comparison between PT1 and PT2 under TD1, the effect size is calculated to be 0.698, indicating a medium effect. This suggests that the difference in mean scores between the two posttests for students in teaching design 1 is of moderate practical importance. Conversely, for the comparison between PT1 and PT2 under TD2, the effect size is notably higher at 1.415, indicating a large effect. This signifies that the disparity in mean scores between the two posttests for students in teaching design 2 is of substantial practical significance, demonstrating a more pronounced impact of the instructional approach on student performance.

Table 2 regarding the mean results of the pre-test, posttest 1, and posttest 2 initially suggested a deviation from normality in the distribution of the data. However, upon conducting a normality test, it became evident that all the calculated p-values associated with the normality test exceeded the threshold of .05, indicating that the data adheres to a normal distribution.

This revelation holds considerable significance as it implies that the data collected from the pre-test, posttest 1, and posttest 2 indeed exhibit a normal distribution pattern. In statistical terms, a normal distribution signifies that the data points are symmetrically distributed around the mean, with the majority of values clustered close to the center and fewer values dispersed towards the extremes. This distribution conforms to the characteristic bell-shaped curve, which is a hallmark of normality.

The confirmation of normality in the data has several implications for subsequent statistical analyses. Firstly, it validates the application of parametric statistical tests, such as t-tests, analysis of variance (ANOVA), or linear regression, which rely on the assumption of normality for accurate inference. These tests are commonly utilized for comparing means, assessing relationships between variables, and making predictions based on regression models, and the assumption of normality is fundamental for their validity.

Table 3 presents a comprehensive analysis of the reliability of pre-test questions utilized in evaluating the GBS Nursing Students. The mean score of 2.79 with a standard deviation of 0.466 indicates the central tendency and variability within the responses. Cronbach's alpha coefficient of 0.926 demonstrates high internal consistency reliability, suggesting that the items in the pre-test questionnaire reliably measure the same underlying construct. Furthermore, McDonald's omega coefficient of .952 reaffirms the robustness of the scale. Notably, the negative correlation of Pre-Test Question 2 with the total scale highlights a potential need for item reversal to align its scoring direction with the intended construct. This comprehensive analysis underscores the reliability and validity of the pre-test instrument in assessing GBS Nursing Students' proficiency effectively.

Table 4 provides a detailed overview of the demographic composition of the nursing school participants, consisting of a total of 16 individuals. The distribution across age groups reveals a varied representation, with 2 participants falling within the 18-24 age bracket, 7 participants aged 25-34, and an equal number aged 35-44, indicating a diverse range of experiences and perspectives. Gender distribution indicates a predominance of female participants, with 13 females compared to 3 males. Marital status diversity is evident, with a significant portion being single or never married (7), followed by married participants (6), and a smaller number in domestic partnerships (2), and one individual who is divorced. Additionally, the ethnic composition showcases a mix of backgrounds, with 5 participants identifying as Hispanic or Latino, while the majority, comprising 11 participants, identify as Black or African American, reflecting the rich diversity within the nursing school cohort.

4. Discussion

Our study investigated the effectiveness of virtual teaching in conjunction with concept mapping and ChatGPT in educating nursing students about caring for the elderly with dementia.

The results, as shown in Table 1, indicate a significant improvement in knowledge retention between posttest 1 and posttest 2, with strong evidence supporting this finding (p = .014 for Average PT1 (TD1), Average PT2 (TD1), and < .001 for Average PT2 (TD1) and Average PT2 (TD2). This demonstrates a degree of hypothesis concept proof and at this point our null hypothesis is not rejected.

Furthermore, the effect sizes for these paired samples suggest medium to high effects (Cohen's d = 0.698 and 1.415, respectively). In the context of the study, the effect sizes for the paired samples provide insights into the practical significance or magnitude of the observed differences between the variables being compared. Cohen's d, a commonly used measure of effect size, quantifies the standardized difference between two means.

In this case, an effect size of 0.698 for one set of paired samples and 1.415 for another set indicates the extent to which the mean scores of the two groups differ, relative to the variability within each group.

A Cohen's d of 0.698 suggests a medium effect size. This means that the difference in mean scores between the two groups is of moderate practical significance. It implies that the teaching provided to students had a noticeable impact on their knowledge.

On the other hand, a Cohen's d of 1.415 represents a high effect size. This indicates that the difference in mean scores between the two groups is substantial and has a significant practical importance. In other words, the teaching on both days had a pronounced and meaningful impact on teaching knowledge.

Table 2 reveals non-normality in the data distribution, indicating potential challenges in interpretation. However, normality tests in Table 2 demonstrate that the data is indeed normal, providing confidence in the statistical analyses, indicating that our hypothesis can be confirmed.

In line with our results, previous research, such as the study by Chen et al. (2020) and Li et al. (2021), has shown the efficacy of virtual reality and online education in improving learning outcomes and health-related metrics. However, contrasting views exist, as evidenced by the study conducted by Hsu (2022), which highlights concerns about the usefulness and adequacy of online education compared to traditional classroom instruction. Despite these conflicting perspectives, our study supports the hypothesis that ChatGPT-assisted virtual instruction enhances knowledge retention among nursing students, particularly regarding dementia care for the elderly. These findings contribute to ongoing debates surrounding the success of online education and pave the way for future research endeavors of broader scope and scale. Although our study proved our hypothesis, it was limited in size and scope and going forward both areas need to be enhanced.

5. Conclusions

Through the use of ChatGPT, an AI-powered virtual teaching tool, nursing care expertise can be improved. According to our research, when important elements are connected, nursing students can improve the compassionate behaviors they use to meet the needs of people with dementia. We have addressed the limits of

our study, which encompass sample size, data collection methods, and research design constraints. By emphasizing the useful applications of AI technology, this study expanded on the body of evidence supporting its use in nursing education. Our hypothesis was proven in that third-year nursing students demonstrated improvement in all three subscales of the questionnaire from pre-test to posttest 2. Pre-test--"Virtual teaching about caring for the older person with dementia.", Post-test 1--"Education on maintaining the highest possible level of wellness and cognition in the dementia patient.", Post-test 2--"Develop strategies for promoting and maintaining daily activities in the elderly patient with dementia, while maintaining a maximum level of safety for the client and the caregiver.".

Informed Consent Statement

Each and every study participant submitted their informed consent.

References

1. Archibald, M. M., & Clark, A. M. (2023). ChatGTP: What is it and how can nursing and health science education use it? Journal of Advanced Nursing. https://doi.org/10.1111/jan.15643
2. Arof, K. Z. M., Ismail, S., & Saleh, A. L. (2018). Contractor's performance appraisal system in the Malaysian construction industry: Current practice, perception and understanding. International Journal of Engineering and Technology (UAE), 7(3), 46-51. https://doi.org/10.14419/ijet.v7i3.9.15272
3. Bean, J. (2021, October 13). Using R in social work research. The Ohio State University. https://bookdown.org/bean_jerry/bookdown_r_for_social_workers/confirmatory-factor-analysis.html
4. Chen, F. Q., Leng, Y. F., Ge, J. F., Wang, D. W., Li, C., Chen, B., & Sun, Z. L. (2020). Effectiveness of virtual reality in nursing education: Meta-analysis. Journal of Medical Internet Research, 22(9), e18290. https://doi.org/10.2196/18290
5. Cheung, G. W., Cooper-Thomas, H. D., Lau, R. S., & Wang, L. C. (2023). Reporting reliability, convergent and discriminant validity with structural equation modeling: A review and best- practice recommendations. Asia Pacific Journal of Management, 1-39. https://doi.org/10.1007/s10490-023-09871-y
6. Dragan, D., & Topolšek, D. (2014, June). Introduction to structural equation modeling: Review, methodology and practical applications. In The International Conference on Logistics & Sustainable Transport (Vol. 6, pp. 19-21). https://www.researchgate.net/profile/Dejan-Dragan-2/publication/267810858_Introduction_to_Structural_Equation_Modeling_Review_Methodology_and_Practical_Applications/links/545acf900cf25c508c319c8b/Introduction-to-Structural-Equation-Modeling-Review-Methodology-and-Practical-Applications.pdf
7. Faradillah, A., & Adlina, S. (2021). Validity of critical thinking skills instrument on prospective Mathematics teachers. Jurnal Penelitian dan Evaluasi Pendidikan, 25(2), 126-137. http://dx.doi.org/10.21831/pep.v25i2.40662
8. Grove, S. K., Burns, N., & Gray, J. (2023). *The practice of nursing research: Appraisal, synthesis, and generation of evidence* (9th ed.). Saunders.
9. Herft, A. (2023). A teacher's prompt guide to ChatGPT aligned with 'What Works Best' guide. https://usergeneratededucation.files.wordpress.com/2023/01/a-teachers-prompt-guide-to-chatgpt-aligned-with-what-works-best.pdf
10. Hsu, P. T., & Ho, Y. F. (2022, August). Effects of online teaching during the COVID-19 pandemic on nursing students' intention to join the nursing workforce: A cross-sectional study. *Healthcare*, 10(8), 1461. https://doi.org/10.3390/healthcare10081461
11. Jennifer R, G., & Susan K, G. (2021). *Burns and Grove's the practice of nursing research appraisal, synthesis, and generation of evidence* (9th ed.) (ISBN: 9780323673174). Elsevier, Inc.
12. Josephsen, J., Ketelsen, K., Weaver, M., & Scheuffele, H. (2023). Dementia care competency model for higher education: A pilot study. *International Journal of Environmental Research and Public Health, 20*(4), 3173. https://doi.org/10.3390/ijerph20043173
13. Li, F., Harmer, P., Voit, J., & Chou, L. S. (2021). Implementing an online virtual falls prevention intervention during a public health pandemic for older adults with mild cognitive impairment: A feasibility trial. *Clinical Interventions in Aging*, 973-983. https://doi.org/10.2147/CIA.S306431
14. Liao, H. C., Huang, C. Y., & Wang, Y. H. (2022). Development and psychometric testing of a scale measuring caring behaviors for healthcare students and providers. Medical Education Online, 27(1), 2066496. https://doi.org/10.1080/10872981.2022.2066496
15. Liu, J., & Liu, S. (2023). The application of ChatGPT in nursing education. https://osf.io/2x7a8/download/?format=pdf
16. Liu, J., Liu, F., Fang, J., & Liu, S. (2023). The application of chat generative pre-trained transformer in nursing education. *Nursing Outlook, 71*(6). https://doi.org/10.2196/48254
17. Li, N., Huang, J., & Feng, Y. (2020). Construction and confirmatory factor analysis of the core cognitive

ability index system of ship C2 system operators. Plos One, 15(8), e0237339. https://doi.org/10.1371/journal.pone.0237339
18. Malkewitz, C. P., Schwall, P., Meesters, C., & Hardt, J. (2023). Estimating reliability: A comparison of Cronbach's α, McDonald's ωt and the greatest lower bound. *Social Sciences & Humanities Open, 7*(1), 100368. https://doi.org/10.1016/j.ssaho.2022.100368
19. McEwan, M., & Wills, E. M. (2021). *Theoretical basis for nursing* (6th ed.) (ISBN: 9781975175658). Lippincott Williams & Wilkins.
20. Nguyen, T. T. P., Truong, G. T. H., Nguyen, H. T. T., Nguyen, C. T. T., Dinh, D. X., & Nguyen, B. T. (2022). An instrument for measuring job satisfaction (VIJS): A validation study for community pharmacists in the context of the COVID-19 pandemic in Vietnam. PLoS One, 17(11), e0276918. https://doi.org/10.1371/journal.pone.0276918
21. Paulsen, J. & BrckaLorenz, A. (2017). Internal consistency. FSSE Psychometric Portfolio. https://scholarworks.iu.edu/dspace/handle/2022/24498
22. Saab, M. M., Hegarty, J., Murphy, D., & Landers, M. (2021). Incorporating virtual reality in nurse education: A qualitative study of nursing students' perspectives. *Nurse Education Today, 105*, 105045. https://doi.org/10.1016/j.nedt.2021.105045
23. Sallam, M., Salim, N. A., Barakat, M., Al-Mahzoum, K., Ala'a, B., Malaeb, D., Hallit, R. & Hallit, S. (2023). Assessing health students' attitudes and usage of ChatGPT in Jordan: Validation study. *JMIR Medical Education, 9*(1), e48254. https://doi.org/10.2196/48254
24. Shorey, S., & Ng, E. D. (2021). The use of virtual reality simulation among nursing students and registered nurses: A systematic review. *Nurse Education Today, 98*, 104662. https://doi.org/10.1016/j.nedt.2020.104662
25. Tam, W., Huynh, T., Tang, A., Luong, S., Khatri, Y., & Zhou, W. (2023). Nursing education in the age of artificial intelligence powered Chatbots (AI-Chatbots): Are we ready yet?. *Nurse Education Today, 129*, 105917. https://doi.org/10.1016/j.nedt.2023.105917
26. Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions: Congeneric and asymmetrical measurements. *Frontiers in Psychology, 7*, 769. https://doi.org/10.3389/fpsyg.2016.00769
27. Verdugo-Alonso, M. A., Henao-Lema, C. P., Córdoba-Andrade, L., & Arias González, V. B. (2017). Dimensionality and internal structure of the Colombian version of the INICO-FEAPS quality of life scale. Journal of Intellectual Disability Research, 61(12), 1094-1103. https://www.researchgate.net/publication/320326481_Dimensionality_and_internal_structure_of_the_Colombian_version_of_the_INICO-FEAPS_quality_of_life_scale
28. World Health Organization. (2017). Integrated care for older people: Guidelines on community-level interventions to manage declines in intrinsic capacity. https://iris.who.int/bitstream/handle/10665/258981/9789241550109-eng.pdf?sequence=1
29. Yee, K., Whittington, K., Doggette, E., & Uttich, L. (2023). ChatGPT assignments to use in your classroom today. UCF Created OER Works. 8. https://stars.library.ucf.edu/cgi/viewcontent.cgi?article=1097&context=oer

Table 1: Paired Sample T-Test

			Statistic	df	p	Mean difference	SE difference		Effect Size
Average Pre-Test	Average PT1 (TD1)	Student's t	0.849	15.0	0.409	0.154	0.182	Cohen's d	0.212
		Wilcoxon W	82.50		0.469	0.117	0.182	Rank biserial correlation	0.213
Average PT1 (TD1)	Average PT1 (TD2)	Student's t	2.791	15.0	0.014	1.300	0.466	Cohen's d	0.698
		Wilcoxon W	89.50 [a]		0.022	1.600	0.466	Rank biserial correlation	0.705
Average PT1 (TD2)	Average PT2 (TD1)	Student's t	-1.571	15.0	0.137	-0.564	0.359	Cohen's d	-0.393
		Wilcoxon W	36.00 [b]		0.182	-0.643	0.359	Rank biserial correlation	-0.400
Average PT2 (TD1)	Average PT2 (TD2)	Student's t	5.661	15.0	< .001	1.491	0.263	Cohen's d	1.415
		Wilcoxon W	105.00 [a]		0.001	1.643	0.263	Rank biserial correlation	1.000

Table 2: Normality Test

Normality Test (Shapiro-Wilk)

			W	p
Average Pre-Test	-	Average PT1 (TD1)	0.971	0.859
Average PT1 (TD1)	-	Average PT1 (TD2)	0.924	0.192
Average PT1 (TD2)	-	Average PT2 (TD1)	0.931	0.254
Average PT2 (TD1)	-	Average PT2 (TD2)	0.965	0.751

Normality Test (Shapiro-Wilk)

			W	p
Average PT2 (TD2)	-	Average PT2 (TD3)	0.910	0.181

Note. A low p-value suggests a violation of the assumption of normality

Table 3: Reliability Analysis of GBS Nursing Students

Scale Reliability Statistics

	Mean	SD	Cronbach's α	McDonald's ω
scale	2.79	0.466	0.926	0.952

Note. item 'PreTest Q2' correlates negatively with the total scale and probably should be reversed

Table 4: Demographical Information

Demographical Factors	***n***	**%**
Nursing Programs		
Associate degree	00	00.0
Generic Bachelors' degree	16	100.0
RN to Bachelors' degree	00	00.0
Age (Years)		
18 – 24	2	12.5
25 – 34	7	43.75
35 – 44	7	43.75
45 – 54	00	00.0
≥ 55	00	00.0
Gender		
Male	3	18.75
Female	13	81.25
Other	00	00.0
Prefer not to say	00	00.0
Marital Status		
Single (never married)	7	43.75
Married	6	37.5
Domestic Partnership	2	12.5
Divorced	1	6.25
Ethnicity		
White	00	00.0
Hispanic or Latino	5	31.25
Black or African American	11	68.75
American Indian or Alaskan Native	00	00.0
Asian	00	00.0
Native Hawaiian or Pacific Islander	00	00.0
Other	00	00.0