journal homepage: https://journal.pktm.com.my/

International Journal of Technical Vocational and Engineering Technology

e-ISSN2710-7094, Vol. 6, No. 2 (2025)

The Effectiveness of an Augmented Reality (AR) Training Module: A Kirkpatrick-Based Framework from Development to Deployment

Nur Dalila Abdullah[1*], Nurul 'Afifah Razali[1], Muhammad Shukran Iman Shafiee[1], Denni Kurniawan[2]

[1]Department of Information Technology, Politeknik Muadzam Shah, Muadzam Shah, 26700 Malaysia

[2]Faculty of Engineering, Universiti Teknologi Brunei, Jalan Tungku Link Gadong, BE1410 Brunei Darussalam

ıding Author email: dalila@pms.edu.my

ARTICLE INFO

Article History:
Received 7 July 2025
Revised 15 September 2025
Accepted 19 October 2025
Published 30 October 2025

Keywords:
Augmented Reality;
AR Training Module;
Kirkpatrick Model;
Educational Technology.

ɛeality (AR) in education has become popular as a novel resource to boost engagement and facilitate experiential learning. Nonetheless, numerous educators have limited experience with AR development tools, which hinders their ability to integrate this technology into their teaching effectively. This study aimed to evaluate the effectiveness of a step-by-step AR Training Module, designed using Unity and Vuforia, through Kirkpatrick's Four-Level Model: Reaction, Learning, Behaviour, and Results. A structured questionnaire of 20 Likert-scale items was administered to 37 lecturers from Politeknik Muadzam Shah and selected Kolej Komuniti institutions. The reliability test was assessed using Cronbach's Alpha, which yielded values between 0.883 and 0.928, indicating high reliability. Descriptive statistics were calculated to determine each Kirkpatrick level's mean and standard deviation, revealing strong agreement on the module's effectiveness, particularly in the Reaction (M = 4.43, SD = 0.54) and Results (M = 4.08, SD = 0.71) domains. The Learning level also showed positive outcomes (M = 3.94, SD = 0.57), while the Behaviour level indicated moderate agreement (M= 3.70, SD = 0.63) regarding the application of acquired skills. These findings support the AR module's ability to deliver engaging, relevant content with measurable impact across different dimensions of learning effectiveness. Independent sample t-tests were performed to evaluate perceptions between institutions, revealing no statistically significant differences, indicating that the module is similarly practical across various educational environments. All statistical evaluations were performed utilising IBM SPSS Statistics version 31.0. The findings demonstrate that the AR module successfully assists beginner users in gaining essential skills in AR development and is positively regarded in diverse institutional settings. The lack of notable differences between institutions indicates that the module is flexible and can be expanded for broader application. This research contributes to the expanding literature on AR integration within education and presents a validated framework for assessing comparable digital training initiatives moving forward.

1.0 Introduction

Augmented Reality (AR) is an innovative interactive technology enabling users to engage with digital content by superimposing digital objects onto physical environments in real-time through a device, such as a smartphone or tablet (Supriyanto et al., 2023). Technology is becoming increasingly popular across various sectors, such as education, healthcare, retail, entertainment, architecture, and manufacturing. In an educational environment, AR provides unmatched possibilities to transform the learning process, bring concepts to life in three dimensions, and ensure interaction. The fact that it changes how the interaction with information takes place, making it dynamic, informative, and intuitive, speaks highly of its great worth.

The efficiency of Augmented Reality in supporting the learning process is closely connected to the quality of the content designed and its practical implementation in daily practice. Short and well-designed AR experience not only keeps the users but also generates satisfaction and enjoyment. According to modern studies (Prasetya et al. 2024; Cheng et al. 2025), AR always increases scholastic achievement, fosters greater motivation, and generates favourable dispositions towards studying, because it provides immersive, interactive and contextual learning experiences that give rise to a better understanding of complex concepts. Therefore, content quality and user experience are the key motivators of engagement and overall satisfaction.

An educational AR module was carefully designed to address this potentially transformational effect and the need for this practical skill development. Intended to take the learner through creating an AR application. The module starts with essential knowledge, progresses to the current industry development tools, like Unity, and up to the specifics needed in creating and deploying a viable Android Package Kit (APK) file. The module is intentionally designed to be understandable and beneficial to people with different experience levels and provides them with the necessary skills to develop AR.

To determine the impact of the module rigorously, the Kirkpatrick model (Donald L. Kirkpatrick, 1998) of training evaluation was used to measure the effectiveness of training, which breaks down the effectiveness into four progressive components: Reaction, Learning, Behaviour, and Results. Such an approach allows for a complete, evidence-based picture of the module's effectiveness by exploring how the participants can acquire new skills, their ability to make use of the skills in a real-life setting, and the ability to demonstrate objective results. The purpose of this research is also to determine whether the module enables people from different backgrounds to learn and follow the steps easily, ensuring that it is accessible and beneficial to a wide range of learners. The information gathered through this method will be crucial for systematically assessing the effectiveness of AR in skill development and addressing knowledge gaps regarding AR's influence on skill learning.

2.0 Kirkpatrick Model

The Kirkpatrick model remains a widely embraced and utilised framework for evaluating training initiatives. Alsalamah and Callinan (2022) stated that the Kirkpatrick model continues to be a prevalent assessment tool across diverse organisations and industries. The research further emphasises the model's utility in computer science, business, medical education, and the social sciences, demonstrating its ongoing relevance for assessing training effectiveness. This robust methodology is well-suited for evaluating the performance of AR elements in workplace training scenarios. Through continuous assessment of the training program's outcomes, specifically Reactions, Learning, Behaviour, and Results, the model effectively captures the comprehensive, long-term effects of training on employee proficiencies and capabilities. When combined with an AR module, organisations acquire vital insights into their program's effectiveness.

2.1 Reaction

A research study done by Nawaz et al. (2022) primarily explores participants' initial reactions to the AR Training Module. It examines their perspective of the information's overall presentation, layout, and design. The participants shared their opinions about the module's organisation, visual attractiveness, and ease of navigation. They also considered how well the presentation style and pace fit their preferred learning methods. This phase also examined the efficiency of these textual and visual aids in keeping participants' attention and improving their understanding of the topics. The module includes 29 modules of detailed work, along with clear illustration graphics that are demonstrated in each module. The feedback is essential because it helps to build a foundation for evaluating more in-depth areas of learning and offers valuable insight into the participant's experiences.

2.2 Learning

The second level assesses the knowledge and skills acquired (Cheung et al., 2023) through the AR module. This includes understanding how to install Unity, set up AR Projects, develop interactive UI elements, such as adding an image or button, setting up Vuforia databases, incorporating 3D models and videos, and ultimately, transforming the project into a fully functional AR mobile app. The assessment focused on the participant's recall of concepts, ability to adhere to procedures, and understanding of essential AR development principles. Their self-evaluation of learning results aids in determining if the module has met its educational objectives and fostered a meaningful learning experience for participants.

2.3 Behaviour

As the analysis proceeds to this level, teaching performance is no longer about what participants know but about how they use what they know. The point is not about behaviour changes but measuring to what extent participants can apply the AR development process within a real or mock-up environment. This encompasses being able to reproduce the project setup steps, debug errors, and develop actual functioning AR applications that go beyond the context of the training. The focus in the module was on encouraging practical work and building up learning; in other words, the goal was for learners not only to remember what they learned but also to gain the confidence and ability to do it. This review also determines whether the training worked in practice and could be implemented.

2.4 Results

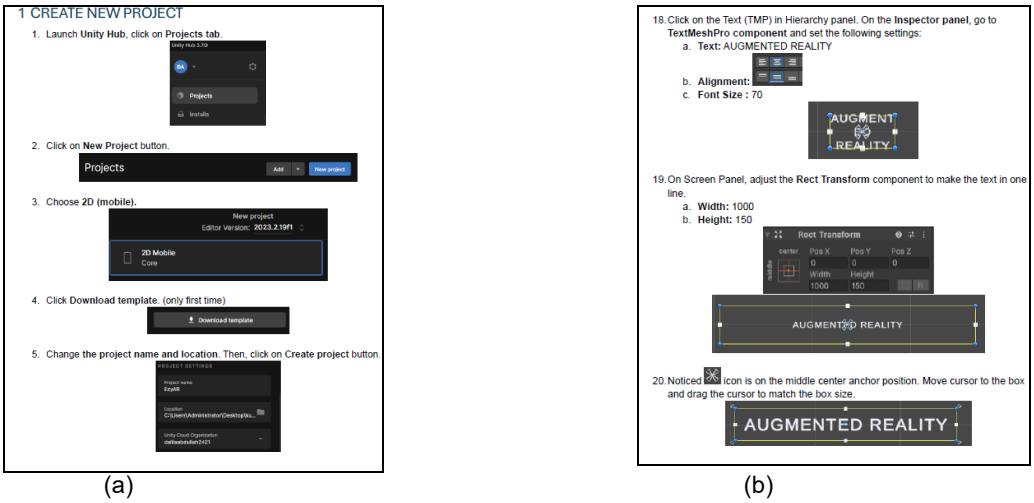
The final level evaluates the AR training module's broader and long-term effect in applying the skills and knowledge for personal and organisational development. It measures explicitly whether the module leads to successfully creating a functioning APK, which has real-world significance. From an institutional perspective, the outcomes could lead to a better curriculum, more informed use of new technologies, and a greater willingness to include new digital tools in teaching and learning. In this phase, the module's total worth and performance is assessed to see if it yields valuable outcomes justifying its continued utilisation.

3.0 Augmented Reality Training Module

The AR Training Module serves users systematically to facilitate the creation of Vuforia and Unity-based augmented reality applications. With the AR Module, users can start by downloading Unity and starting an AR project in Unity Hub. Then, they can iteratively work towards more complex tasks like building interactive homepages with image and button menus and other widgets. It also contains the AR components like setting up the Vuforia database, adding 3D models, and video integration into the AR space, equipping users with the technological skills necessary to actualise their augmented reality concepts.

The AR Training Module contains 29 illustrated modules, and the training materials are intended to help all learners, particularly novices with little to no experience taking part in AR

development, to be able to participate. All modules have clear, step-by-step instructions and relevant visuals to help understand and enable effective independent pacing. The participants can effectively convert their Unity project into a complete AR mobile application (APK) ready for deployment by the end of the training module. This module provides users with technical skills and cultivates their creativity and innovation through the hands-on use of AR technology. Figure 1 (a) shows the creation of a new project module, and Figure 1 (b) shows the text element settings in the interface.



(a) (b)

Fig. 1. (a) Create new project module, (b) Text element in the interface.

4.0 Evaluation of Training Module Effectiveness

A total of 37 participants completed the Augmented Reality (AR) training module evaluation questionnaire. These respondents were primarily from Politeknik Muadzam Shah, representing various departments including the Department of Information and Communication Technology, Department of Commerce, Department of Mathematics and Science, and the Department of Tourism and Hospitality. Regarding age distribution, the majority fell within the 35–44 and 45–54 age ranges, indicating participation from experienced academic staff. About familiarity with AR technology, 24 respondents reported being "Not familiar" and 13 respondents reported as "Somewhat familiar", and a majority of 30 respondents also indicated no prior experience, and 7 respondents had little experience in building AR applications before engaging with this training. This highlights that the module was evaluated primarily by novice to intermediate-level users, making it appropriate for assessing the effectiveness of a foundational AR learning module.

This study had a total population of 43 participants, and valid responses from 37 participants were obtained. In this case, the Krejcie and Morgan (1970) sample size determination table was used to ascertain if the sample size was appropriate. They recommend that for a population of 45, the required sample size for a confidence level of 95% and margin of error of 5% is approximately 40 respondents. The 37 responses represent nearly 86% of the population, sufficient for pilot testing and thorough data analysis. Recent studies continue to recognise the relevance and effectiveness of the Krejcie and Morgan method in research. Ahmad and Halim (2017), Chanuan Uakarn et al. (2021) and Azami et al. (2023) have utilised this technique, demonstrating its applicability in determining appropriate sample sizes. The mentioned studies highlight that, with small, well-defined populations, achieving over 80% as was the case in this study is more than adequate for ensuring generalizability and internally consistent results.

To evaluate instrument reliability using Cronbach's Alpha, a minimum of 30 participants is sufficient (Bujang et al.,2018). The current study's sample of 37 participants meets and exceeds this requirement, validating the dataset's eligibility for testing internal consistency and conducting thorough data analysis. Furthermore, it confirms that no changes were made to the

instrument or processes after the pilot phase. This approach aligns with the concept of an internal pilot design, as suggested by Bond et al. (2023) and Avery et al. (2017). Pilot data can be included in the final analysis if the study protocol remains unchanged. All quantitative analyses, including reliability, descriptive, and inferential statistics, were conducted using IBM SPSS Statistics version 31.0, which is commonly used in educational and social science research.

Lindner and Lindner (2024) five-point interpretation scale is used to assist in interpreting the mean scores for this research. There are five descriptive categories for the mean scores, which are 1.00–1.50 = Strongly Disagree, 1.51–2.50 = Disagree, 2.51–3.50 = Neutral, 3.51–4.50 = Agree, and 4.51–5.00 = Strongly Agree. This interpretation framework is used throughout the analysis of all four levels of the Kirkpatrick model to evaluate participants' perceptions of the AR training module. It offers a stable base for evaluating and debating the success of each level, allowing for accurate, evidence-driven conclusions about learner satisfaction, acquired knowledge, behavioural application, and attainment of results.

4.1 Reliability Test: Cronbach's Alpha

Cronbach's Alpha was used to measure the internal consistency of the survey used in this study as part of a reliability assessment. The tool assessed the AR Training Module's efficacy at the four Kirkpatrick model levels of behaviour, learning, reaction, and results. Five items are evaluated on a 5-point Likert scale for each level. The analysis results are shown in Table 1, which shows that all four sections received exceptional reliability values. Cronbach's Alpha for the Reaction segment was 0.920, Learning was 0.906, Behaviour was 0.883, and Results was 0.928. The instrument showed an extremely high internal consistency with a total Cronbach's Alpha of 0.947 for the four levels. These values exceed the commonly accepted minimum threshold of 0.70, as suggested by Nunnally (1978) and are well-aligned with more recent guidance by Bujang et al. (2018), who noted that a Cronbach's Alpha value above 0.80 is highly reliable in educational research involving Likert-scale instruments.

Table 1: Reliability coefficient of the items

Level	No. of Items	Cronbach's Alpha (α)
Level 1 – Reaction	5	0.920
Level 2 – Learning	5	0.906
Level 3 – Behaviour	5	0.883
Level 4 – Results	5	0.928

The high internal consistency across all four sections indicates that the items within each construct reliably measure the same underlying concept. In other words, participants responded to the items in a manner that reflects strong coherence and alignment within each evaluation level. This finding further validates the use of the instrument in assessing the AR module's effectiveness among technical and vocational educators. Since no changes were made to the questionnaire during the data collection, and the pilot sample size exceeded 30 respondents, the reliability results are considered statistically sound and methodologically appropriate.

4.2 Level 1 – Reaction: Participants' Satisfaction and Engagement

Table 2 shows the mean and standard deviation for Level 1 – Reaction, with a mean score of 4.43 and a standard deviation (SD) of 0.54. This suggests that participants reacted positively and consistently to the AR training module. Based on the evaluation scale established by Lindner and Lindner (2024), a mean score within the range of 3.51–4.50 is categorised as "Agree," indicating that learners generally concurred that the module was practical in terms of its structure, visuals, and delivery. While this score is close to the upper limit of the "Agree" category and only slightly below the "Strongly Agree" threshold (≥ 4.51), it nevertheless represents a very positive response, according to findings from consensus-based research, such as Delphi studies (von der Gracht, 2012). A Coefficient of Variation (CV) below 0.5 is typically considered to reflect acceptable internal agreement among respondents. While the CV in this case is slightly above

that threshold, it still indicates a reasonable level of consensus. Additionally, Balalle (2024) emphasised that positive learner reactions are important in technology-based learning, as they help increase engagement and motivation to explore the topic further. Therefore, the reaction demonstrates that the AR module was engaging and well- received by learners across different institutions.

Table 2: Mean and Standard Deviation for Level 1 – Reaction

No	Item	Mean	Standard Deviation
1	The module instructions were clear and easy to understand.	4.27	0.65
2	I enjoyed using the step-by-step guide to build the AR app.	4.27	0.65
3	The visual aids (screenshots, videos) helped me follow the process better.	4.49	0.61
4	I would recommend this module to others learning AR.	4.54	0.61
5	The learning experience was engaging and motivating.	4.57	0.56
	Overall Mean Score and Standard Deviation	4.43	0.54

4.3 Level 2 – Learning: Knowledge and Skill Acquisition

As shown in Table 3, the mean and standard deviation for Level 2, the analysis yielded a mean score of 3.94 with a standard deviation of 0.57, indicating that participants perceived the AR training module as effective in enhancing their knowledge and skills. Based on recent guidelines for interpreting 5-point Likert scales by Lindner and Lindner (2024), this score falls within the "Agree" range (3.51–4.50), suggesting that most respondents agreed they had acquired new learning from the module. The moderate standard deviation indicates relatively consistent responses, reflecting participants shared and uniform learning experience. As originally emphasised by Donald L. Kirkpatrick (1998), learning is a core dimension of evaluating training effectiveness. More recent insights by Simsek and Direkçi 2023) and AlGerafi et al. (2023) further support the notion that AR-based learning can significantly improve learning outcomes, enhance comprehension, and foster positive student perceptions. Their findings of positive learning impacts align with the high mean score and low variability observed in this study. This suggests that the AR training module successfully conveyed essential technical knowledge, particularly in Unity installation, Vuforia integration, and AR interface development.

Table 3: Mean and Standard Deviation for Level 2 - Learning

No	Item	Mean	Standard Deviation
1	I now understand the basic steps to develop an AR app using Unity.	4.08	0.65
2	I can confidently navigate the Unity interface after using the module.	3.84	0.65
3	I understand how to integrate AR SDK (e.g., Vuforia) in Unity.	3.81	0.61
4	I know how to configure Unity settings to build an APK.	3.81	0.61
5	The module improved my overall knowledge of AR development.	4.14	0.56
	Overall Mean Score and Standard Deviation	3.94	0.57

4.4 Level 3 – Behaviour: Application of Skills in Practice

Table 4 demonstrates the mean and standard deviation for Level 3 (Behaviour), the mean score is 3.70 with a standard deviation of 0.63, indicating that participants generally agreed to some extent that they had begun applying the skills acquired through the AR training module. The slightly wider standard deviation indicates variability in participants' ability or opportunity to transfer knowledge into practice. Based on the interpretation scale, this score falls within the "Agree" range (3.51–4.50), suggesting a positive but moderate level of behavioural transfer. The slightly higher standard deviation reflects some variability among participants, likely due to differences in individual readiness, opportunity, or environment for applying the skills. Behavioural change following training is rarely immediate, heavily relying on organisational support, individual motivation, and access to tools. Studies consistently show that factors like supervisor support (Mohamad Yunus et al., 2022), peer feedback, application opportunities, and incentives (Gautam and Basnet, 2021) are critical for successful training transfer. Indeed, work environment factors significantly influence post-training behaviour (Braun et al. 2019),

underscoring the need for supportive systems and a positive organisational culture (Ha and Vanaphuti, 2022). In this study, the moderate Behaviour score indicates many participants are applying AR skills (Unity, Vuforia), but others still need more hands-on practice and institutional support for full integration.

Table 4: Mean and Standard Deviation for Level 3 - Behaviour

No	Item	Mean	Standard Deviation
1	I was able to follow the steps independently without much help.	3.60	0.65
2	I can now develop a simple AR app on my own.	3.57	0.65
3	I applied the knowledge gained to troubleshoot errors during the build.	3.60	0.61
4	I feel confident in using Unity and AR tools in future projects.	3.70	0.61
5	I would use this process again in my next AR development task.	4.03	0.56
	Overall Mean Score and Standard Deviation	3.70	0.63

4.5 Level 4 – Results: Achievement of Learning Objectives

Table 5 shows the mean and standard deviation for Level 4 (Results), and the mean score is 4.08, with a standard deviation of 0.71, indicating that participants generally perceived the AR training module as beneficial in achieving broader outcomes. These outcomes include improved readiness to integrate AR into teaching practices, enhanced professional confidence, and overall educational value. Based on the interpretation scale, the mean falls within the "Agree" range (3.51–4.50), suggesting that most participants viewed the training as relevant and practical. While the slightly higher SD reflects some response variation, the overall trend remains positive. In alignment with the Kirkpatrick model's Level 4, which focuses on training impact in terms of institutional or personal goals, this finding implies that the module delivered meaningful long-term benefits. This is further supported by (AlGerafi et al., 2023), who observed that when participants report mean scores above 4.0 in digital learning contexts, it generally indicates that the training aligns well with real-world educational needs and expectations.

Table 5: Mean and Standard Deviation for Level 4 - Results

No	Item	Mean	Standard Deviation
1	I successfully built an APK file using Unity.	3.92	0.92
2	The APK worked as expected when tested on my device.	3.95	0.94
3	The AR content (e.g., marker tracking, 3D objects) functioned correctly.	4.10	0.84
4	I achieved the learning objectives stated at the beginning of the module.	4.21	0.63
5	This module helped me complete a working AR prototype from start to finish.	4.21	0.67
	Overall Mean Score and Standard Deviation	4.08	0.71

4.6 Independent Samples T-Test

An independent samples t-test compares the mean scores of two distinct groups of respondents from Politeknik Muadzam Shah (PMS) and the combined Kolej Komuniti (KK) institutions, including Raub, Klang and Bera. This study sought to identify whether there were any statistically significant differences in the effectiveness of the AR training module. The assessment was carried out according to the four tiers of the Kirkpatrick framework: Reaction, Learning, Behaviour, and Results. This t-test was suitable for this research because the groups (PMS vs. KK) were independent and did not overlap. William Sealy Gosset initially developed this method under the alias "Student" (Student, 1908), and it continues to be a fundamental resource in educational research for assessing means among different groups. Arman et al. (2022) stated this statistical method was employed to compare two groups and effectively identified significant differences. Thus, the independent t-test offers a reliable and validated approach to assess the perceived effectiveness of the AR training module among various institutions.

Before performing the independent samples t-test, preliminary checks were conducted to confirm fundamental assumptions. The Shapiro-Wilk test for normality indicated that the data for both Kolej Komuniti ($p = 0.951$) and Politeknik ($p = 0.694$) were normally distributed, with both

p-values exceeding 0.05. This confirms that the assumption of normality was satisfied for each group. Furthermore, Levene's Test for Equality of Variances yielded a non-significant outcome (p = 0.701), suggesting that the variances between the two institutions were statistically comparable. This meets the requirement for homogeneity of variance. As both assumptions were satisfied, the dataset was deemed suitable for further analysis using an independent samples t-test.

Table 6 presents the group comparison of Kirkpatrick Levels through an independent samples t-test. The results revealed no significant differences between the two institutional groups at the four levels. For the Reaction level, the mean score for PMS was 4.53, compared to 4.34 for KK, with a t-value of 1.074 and a p- value of 0.29. For the Learning level, PMS scored a mean of 4.01 while KK scored 3.87 with a t-value of 0.745 and a p-value of 0.46. Similarly, the Behaviour level showed a PMS mean of 3.80 and a KK mean of 3.61 (t = 0.92, p = 0.37), and for the Results level, the mean scores were 3.94 for PMS and 4.20 for KK (t = –1.10, p = 0.28). In each case, the p-values exceeded the commonly accepted significance threshold of 0.05, indicating no statistically significant differences between the two groups.

Table 6: Group comparison of an independent sample t-test

No Kirkpatrick Level	PMS Mean	KK Mean	t-value	p-value
Level 1 – Reaction	4.53	4.34	1.074	0.29
Level 2 – Learning	4.01	3.87	0.745	0.46
Level 3 – Behaviour	3.80	3.61	0.92	0.37
Level 4 - Results	3.94	4.20	- 1.10	0.28

The findings indicate that both PMS and KK participants had similar perceptions of the AR module regarding satisfaction, learning outcomes, behavioural application, and perceived results. This positive outcome suggests that the training module is effective and reliable across different educational contexts. Such consistency is crucial in educational technology, ensuring fair access and usability for all learners. When no significant differences are found among groups, it may imply that the educational resources are universally accessible and well-designed, catering to the needs of various learners. This notion is supported by Gardner et al. (2023), who discovered that educational models performed consistently across multiple institutions, highlighting the adaptability and generalizability of well-designed interventions.

5.0 Conclusion

Findings from this study indicate that the AR Training Module successfully imparted necessary AR development skills to instructors, even those with minimal prior exposure. The instrument's high internal consistency across all four levels of the Kirkpatrick framework confirms its reliability. Average scores also point to a favourable reception of the module's design, content, and real-world applicability. The consistent effectiveness of the module across varied educational settings is underscored by the absence of significant outcome disparities between participants from PMS and different KK institutions. These results are consistent with previous academic work, Simsek and Direkçi (2023), which found that clear instructional design can foster consistent learning experiences across diverse institutions, and Balalle (2024), that positive feedback from learners is crucial for sustaining motivation and engagement.

Even though the results remained optimistic, the study notes some aspects requiring additional focus. The average scores alongside considerable variation within the Behaviour level indicate that participants may possess the knowledge. However, a few of them may not have the opportunity or confidence to utilise the skills independently. Further research could explore the remote retention and actual application of AR development skills, perhaps through longitudinal studies or subsequent project implementations. In addition, expanding the scope of the participants to include teachers from polytechnics, colleges, private universities, and other higher

learning institutions could help assess the overarching academic relevance of the module beyond its primary discipline. Combining self-reported data with observational or other forms of criterion-referenced evaluation as pointed out by Donald L. Kirkpatrick (1998) and Avery et al. (2017) could strengthen credibility and accuracy. In summary, the AR module exhibits remarkable promise as an agile and effective tool for equipping educators with the requisite digital competencies in Industry 4.0 education.

Acknowledgements

I would like to extend my sincerest appreciation to the Department of Information Technology and Communication (JTMK), together with the administration of Politeknik Muadzam Shah, for their support and willingness to provide the facilities for this research. Special thanks are given to the lecturers from Politeknik Muadzam Shah and the participating Kolej Komuniti, namely Raub, Klang and Bera for their participation in this research and critical input on the AR Training Module. At this moment, I would also like to acknowledge all other persons who contributed in one way or another towards completing this research.

Author Contributions

Abdullah N.D.: Conceptualization, Methodology, Data Analysis**; Razali N.A.:** Validation, Supervision**;** **Shafiee M.S.I.:** Writing-Reviewing and Editing: **Kurniawan D.:** Validation**.**

Conflicts of Interest

The manuscript has not been published elsewhere and is not being considered by other journals. All authors have approved the review, agree with its Submission, and declare no conflict of interest in the manuscript.

6.0 References

Ahmad, H., & Halim, H. (2017). Determining Sample Size for Research Activities: The Case of Organizational Research. In Business Review (Vol. 2, Issue 1).

AlGerafi, M. A. M., Zhou, Y., Oubibi, M., & Wijaya, T. T. (2023). Unlocking the Potential: A Comprehensive Evaluation of Augmented Reality and Virtual Reality in Education. Electronics, 12(18), 3953.

Alsalamah, A., & Callinan, C. (2022). The Kirkpatrick model for training evaluation: bibliometric analysis after 60 years (1959–2020). Industrial and Commercial Training, 54(1), 36–63. https://doi.org/10.1108/ICT-12-2020-0115

Arman, A., Rianto, S., & Yuherman, Y. (2022). Perbandingan Aktivitas Belajar Geografi Antara Siswa Kelas Xi Kelompok Peminatan Dengan Lintas Minat Di Sma Negeri 1 Tigo Nagari. Jambura Geo Education Journal, 3(1), 1–10.

Avery, K. N. L., Williamson, P. R., Gamble, C., Francischetto, E. O. C., Metcalfe, C., Davidson, P., Williams, H., Blazeby, J. M., Blencowe, N., Bugge, C., Campbell, M., Collinson, M., Cooper, C., Darbyshire, J., Dimairo, M., Doré, C., Eldridge, S., Farrin, A., Foster, N., Treweek, S. (2017). Informing efficient randomised controlled trials: Exploration of challenges in developing progression criteria for internal pilot studies. BMJ Open, 7(2).

Azami, N., Azmee, D., Keng, N., Qing, N., Shen, N., & Han, L. (2023). Factors Affecting Employee Retention Among SMES' Millennial Employees in Klang Valley, Malaysia. International Journal of Academic Research in Business and Social Sciences, 13.

Balalle, H. (2024). Exploring student engagement in technology-based education in relation to gamification, online/distance learning, and other factors: A systematic literature review. In Social Sciences and Humanities Open (Vol. 9). Elsevier Ltd.

Bond, C., Lancaster, G. A., Campbell, M., Chan, C., Eddy, S., Hopewell, S., Mellor, K., Thabane, L., & Eldridge, S. (2023). Pilot and feasibility studies: extending the conceptual framework. Pilot and Feasibility Studies, 9(1).

Braun, M. T., Converse, P. D., & Oswald, F. L. (2019). The accuracy of dominance analysis as a metric to assess relative importance: The joint impact of sampling error variance and measurement unreliability. Journal of Applied Psychology, 104(4), 593–602.

Bujang, M. A., Omar, E. D., & Baharum, N. A. (2018). A review on sample size determination for cronbach's alpha test: A simple guide for researchers. Malaysian Journal of Medical Sciences, 25(6), 85–99.

Chanuan Uakarn, Nittaya Sintao, & Kajohnsak Chaokromthong. (2021). Sample Size Estimation using Yamane and Cochran and Krejcie and Morgan and Green Formulas and Cohen Statistical Power Analysis by G*Power and Comparisons.

Cheung, V. K. L., Chia, N. H., So, S. S., Ng, G. W. Y., & So, E. H. K. (2023). Expanding scope of Kirkpatrick model from training effectiveness review to evidence-informed prioritization management for cricothyroidotomy simulation. Heliyon, 9(8).

Cheng, C. Y., Quah, W. B., Kong, S. L., Loo, K. Z., Tan, Y. X., Koay, J. M., & Asmak Ismail, N. H. (2025). Do You Prefer Virtual Tourism (VT)? The Use of Minecraft in Developing Titania Virtual Theme Park. International Journal Of Technical Vocational And Engineering Technology, 6(1), 93-100.

Donald L. Kirkpatrick. (1998). Another Look at Evaluating Training Programs: Fifty Articles from Training & Development and Technical Training: Magazines Cover the Essentials of Evaluation and Return-on- investment (Donald L. Kirkpatrick, Ed.). American Society for Training & Development.

Gardner, J., Yu, R., Nguyen, Q., Brooks, C., & Kizilcec, R. (2023). Cross-Institutional Transfer Learning for Educational Models: Implications for Model Performance, Fairness, and Equity. 2023 ACM Conference on Fairness Accountability and Transparency, 1664–1684.

Gautam, D. K., & Basnet, D. (2021). Organizational culture for training transfer: the mediating role of motivation. International Journal of Organizational Analysis, 29(3), 769–787.

Ha, H., & Vanaphuti, V. (2022). Training transfer to the workplace: a study of corporate English language training in hospitals. Journal of Workplace Learning, 34(1), 110–131.

Krejcie, R. V, & Morgan, D. W. (1970). Determining Sample Size for Research Activities. In Educational And Psychological Measurement (Vol. 30).

Lindner, J. R., & Lindner, N. (2024). Interpreting Likert type, summated, unidimensional, and attitudinal scales: I neither agree nor disagree, Likert or not. Advancements in Agricultural Development, 5(2), 152–163.

Mohamad Yunus, ain, Sharuddin, N., & Khalid Mohd Abas, M. (2022). The Impact of Perceived Organisational Support, Supervisor Support and Self-Efficacy on Transfer of Training among Public Service Officers. In Journal of International Business, Economics and Entrepreneurship (Vol. 7, Issue 2).

Nawaz, F., Ahmed, W., & Khushnood, M. (2022). Kirkpatrick Model and Training Effectiveness: A Meta - Analysis 1982 To 2021. Business & Economic Review, 14(2), 35–56.

Nunnally, J. C. (1978). Psychometric Theory (2nd Edition). McGraw-Hill.

Prasetya, F., Fortuna, A., Samala, A. D., Rawas, S., Mystakidis, S., Syahril, Waskito, Primawati, Wulansari, R. E., & Kassymova, G. K. (2024). The impact of augmented reality learning experiences based on the motivational design model: A meta-analysis. Social Sciences & Humanities Open, 10, 100926.

Simsek, B., & Direkçi, B. (2023). The Effects of Augmented Reality Storybooks on Student's Reading Comprehension. British Journal of Educational Technology, 54(3), 754–772.

Student. (1908). The Probable Error of a Mean. Biometrika, 6(1), 1–25.

Supriyanto, S., Joshua, Q., Abdullah, A. G., Tettehfio, E. O., & Ramdani, S. D. (2023). Application of Augmented Reality (AR) in vocational education: A systematic literature review. Jurnal Pendidikan Vokasi, 13(2), 205–213.

Von der Gracht, H. (2012). Consensus measurement in Delphi studies Review and implications for future quality assurance. Technological Forecasting and Social Change, 79, 1525–1536.