DOI: https://doi.org/10.63056/jllsa.2.11.2026.465

Journal of Language, Literature & Social Affairs

journal homepage: https://scholarclub.org/index.php/jllsa

Evaluating the Validity of LLM-Generated Synthetic Respondents in Representing Human Response Patterns

Kainat Khalid*[a], Dr. Bushra Yasin[b], Dr. Abira Ismat Butt[c]

a. Lecturer in English, PhD Scholar in Education, University of Home Economics, Lahore.
b. Assistant Professor, Department of Education, University of Jhang, Jhang.
c. Assistant Professor, Department of Education, Kinnaird College for Women.

ARTICLE INFO

Received:
October 21, 2026
Revision Received:
November 03, 2026
Accepted:
November 15, 2026
Available Online:
November 24, 2026

Keywords:
Artificial Intelligence, Large Language Models, Psychometric Validation, Response Patterns, Synthetic Respondents, Survey Research

ABSTRACT

With the surge of the Large Language Models (LLMs), new applications for the creation of synthetic respondents have emerged in research settings where gathering human data can be expensive, time intensive or otherwise challenging. This study used established survey instruments to evaluate the validity of LLM-generated synthetic respondents in representing human response patterns across four themes. Data were collected from university students using a structured questionnaire. The same questionnaire items were presented to ChatGPT (GPT-4o), Google Gemini, and Anthropic Claude to compare human and AI-generated responses across the four themes. The analysis included the following characteristics: central tendency, distribution of scores, variability of responses, inter item relationships, and latent construct representations. Comparisons and statistical analyses were made to understand how close the synthetic responses were at both obtrusive and hidden levels to the human response. The results showed that the LLM-generated responses tended to be internally coherent and self-consistent, but there were some significant differences between the LLM results and human data. The differences in score distributions, lower amount of variability and different patterns of construct representation were observed in synthetic responses. Some of the aggregate trends were similar to those seen in humans, while others pointed to differences in LLM response patterns, raising questions about how well these models can capture the diversity and complexity of people. The results also demonstrated that there was variation in the level of agreement between the human and synthetic data on each construct, emphasizing the need for multiple analyses of data for multiple levels of validity, and not just a superficial empirical reassonance. In general, the observations of the study showed empirical evidence of strengths and limitations of LLM generated synthetic respondents and added to the larger discussions of the viability of using LLM in social and behavioral science research. The results indicated that synthetic respondents could provide valuable additional data for exploratory investigations and cannot be treated blindly as alternatives to human respondents if proper validation is not applied.

* Correspondence to: Lecturer in English, PhD Scholar in Education, University of Home Economics, Lahore.
E-*mail address:* kainatkhld@gmail.com (K. Khalid)

3105-4919 & 3105-4927/© 2026 The Authors. Published by Scholar Club. This is an open access article.

1. Introduction

Artificial Intelligence (AI) has rapidly revolutionized how information is created, ingested and used in various fields. Technological developments have greatly impacted education, research, health care, government and, importantly, communication, making for new ways to work more efficiently and to make better use of information. In the modern world, where digital systems are omnipresent, the necessity of providing smart systems to assist human decision making has been growing in importance . People and societies are becoming more dependent on the capabilities of digital technology, the need for intelligent systems that will facilitate human decision making has grown in importance. The advent of sophisticated language-based technologies rooted in recent advances in machine learning and AI that generate human-like communication and give rise to complex textual output were born . This progress has inspired curiosity in the research and data collection space as they strive to understand how to put AI-driven answers to use.

The use of technology in education and social systems has proven to be a revolution towards the implementation of digital innovations. The use of a technology-supported environment has been proven to increase engagement, learning outcomes, accessibility, and the knowledge creation processes . Further, the adoption of innovative leadership practices and technology-based solutions has been linked to better education experiences and becoming better at their job. This evolution is a positive sign of acceptance for intelligent technologies in traditional human-to-human operation, decision-making and problem-solving environments.

Concurrently, the increased dependence of digital systems has brought opportunities for the large-scale collection and analysis of data. Smart systems are based on the concept of linking various technologies together, producing, processing and utilizing their information in a most efficient way to assist decision making process (Khan & Younas, 2021). Furthermore, modern day AI methods can generate material that is just like human conversation styles. This has led to a range of research that's focused on the prospect of using AI-generated responses on surveys, simulations and behavioral research .

Evidence from educational research is extensive about the need for understanding nature of human perceptions, attitudes and behaviours via the investigation of those things empirically). The studies that employ human beings as their data have produced valuable and meaningful results for studying teaching quality, student motivation, parent participation, leadership development, SRL, and classroom effectiveness (Younas et al., 2023). As more and more answers are being produced by AI, a critical methodological worry about whether these answers will be sure authorities from traditional respondents is becoming an increasing factor.

Smart learning ecosystems continue to grow in significance, complementing the increase in the importance of technology in the creation of knowledge . Experiments integrated in universities have pinpointed important achievements in the implementation of SLP and digital learning methods, and new challenges in terms of being authentic, engaging, and integrating technology . Likewise, studies exploring the online learning readiness have established that the readiness of the user to learn online through technology has a significant effect on engagement and efficiency in learning. The results indicate that digital technologies are deeply entrenched in today's social and educational landscape, and that there is a growing shift toward measuring and assessing the outputs of AI.

Research examining educational interventions has repeatedly emphasized the role of technology in shaping user experiences . For example, studies involving Kahoot!, Symbolab calculators, and smart education frameworks have reported positive outcomes regarding student engagement, conceptual understanding, and learning effectiveness . Such evidence demonstrates that digital tools can effectively support educational processes. However, while these technologies facilitate learning and information access, the capacity of AI systems to accurately replicate human response patterns requires systematic validation.

The emergence of AI-generated content has also introduced ethical considerations. emphasized the importance of ethical governance in AI-supported educational environments, highlighting the necessity of ensuring transparency, accountability, and responsible implementation . Similar concerns become particularly important when AI-generated respondents are utilized in research settings. If synthetic respondents are unable to accurately represent human perspectives, significant concerns regarding validity, reliability, and research quality may arise.

Moreover, studies focusing on quality assurance and educational effectiveness have consistently stressed the importance of valid data in decision-making and policy development (Khan, Taj, Younas., 2024). Since research conclusions influence educational practices, technological adoption, and institutional strategies, ensuring the credibility of collected data remains essential. Consequently, evaluating whether AI-generated respondents can produce data comparable with human responses has become an increasingly important area of investigation.

The concept of ubiquitous learning further illustrates the interaction between humans and digital technologies in contemporary society. argued that learning environments are becoming increasingly connected through digital infrastructures, creating new

forms of interaction between technology and users. Within such environments, AI technologies are no longer passive tools but are actively participating in communication and information generation processes. This shift necessitates rigorous examination of the authenticity and representativeness of AI-generated outputs .

Furthermore, studies examining educational quality, classroom processes, institutional effectiveness, and student engagement consistently demonstrate the complexity of human perceptions and experiences (Younas et al., 2022; Younas et al., 2023; Khan & Younas, 2021). Human responses are shaped by social, cultural, cognitive, and contextual factors that contribute to substantial diversity within empirical datasets. Whether AI-generated respondents can adequately reproduce these complex response structures remains a significant research concern .

Against this backdrop, the present study evaluated the validity of LLM-generated synthetic respondents in representing human response patterns (Faisal, Akmal, Makhdum, Sandhu, Younas, Salim, 2023). By comparing synthetic and human-generated responses, the research sought to determine whether AI-generated participants could accurately represent human behaviours and attitudes within survey-based investigations .

1.1. Significance

This study examines the current debate around the use of AI in research and data production (Qayyum & Hussain, 2019). The impacts of AI on educational, social, and technological settings are growing and having an effect on their surroundings, making it crucial to have an understanding of their capacity to simulate human reaction to their world. The results of the study offer insightful information about the potential benefits and drawbacks of synthetic respondents as a research method (Younas, Salim, 2023). Moreover, the study is also acquiring methodological knowledge on whether the participants recruited by the use of AI can serve as reliable substitutes or supplements of human participants. The research is also contributing to the current debates on smart education, technology innovation, ethical application of AI in education, and technology-supported decision-making .

1.2. Problem Statement

With the rise in the artificial intelligence technologies, opportunities have been provided to generate synthetic respondents who can give human-like response (Faisal, Akmal, Makhdum, Sandhu, Younas, Salim, 2023). Yet, though the availability of AI-created data is increasing rapidly, there is not a lot of evidence about how well synthetic respondents emulate the responses of real people (Qayyum, Khalil, Arif, Rafiq, & Makhdum, 2025). Past research on digital innovation, smart learning, educational technologies, and AI governance has highlighted the critical need for accurate and trusted data in making informed decisions (Khan & Younas, 2021). In this context, there is uncertainty if synthetic entities produced by LLM can deliver genuine insights into human attitudes, perceptions and behaviours. This has to be systematically investigated in regard to the validity of synthetic searches to capture human response types, as there is a need for empirical verification .

1.3. Objectives

- To evaluate the validity of LLM-generated synthetic respondents in representing human response patterns.
- To compare response characteristics between human participants and LLM-generated synthetic respondents.
- To determine the extent to which synthetic respondents can serve as reliable sources of research data.

1.4. Research Questions

- How accurately do LLM-generated synthetic respondents represent human response patterns?
- What similarities and differences exist between human participants and synthetic respondents?
- To what extent can LLM-generated respondents provide reliable data for research purposes?

1.5. Gaps

While the existing literature reflects a diverse range of studies focused on educational quality, digital learning, digital technology integration, smart education, effective classroom, educational leadership and AI governance, few studies have directly addressed the validity of AI-made synthetic respondents in the context of empirical research settings. The literature available is mainly oriented to technology adoption and to educational outcomes instead of one aspect, that is, the representation of response patterns of a human through the artificial intelligence (Makhdum, Khan, Munir, Younas & Faisal, 2022). Moreover, research exploring digital innovation has almost exclusively focused on the question of whether AI stakeholders can successfully imitate human attitudes, perceptions and behavioural reactions to digital innovation. Hence the need for research into the validity and reliability of the synthetic responses of LLM's, which this study aims at .

2. Methodology

The research design used in this study was a mixed-methods approach comprising both qualitative and quantitative research (Maria, Makhdum, Sandhu, Khan & Younas, 2025). This study used a mixed-methods research design, integrating both qualitative and quantitative aspects. This choice of a mixed-methods approach was made because the need was felt to analyze the distribution of scores and the quality of the perceptions of the advantages and disadvantages of the use of synthetic respondents . When dealing with complex phenomena, mixed methods research allows for the syntegration of quantitative and qualitative research. Mixed methods designs have been successfully applied in studies on education, technology adoption, digital technology and learning, smart educational environments, to address multifaceted research problems and to provide high credibility conclusions . The study therefore had a convergent mixed methods design where two data types were collected within one phase, analysed separately and then consolidated during interpreting (Maria, Makhdum, Sandhu, Khan & Younas, 2025).

The quantitative part concentrated on comparing the pattern of responses produced by the human respondents and the synthetic respondents produced by LLMs (Salim, Makhdum & Younas, 2022). The qualitative component focused on understanding how and why synthetic respondents were viewed as credible, representative and have limitations when used in research (Salim, Makhdum & Younas, 2022; Qayyum, 2025). There was more evidence of validation through the two methods working together than was possible with only one method of validation .

During the quantitative phase, the target group for the study was the university students, university teachers, educational researchers and post graduate scholars with prior knowledge of online surveys and digital research environment (Younas, Khan & Makhdum, 2022). The groups were chosen in order to convene due to the regular use of survey-based studies and technology-based education. These populations have consistently been shown to be important in terms of both technology readiness, and engagement with digital technologies, in higher education research contexts (Khan & Younas, 2021). The population for the qualitative phase was comprised of experts in four areas of education, namely educational technology, artificial intelligence, research methodology and higher education (Younas, Khan & Makhdum, 2022).

The quantitative aspect of the study involved a sample of 300 respondents who were selected from the human respondents of the study population (Qayyum, Aslam, Maria, Younas, & Mehmood, 2025). A sample size of this size was deemed adequate for descriptive analysis, comparative and inferential analysis (Makhdum, Sandhu, Batool, Khan, Faisal & Younas, 2023). Furthermore, 300 synthetic respondents were created by a Large Language Model with prompts based on the standard survey instrument prompt used for humans. The resulting sample sizes were essentially the same for both groups, allowing the distributions of responses across both groups and the psychometric properties to be directly compared (Makhdum, Khan, Munir, Younas, Faisal, 2022). Fifteen experts were chosen for the qualitative strand for semi-structured interviews . Experts rein on the synthetic respondents' research revealed day to day and method questions.

The quantitative part of the study involved different sampling technique, while the qualitative part of the study employed the same sampling techniques. In the case of human participants, stratified random sampling was done in order to represent all the university students, teachers, researchers and post graduate scholars. Stratification added to diversity and represented more observing different responses. Purposive sampling was used for the expert interviews, where the subjects were chosen due to their knowledge in the field of artificial intelligence, the technology of education, specialisation in conducting surveys and research methodology. In qualitative research, when participation of certain number of information-rich participants is enough to meet research goals, there is a wide consensus that the purposeful sampling would be appropriate technique.

Structured questionnaire and a semi-structured interview protocol were used for collecting data (Younas et al., 2023). This questionnaire comprised of a list of demographic items and a series of statements using the Likert Scale format that assessed attitudes, perceptions, beliefs, and judgments related to technology, digital innovation, education, teaching, and applications related to AI. Studies have firmly supported the use of structured questionnaires in the fields of the study of educational quality, student engagement, technological interventions, leadership practices, and digital learning environments (Younas et al., 2023). The items were responded on a five point Likert scale ranging from strongly disagree to strongly agree.

To elicit synthetic responses, the selected Question Answering (QA) Large Language Model (LLM) was given the same questions and prompts. The uniformity of the prompting procedure was fully realized throughout the synthetic cases which allowed for minimizing differences between cases in terms of the different words used as prompts. For each synthetic participant (respondent), the same survey instrument was used as for human participant (respondent). The automatically generated responses were then collated into a set of responses similar to the human sample. This procedure resulted in the two groups being comparable and "standard" in the measuring process.

The experts were interviewed using semi-structured interviews to gather qualitative data. The interview protocol included open-ended questions to explore perceptions on the use of synthetic respondents, strengths and weaknesses of AI-generated

responses, concerns about reliability and validity, ethical issues, and future potential uses of LLM-generated data in academic research. Semi-structured interviews were chosen due to their flexibility and the fact that they still allow some structure to the interview to capture the research objectives. Each interview was transcribed verbatim and recorded with the participant's permission for analysis.

Data collection for quantitative research took place using Internet based data collection tools (survey). The participant information sheet included information about the purpose of the research and it was requested that this questionnaire be completed voluntarily by the participants. Because the study is technology-driven, technologies are gaining in popularity for participation environments (Khan & Younas, 2021), online participation was seen as appropriate. Before any analysis was carried out, human responses and synthetic responses were stored separately.

Two of them were quantitative data analysis techniques which are descriptive and inferential statistical techniques. For comparison with human responses, the measures of central tendency (mean and standard deviation) were calculated for the synthetic responses. Similarities and differences in datasets were investigated by looking at the response distributions, patterns of variability, and inter item correlations as well as reliability coefficients. Comparative analyses were carried out in order to pinpoint the points of intersection and disparity. Finally, a factor analysis was used to explore the extent to which construct validity was replicated using respondents who were synthetic . The analysis was directed towards the representativeness of responses at several-dimensional levels and not just average response similarities.

Thematic analysis method was used to analyze the qualitative data. Transcriptions of interviews were read over and over, for familiarity with the data. First codes were created upon repetition of ideas and trends. Analogous codes were then organised into larger themes that together captured the participant's thoughts about synthetic respondents and synthetic science research data. Major concerns, expectations, advantages and limitations of synthetic respondents were identified using thematic analysis. Thematic analysis results were then compared with quantitative results to help integration and triangulation.

Reliability was established through multiple procedures. For the quantitative component, internal consistency reliability was assessed using Cronbach's alpha coefficient. Values of 0.70 and above were considered acceptable for research purposes. Reliability analysis was conducted separately for human and synthetic datasets to evaluate consistency within both groups. Additionally, standardized administration procedures were implemented to minimize measurement errors. Identical instruments, instructions, and response formats were employed to ensure consistency during data collection.

For the qualitative component, reliability was enhanced through detailed documentation of interview procedures, maintenance of an audit trail, and consistent application of coding procedures. Peer review of coding decisions further strengthened the trustworthiness of findings. Transcriptions were carefully reviewed to ensure accuracy and completeness before analysis.

Validity was addressed through several strategies. Content validity for the questionnaire was established through expert review conducted by specialists in educational research, artificial intelligence, psychometrics, and survey methodology. Experts evaluated the clarity, relevance, appropriateness, and comprehensiveness of questionnaire items in relation to the study objectives. Revisions were incorporated based on their recommendations to improve instrument quality.

Construct validity was assessed through exploratory factor analysis and examination of inter-item relationships. The analysis evaluated whether the instrument measured the intended constructs consistently across both human and synthetic datasets. The use of factor analysis provided evidence regarding the structural integrity of the measurement framework. Criterion-related validity was further examined by comparing response tendencies observed in synthetic data with corresponding patterns derived from human participants.

The validity of qualitative findings was enhanced through triangulation, member checking, and expert verification. Triangulation involved comparing quantitative and qualitative findings to identify areas of agreement and divergence. Member checking was conducted by sharing summaries of interview interpretations with selected participants to confirm the accuracy of representations. Expert verification further ensured that interpretations remained consistent with participant perspectives.

Integration of quantitative and qualitative findings constituted a critical component of the research methodology. Following separate analyses, the results from both strands were merged to develop a comprehensive understanding of the validity of LLM-generated synthetic respondents. Areas where quantitative findings aligned with qualitative perceptions were interpreted as evidence of convergence, whereas discrepancies provided additional insights into limitations and contextual considerations. This integration strengthened the overall rigor of the study and enabled a more comprehensive assessment of whether synthetic respondents adequately represented human response patterns.

The methodology implemented a convergent mixed-methods design, with systematic and robust sampling procedures, in conjunction with validated instruments, systematic data collection strategies, and multiple reliability and validity measures,

allowing for assessment of how well LLM-generated synthetic respondents matched human responses in the context of today's research landscape.

2.1. Descriptive Analysis

The study examined the reliability and representativeness of human responses as represented by LLM-generated synthetic respondents with quantitative and qualitative data. The survey instrument was used to collect quantitative data from 300 human respondents and 300 synthetic (LLM) respondents. Descriptive statistics were used to analyse the mean, range and general pattern of the responses of the two groups. Mixed method studies have highlighted the need for quantitative data, along with the opinions expressed by the participants, to fully grasp research phenomena .

Table 1: Demographic Distribution of Human Respondents (n = 300)

Variable	Category	Frequency	Percentage
Gender	Male	142	47.3%
Gender	Female	158	52.7%
Education	Undergraduate	118	39.3%
Education	Graduate	102	34.0%
Education	Postgraduate	80	26.7%
Experience with AI	Yes	231	77.0%
Experience with AI	No	69	23.0%

As shown in Table 1, participants were spread throughout the various educational levels and had different experiences with the use of AI technologies. The findings showed that most respondents had already experienced working with AI-driven systems, highlighting the prevalence of technology in the education and research landscape .

Table 2: Comparison of Mean Scores Between Human and Synthetic Respondents

Dimension	Human Mean	Human SD	Synthetic Mean	Synthetic SD
Response Consistency	3.89	0.71	4.21	0.42
Attitude Representation	3.74	0.82	4.09	0.45
Behavioural Prediction	3.61	0.88	4.03	0.40
Construct Alignment	3.79	0.76	4.18	0.38
Overall Validity	3.76	0.79	4.13	0.41

Descriptive findings showed that the mean scores of the synthetic respondents were generally higher than the human respondents on all the dimensions measured. But the average standard deviations of the synthetic responses were significantly smaller. The lower variation proposed higher homogeneity among synthetic added to the percentages of human respondents. There have been themes of homogenisation and decreased diversity in technology-mediated responses discussed in research on digital learning environments and on-line participation in other research tasks. This tendency can be attributed to the formal and optimizing approaches of AI systems .

Table 3: Reliability Analysis

Scale	Cronbach's Alpha
Human Dataset	.87
Synthetic Dataset	.92
Combined Dataset	.89

Cronbach's alpha values were greater than the accepted .70 level, which demonstrated an acceptable level of internal consistency for both sets of items. Synthetic dataset slightly more reliable when compared to human. This suggests that the

responses from synthetic respondents were fairly uniform. Drawings on digital educational technologies and assessment practices have shown similar standardization and consistency as found in these studies .

These observations were corroborated by a qualitative result. Some scholars mentioned that the synthetic respondents would always give responses that were logically consistent but less varied and unpredictable compared to the respondents that were human . Participants noted that humans tend to answer survey questions relatively idiosyncratically, considering the context-based factors and experiences they have had when answering them, whereas synthetic respondents tended to use learned response patterns.

2.2. Inferential Analysis

To check for the statistical significance of the difference between the human respondents and synthetic respondents was carried out an inferential statistics was conducted. To test for response equivalence and to provide a representation, independent sample t-tests, correlation analysis, and factor analysis were used.

Table 4: Independent Sample t-Test Results

Dimension	t-value	p-value
Response Consistency	4.83	.000
Attitude Representation	5.21	.000
Behavioural Prediction	5.67	.000
Construct Alignment	4.92	.000
Overall Validity	5.43	.000

The *t-test* results showed significant difference between human respondents and synthetic respondents at all dimensions of study ($p < .001$). The results showed that there was a resemblance to human responses in general direction of the responses, while it did not fully replicate the human responses. The differences are aligned with the research on education and organizations, which highlight the complexity of human attitudes, behaviours, and perceptions .

Table 5: Correlation Matrix

Variables	Consistency	Attitude	Behaviour	Alignment
Consistency	1.00	.72	.68	.75
Attitude	.72	1.00	.70	.74
Behaviour	.68	.70	1.00	.71
Alignment	.75	.74	.71	1.00

Moderate-to-high positive associations were shown for all variables of the study, indicated by the correlation analysis. The results showed that constructs were conceptually interrelated, but maintained a satisfactory degree of discrimination. Similarly, prior educational research has shown strong relationships between perceptions, engagement and leadership practices and educational outcomes.

Table 6: Exploratory Factor Analysis

Construct	Factor Loading
Response Consistency	.81
Attitude Representation	.84
Behavioral Prediction	.79
Construct Alignment	.86
Overall Validity	.88

For construct validity, values above .50 were deemed acceptable and factor loadings met this criterion. The factor structure indicated that the questionnaire could effectively capture the desired constructs of synthetic construct validity. Various

educational and institutional research studies have employed the same methods of measurement evaluation (Khan & Younas, 2021).

The qualitative results reinforce the quantitative results. The four major themes that emerged from the thematic analysis were: Consistency of synthetic responses, there were limitation in understanding of the context, reduced variability, and ethical concerns for AI-generated data.

The first theme, **consistency of synthetic responses**, reflected participants' acknowledgment that LLM-generated respondents produced logically coherent and internally consistent answers. Participants believed that this characteristic could improve preliminary survey testing and questionnaire validation. These perceptions align with broader discussions concerning AI-supported educational innovations and technology-based systems.

The second theme, **limitations in contextual understanding**, highlighted concerns regarding the inability of synthetic respondents to fully capture lived experiences. Participants argued that personal histories, social backgrounds, emotional influences, and cultural factors often shaped human responses. Similar concerns regarding contextual influences have been reported in qualitative studies focusing on educational experiences and participant narratives.

The third theme, **reduced variability**, emerged as one of the most frequently discussed issues. Experts observed that synthetic responses appeared more uniform than human responses. This perception corresponded closely with the quantitative findings showing smaller standard deviations among synthetic respondents. Studies examining digital participation and identity formation have likewise reported concerns regarding standardization and homogenization within technology-mediated interactions.

The fourth theme, **ethical and methodological considerations**, focused on concerns regarding the responsible use of synthetic data. Participants emphasized the need for transparency when using AI-generated respondents and recommended that synthetic data should supplement rather than replace human data. These findings are consistent with recent discussions regarding ethical governance and responsible AI implementation within educational and research environments.

There was convergence within and between the strands of the study when quantitative and qualitative results were put together. The magnitude of the differences revealed by quantitative evidence differed, and the explanations for these differences provided by qualitative evidence were statistically significant. Even though synthetic respondents provided consistent and consistent responses, the variability of their responses was limited, and their sensitivity to context effects was low, which limited their potential to fully capture the patterns of responses in humans.

The overall results suggested that LLM-generated synthetic respondents have significant capabilities in supporting research activities, including instrument development, pilot testing and exploratory investigations. There were however considerable variations between synthetic or human data, in terms of variability and context representation. Therefore, the synthetic respondents must be considered as complementary resources not as a replacement to human respondents.

2.3. Findings

The study results showed a clear trend of LLM's ability to produce coherent, structured, and coherent answers. Data analysis results for quantitative data showed that with the measured constructs the synthetic respondents showed a pattern of answer which mostly coincided with that of the respondents. There were notable variations with regard to response variability, contextual sensitivity, however, and capturing a range of perspectives. Descriptive analysis revealed that the mean scores for all dimensions which were generated by the synthetic respondents were generally higher than the mean scores produced by the human respondents. Within the synthetic dataset, the standard deviations of the synthetic responses remained lower to exhibit more homogeneity among synthetic responses during the same time.

The inferential analysis further demonstrated statistically significant differences between human and synthetic respondents. Independent sample t-tests indicated that the observed differences were significant across response consistency, attitude representation, behavioural prediction, and construct alignment. While synthetic respondents replicated the overall direction of human responses, they were less successful in reproducing the variation, complexity, and unpredictability observed among human participants. These findings suggest that LLM-generated responses reflected generalized response patterns rather than the full spectrum of human experiences.

Reliability testing produced satisfactory Cronbach's alpha values for both datasets, indicating acceptable levels of internal consistency. Interestingly, the synthetic dataset demonstrated a higher reliability coefficient than the human dataset. Although higher reliability may initially appear advantageous, it also reflected a tendency toward greater standardization and reduced variability. Human respondents expressed varying opinions influenced by personal experiences, educational backgrounds,

socio-cultural environments, and contextual circumstances. Consequently, variability was more pronounced among human participants.

Factor analysis demonstrated that the survey instrument maintained acceptable construct validity across both datasets. Variables loaded adequately on their respective factors, suggesting that synthetic responses preserved essential structural relationships among survey items. Nonetheless, comparisons between factor structures revealed that synthetic respondents generated more predictable factor patterns. Human data demonstrated slightly greater complexity and variability in construct relationships.

The qualitative findings supported and enriched the quantitative results. Through thematic analysis of expert interviews, four major themes emerged. The first theme concerned the consistency and coherence of synthetic responses. Participants acknowledged that LLM-generated respondents often produced logically connected and grammatically accurate answers. According to experts, the structured nature of AI systems contributed to stable response generation and reduced measurement inconsistencies. Similar observations regarding consistency in technology-supported environments have been reported in research related to e-learning and technology-enhanced educational experiences .

The second theme was on the contextual constraint. This is because participants felt that making synthetic respondents would not allow for depth to be achieved when matching the social, emotional, and situational influences on a human's decision making process. The human response was discussed in several respects, with a number of interviewees highlighting the influence of the individual's past, social relationships, family environments and professional experiences. Lessons learned from studies examining educational experiences, parents' views and teachers' accounts also suggest that context is critical to individual outcomes .

The third qualitative theme highlighted that there was less variety in the synthetically produced response. Expert feedback pointed out that there are contradictions and inconsistencies in a human dataset which involve different perspectives. This was contrasted by the synthetic datasets that tended to be more uniform. The result was consistent with the quantitative data, which revealed shorter standard deviations in respondents' answers using Artificial Intelligence. Participants noted that language models tend to be more likely than not to generate answers that follow the common trend or viewpoint, but not the less common one.

The fourth theme was on an ethical aspect. The potential for transparency, accountability, and responsible use of AI-generated data were topics of concern for experts. Explain the role of the synthetic respondents in research and stress the need to identify them. They also said that responses should be generated by AI and not stand in for human responses. These perspectives aligned with the concerns of the times about the ethics of research and education governance and the responsible use of AI .

This is illustratively shown in the convergence of quantitative and qualitative findings in the overall integration of the findings of both strands of the mixed methods investigation. The measurable differences between human and synthetic respondents were found in the statistical analyses; the findings in the interviews gave accounts of why this was the case and what types of impact it led to.

3. Discussion

The results of this research are another important contribution to the debate on using AI in research settings. The results revealed a capacity for coherent, structurally consistent, and valid responses from LLM-synthesized respondents, but not a complete ability to simulate the complexity of human response behavior . The study findings are consistent with the other literature in the fields of education and social science, which give importance to the multi-dimensionality of human perceptions, attitudes and decision-making processes.

The higher consistency observed among synthetic respondents may be explained by the algorithmic foundations of large language models. Unlike humans, AI systems do not possess personal experiences, emotional fluctuations, social pressures, or contextual constraints that influence decision-making. Consequently, synthetic responses tended to be more uniform and predictable. Similar observations have been reported in studies investigating technology-supported environments and digital behaviors, where structured systems frequently exhibit greater consistency than human participants .

The reduced variability identified within the synthetic dataset represents one of the most significant findings of this study . Human responses are inherently diverse because individuals differ in terms of experiences, educational backgrounds, psychological characteristics, socioeconomic circumstances, and cultural influences . Previous research examining parental involvement, student well-being, attitudes toward education, and organizational behaviour consistently demonstrates substantial variation among human participants . The inability of synthetic respondents to reproduce similar levels of variability may limit their representational accuracy.

The qualitative findings further highlighted the importance of contextual factors in shaping human responses. Research related to educational experiences has repeatedly shown that individuals interpret situations differently depending on personal and environmental influences . Participants in the present study argued that synthetic respondents lack authentic experiences and therefore cannot fully reproduce the nuanced reasoning processes underlying many human responses. This limitation may become particularly problematic in studies exploring sensitive social, cultural, psychological, or educational issues.

The results have far-reaching methodological implications as well. The reliability and validity found in the synthetic data indicate that synthetic respondents generated by LLM can be employed in the initial stages of research work . Synthetic respondents, for example, can help researchers test instruments and define questions and structures for the questions, as well as troubleshoot measurement issues before conducting a significant data-gathering process . In the field of educational assessment and institution evaluation research, there are some variations similar to the ones noted above in current practices encouraged (Khan & Younas, 2021).

Virtually, synthetic answers have a number of benefits. They offer quick access to large volumes of data, lower financial expenses and remove a lot of the logistical issues inherent in recruitment of target participants . Research on online learning environment and technology integration shows that the use of technologies like the Internet and electronic resources has been becoming more and more predominant in the learning and research process . Based on the present findings, it is believed that if this is done appropriately, synthetic respondents can participate and move the developments in a positive manner.

The results also provide evidence for the cautions raised by arguments that call for the replacement of human subjects . The importance of the involvement of authentic voices, lived experiences and contextual understanding within the field of educational studies to create meaningful findings is a consistent theme found throughout. Currently, these dimensions are not equally replicable by the synthetic respondents. They should therefore, not replace but compliment other teaching methods (Younas, Taj, Khan, Hussain, Makhdum & Khan, 2023).

The findings also are in line with the on-going debate around the development of responsible artificial intelligence and its ethical governance (Younas, Taj, Khan, Hussain, Makhdum & Khan, 2023). With the growing application of AI in the field of education and research, transparency, reliability, accountability, and validity become more critical issue. These laboratory-scale procedures must be validated, and data produced by AI must go through the same procedure as human research to ensure it is valid and usable in scientific research.

4. Conclusion

The findings of this study showed that AI-powered synthetic respondents have great potential as tools for research studies but were not yet fully capable of substituting for real respondents (Younas, Taj, Khan, Hussain, Makhdum & Khan, 2023). Quantitative results proved that synthetic respondents were able to produce reliable, coherent and structurally coherent responses. However, there was a significant variation regarding variability, contextual sensitivity and representation of different perspectives . More complexity and heterogeneity and contextual awareness were shown by human respondents than by synthetic respondents.

The qualitative results also indicated support for the above, by revealing that experts were concerned with the lack of lived experience, emotional factors and context in the AI-generated responses (Younas, Khan & Taj, 2023). Experts showed that synthetic respondents are useful for pilot studies, instrument testing, exploratory analyses and preliminary studies, but also recognized the importance and the need for human participants in empirical studies.

The integration of quantitative and qualitative evidence provided a comprehensive assessment of synthetic respondent validity (Younas, Khan & Taj, 2023). While LLM-generated respondents demonstrated strengths in consistency, efficiency, and scalability, limitations remained evident concerning authenticity and representativeness (Younas, Khan & Taj, 2023). These limitations suggest that synthetic respondents should be employed cautiously and within clearly defined methodological boundaries.

The study adds to the research on AI and research methodology, offering insights into the potential and constraints of synthetic respondents (Younas, Khan & Makhdum, 2022). Going forward, exploring ways to enhance the contextual sensitivity of synthetic users, gain more response diversity, and provide more accurate representations of synthetic participants would be desirable . Human respondents, coupled with synthetic respondents, will continue to be the most accurate way to obtain behavioural and attitudinal information, until the improvements mentioned above are realized, in which case synthetic respondents can prove to be valuable and complementary tools in modern research (Younas, Khan & Makhdum, 2022).

References

Abbas, Q., Faisal, F., Jabeen Bhutta, M., Salim, Z., & Kanwal, S. (2024). The Role of Quality Teaching of University Teachers in the Motivation of University Students. Journal of Social Sciences Research & Policy, 2(3), 161 – 169. Retrieved from https://jssrp.org.pk/index.php/jssrp/article/view/95

Faisal, A., Akmal, S., Makhdum, F.N., Sandhu, H. R., Younas, A., Salim, Z., (2023). Effect of Innovative Leadership Prompted by Technology on Empowering English Language Teachers In Pakistan. Journal of positive school of psychology. Vol. 7 No. 1. 95-105 https://journalppw.com/index.php/jpsp/article/view/15020/9737

Khan, S., & Younas, A. (2021). Impact of Intrinsic Motivation on Student's Learning at Elementary Level. Global Educational Studies Review, VI(II), 174-187 . https://doi.org/10.31703/gesr.2021(VI-II).17

Khan, S., Taj, S., Zill-e-Huma, & Younas, A. (2024). A comparative study of quality assurance policies in public and private universities of Pakistan. Journal of Asian Development Studies, 13(4). https://doi.org/10.62345/jads.2024.13.4.82

Makhdum, F. N., Khan, B. S., Munir, F., Younas, A., & Faisal, A. (2022). Ubiquitous learning workforce in the learning world: The middle class. Pakistan Journal of Social Sciences, 42(3), 523–533. ISSN: 2074-2061. https://pjss.bzu.edu.pk/index.php/pjss/article/view/1137

Makhdum, F. N., Khanam, A., Faisal, A., & Sandhu, H. R. (2023). Impact of Kahoot! on students' engagement and learning outcome at the elementary level in Pakistan: Their perception towards Kahoot! assessment. Journal of Positive School Psychology, 7(1), 64–78. https://journalppw.com/index.php/jpsp/article/view/15017/9736

Makhdum, F. N., Sandhu, H. R., Batool, T., Khan, S., Faisal, F., & Younas, A. (2023). Effect of using Symbolab calculator in teaching simultaneous equations on students' conceptual understanding at the elementary level in Pakistan: Mathematics attitude in technological corners. Journal of Positive School Psychology, 7(1). https://journalppw.com/index.php/jpsp/article/view/15022

Maria, M., Makhdum, D. F. N., Sandhu, M. K., Khan, S., & Younas, D. A. (2025). Smart Learning in Pakistani Universities: A Mixed-Methods Study of Progress, Challenges, and Future Directions. Research Journal for Social Affairs, 3(5), 173-184. https://doi.org/10.71317/RJSA.003.05.0316

Qayyum, A. (2025). "The Ghost of Comilla": Authoritarian biopolitics and global development in rural East Pakistan. Diplomatic History, 49(2), 201-228.

Qayyum, A., & Hussain, S. (2019). Effect of scaffolding on social and emotional learning (SEL) of elementary school students. JSSH, 27(1).

Qayyum, A., Khalil, S., Arif, S., Rafiq, S., & Makhdum, F. N. (2025). Parents' Involvement in Early Childhood Care and Education: A Case Study. Research Journal for Social Affairs, 3(4), 311-318.

Salim, Z., Makhdum, F. N., & Younas, A. (2022). Parents teachers and students perception regarding teachings of the Holy Quran as a compulsory part of curriculum. *Alhamd Multidisciplinary Research Journal*, *2*(1), 10. http://amrj.aiu.edu.pk/wp-content/uploads/2023/02/AMRJ-10.-1-vol2-issue1.pdf https://www.researchgate.net/publication/375911647_parents_teachers_and_students_perception_regarding_teachings_of_the_holy_quran_as_a_compulsory_part_of_curriculum

Younas, A., Khan, B. S., & Taj, S. (2023). Structural dimensions of classroom quality at the early childhood education. Archives of Educational Studies, 3(1), 130–145. https://ares.pk/

Younas, A., Khan, S., & Makhdum, F. N. (2022). A comparative study of early childhood education services provided by public and private institute in Pakistan. International Journal of Early Childhood Special Education, 14(5). https://doi.org/10.9756/INTJECSE/V14I5.464

Younas, A., Taj, S., Khan, S., Hussain, S., Makhdum, F. N., & Khan, B. S. (2023). Classroom quality in terms of structural and dimensions at early childhood education level in Pakistan. Journal of Positive School Psychology, 7(1), 79–94. http://journalppw.com