Scientific Research Publishing

International Journal of Intelligence Science, 2027, 17(1), 1-12
https://www.scirp.org/journal/ijis
ISSN Online: 2163-0356
ISSN Print: 2163-0283

Perceived Quality of Siloed versus Time-Synchronized Data Architectures for AI-Based Digital Mental Health: A Within-Subjects Scenario Study

Saranya Vaithilingam

University of the Cumberlands, Williamsburg, KY, USA
Email: sara.lingam4@gmail.com

How to cite this paper: Vaithilingam, S. (2027) Perceived Quality of Siloed versus Time-Synchronized Data Architectures for AI-Based Digital Mental Health: A Within-Subjects Scenario Study. *International Journal of Intelligence Science*, **17**, 1-12.
https://doi.org/10.4236/ijis.2027.171001

Received: October 14, 2026
Accepted: November 22, 2026
Published: November 25, 2026

Copyright © 2027 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/

Open Access

Abstract

Fragmented data architectures may limit how artificial intelligence (AI)-based digital mental health systems combine and communicate multimodal information. This scenario-based study tested whether users perceived the Integrated framework more favorably than the Traditional framework. U.S. adults familiar with digital mental health technologies reviewed standardized descriptions of both conditions and rated system quality, information quality, and perceived transparency and explainability. The constructs covered data quality, portability, response time, accuracy, timeliness, completeness, transparency, and explainability. Paired-samples t tests were supplemented by Wilcoxon signed-rank tests. Valid paired samples were 277, 267, and 258 for the three constructs, respectively. Ratings favored the Integrated condition for system quality (M = 3.47 vs. 3.38; t (276) = −2.66, P = 0.008, $|d_z|$ = 0.16), information quality (M = 3.48 vs. 3.34; t (266) = −4.49, P < 0.001, $|d_z|$ = 0.28), and perceived transparency and explainability (M = 3.79 vs. 3.65; t (257) = −3.04, P = 0.003, $|d_z|$ = 0.19). Nonparametric analyses supported the same conclusions. Absolute paired effect magnitudes ranged from $|d_z|$ = 0.16 to $|d_z|$ = 0.28. Because participants evaluated descriptions rather than functioning systems and temporal alignment was bundled with integration, the findings identify a modest perceptual advantage rather than causal, technical, or clinical-performance benefits. The results support further user-centered prototype and implementation testing.

Keywords

Artificial Intelligence, Time-Synchronized, Digital Mental Health, Data Integration, Multimodal Data, Explainable Artificial Intelligence,

Information Systems, System Quality, User Perceptions, Temporal Alignment, Within-Subjects Design

1. Introduction

Digital mental health systems draw on electronic health records, mobile applications, self-report measures, therapy interactions, smartphones, and wearable sensors. These sources can provide complementary clinical, behavioral, linguistic, and physiological signals for continuous assessment and personalized care [1]-[3]. Their value, however, depends on whether information can be combined coherently. Fragmentation across vendors, devices, timestamps, sampling frequencies, preprocessing pipelines, and annotation conventions can obstruct multimodal interpretation [4]-[6]. In mental health, a signal may acquire meaning only in relation to other events: a physiological change, linguistic cue, sleep disruption, or activity shift can be misleading when interpreted without temporal and behavioral context.

Time-synchronized integration in this study meant aligning information streams by sequence and timing so that related observations could be considered together. The two framework-description conditions are labeled Traditional and Integrated throughout this report. Traditional described reliance on a single model; Integrated described combining multiple sources with time synchronization and cross-system coordination. Prior work has described the potential of multimodal AI and integrated digital interventions to support contextual interpretation and adaptive workflows [4] [7]. Yet technical integration does not establish that a system will be perceived as accurate, responsive, understandable, or useful. Explainable-AI research likewise emphasizes that explanations must be intelligible and decision-relevant to people, not merely available as algorithmic outputs [8] [9]. A systems design must therefore communicate how information is organized and how outputs can be understood.

The study was grounded in the DeLone and McLean Information Systems Success Model [10]. System quality was operationalized through perceived data quality, portability, and response time; information quality through perceived accuracy, timeliness, and completeness. The third outcome assessed perceived transparency and explainability using items adapted from the System Causability Scale [9]. DeLone and McLean's net-benefits construct concerns impacts on defined stakeholders; the present explanation ratings address a narrower aspect of user perception [10]. System-quality and information-quality items were adapted from the IS-Impact Model [11].

Evidence about integrated digital health has largely emphasized technical potential, conceptual models, or implementation barriers. Less is known about whether non-specialist users recognize quality differences when the same evaluation criteria are applied to Traditional and Integrated framework descriptions.

The present study therefore compared within-person ratings of the two descriptions across three evaluation constructs. Its contribution is not a validation of an architecture; it is an empirical test of whether proposed architectural features are visible to users as differences in information-system quality. The study further identifies which evaluation construct shows the largest perceptual difference and can inform subsequent prototype evaluation.

2. Methods

2.1. Study Design

This quantitative, nonexperimental, within-subjects study compared ratings of two standardized AI-based digital mental health framework descriptions presented in a fixed order. Framework description was the within-participant factor; system quality, information quality, and perceived transparency and explainability were the evaluation constructs. Each participant evaluated Traditional before Integrated. The fixed order was not randomized or counterbalanced, so condition differences cannot be separated from order and wording effects.

2.2. Participants, Recruitment, and Survey Administration

Eligible participants were U.S. residents aged 18 years or older. The consent form specified the experience criterion as "Have experience using or evaluating digital mental health platforms." The survey questionnaires did not reproduce or test separate familiarity screening questions or their qualifying responses. Participants were recruited through a paid SurveyMonkey Audience panel using eligibility criteria for U.S. residence, adulthood, and experience using or evaluating digital mental health platforms. This was a nonprobability panel sample; specific panel quota settings were not available in the archived recruitment materials. The invitation-only, self-administered questionnaire was hosted on SurveyMonkey. Electronic consent preceded the study items; page logic ended the survey for individuals who declined consent. Survey items were nonmandatory, so participants could skip questions or discontinue without penalty.

The platform recorded 422 survey starts, of which 406 included recorded consent and 283 were classified as complete across all starts. Construct-level paired samples depended on valid scores for both conditions rather than whole-survey completion.

2.3. Framework Conditions and Procedure

The Traditional definition described reliance on a single AI model without multiple data sources; the Integrated definition described combining diverse sources for comprehensive, personalized, explainable insights with time synchronization and cross-system coordination. The Traditional definition preceded the Integrated definition, and each matrix placed Traditional above Integrated. The survey provided both verbatim descriptions and instructions in a fixed order. Both conditions used identical rating items in one survey version without randomiza-

tion or counterbalancing. No working platform, AI model, clinical intervention, or personal health-data integration was presented. Temporal alignment was not independently manipulated; differences apply to the complete descriptions.

2.4. Measures

Items were adapted with permission from the IS-Impact Model and System Causability Scale [9] [11]. Wording was revised for AI-based digital mental health while retaining the intended concepts. Responses used a five-point Likert scale from 1 (strongly disagree) to 5 (strongly agree), and the same items were presented for both conditions. Negatively worded items were reverse-coded before aggregation. The system-quality composite comprised nine items across the labeled domains of data quality, portability, and response time. The information-quality composite comprised nine items across accuracy, timeliness, and completeness. The three-item composite originally labeled "net benefits" is reported here as perceived transparency and explainability to reflect its item content. It did not measure clinical effectiveness, organizational return, or objective performance.

Internal consistency was estimated separately by construct and condition. Cronbach's alpha was 0.83 for system quality in each condition; 0.77 and 0.85 for information quality in the Traditional and Integrated conditions, respectively; and 0.78 and 0.76 for perceived transparency and explainability. These values supported score consistency but did not by themselves establish construct validity in this scenario-based context.

2.5. Statistical Analysis

A G*Power 3.1.9.7 analysis for a two-tailed paired-samples t test assumed $|d_z| = 0.24$, alpha = 0.05, and power = 0.95 and produced a minimum of 228 paired observations [12]. Analyses were conducted in JASP 0.19.3. The primary tests were two-tailed paired-samples t tests for the three evaluation constructs. Differences retained the Traditional minus Integrated order. Paired-samples Cohen's d_z was the mean paired difference divided by the standard deviation of paired differences; $|d_z|$ denotes its magnitude [13] [14]. Magnitudes are reported numerically alongside the response scale, without assigning categories based on between-subject benchmarks [14]. Shapiro-Wilk tests assessed the paired-difference distributions, and Wilcoxon signed-rank tests served as robustness analyses.

Missingness was handled separately for each construct using pairwise deletion at the construct-comparison level. Responses used a five-point Likert scale from 1 (strongly disagree) to 5 (strongly agree); values of 0 were treated as missing. Negatively worded items were reverse-coded before aggregation. Within each framework condition, a construct score was calculated from the available valid item responses when at least one item assigned to that construct had been answered. Each paired comparison required valid construct scores for both the Traditional and Integrated conditions. If either construct score was missing or invalid, the participant's score for the other condition was also excluded from that construct com-

parison; valid paired scores for other constructs remained eligible. This procedure yielded 277 paired observations for system quality, 267 for information quality, and 258 for perceived transparency and explainability. Statistical significance was evaluated at alpha = 0.05. The primary analyses used unadjusted P values. A post hoc Holm procedure assessed whether the three paired-samples *t*-test conclusions held at a familywise alpha of 0.05.

2.6. Ethics and Data Protection

The University of the Cumberlands Institutional Review Board approved the procedures (approval no. 0925-250932; September 21, 2025). Electronic informed consent was required to proceed. Participation was voluntary. The investigator did not collect or access names, email addresses, IP addresses, institutional affiliations, or other direct identifiers, and received anonymized responses. Exported survey data were stored in password-protected files accessible only to the principal investigator and are scheduled for destruction after the approved three-year retention period.

3. Results

3.1. Participant Flow and Sample Characteristics

Of 422 recorded starts, 406 had recorded consent (96.2%). The platform classified 283 records as complete across all starts (67.1%). Usable paired samples were 277 for system quality, 267 for information quality, and 258 for perceived transparency and explainability; all exceeded the power-analysis minimum of 228. Demographic table data summarizes all survey starts rather than the consenting subset. Demographic variables were descriptive and were not modeled as moderators.

3.2. Primary Comparisons

Shapiro-Wilk tests indicated departures from normality for the paired-difference distributions. Observed values within the response scale were retained. Paired-samples *t* tests and Wilcoxon signed-rank robustness tests are presented together in Table 1. Paired score differences across the three constructs are shown in Figure 1.

Table 1. Paired comparisons and Wilcoxon robustness analyses.

Statistic	System quality	Information quality	Perceived transparency and explainability
Items per condition	9	9	3
Paired *n*	277	267	258
Traditional descriptive *n*	277	267	267
Integrated descriptive *n*	277	267	258
Traditional *M* (SD)	3.38 (0.57)	3.34 (0.59)	3.65 (0.86)
Integrated *M* (SD)	3.47 (0.59)	3.48 (0.58)	3.79 (0.87)
t (df)	−2.66 (276)	−4.49 (266)	−3.04 (257)
P for *t*-test	0.008	<0.001	0.003

Continued

$\lvert d_z \rvert$ [95% CI]	0.16 [0.04, 0.27]	0.28 [0.15, 0.39]	0.19 [0.06, 0.31]
W positive-rank sum	7276.50	5116.00	3902.00
z	−2.78	−5.12	−3.30
P for Wilcoxon test	0.005	<0.001	<0.001
Rank-biserial r [95% CI]	−0.23 [−0.37, −0.07]	−0.43 [−0.55, −0.28]	−0.31 [−0.46, −0.13]

Note. CI = confidence interval. d_z uses the SD of paired differences; $\lvert d_z \rvert$ reports its magnitude. W is the sum of positive ranks for Traditional minus Integrated. Descriptive sample sizes are reported separately because they can differ from the paired sample size.

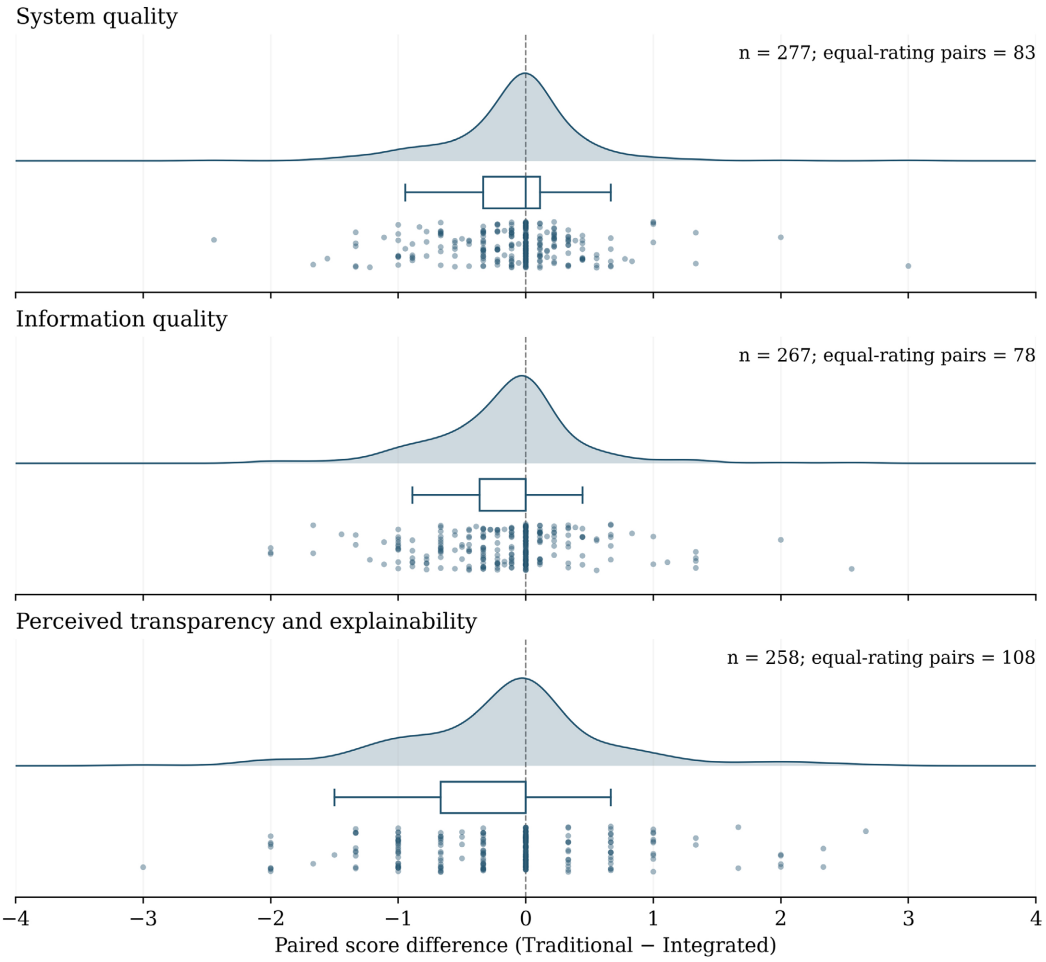

Figure 1. Distributions of paired construct-score differences (Traditional minus Integrated) for system quality (n = 277), information quality (n = 267), and perceived transparency and explainability (n = 258). Density curves, boxplots, and individual points depict the differences; negative values indicate higher ratings for Integrated.

System quality was higher for the Integrated condition (M = 3.47, SD = 0.59) than for the Traditional condition (M = 3.38, SD = 0.57), t (276) = −2.66, P = 0.008, $\lvert d_z \rvert$ = 0.16 (**Figure 2** and **Figure 3**). Information quality showed the largest paired separation: integrated M = 3.48 (SD = 0.58) versus traditional M = 3.34 (SD = 0.59), t (266) = −4.49, P < 0.001, $\lvert d_z \rvert$ = 0.28 (**Figure 4** and **Figure 5**). Perceived

transparency and explainability ratings were also higher for the Integrated condition ($M = 3.79$, SD = 0.87) than the Traditional condition ($M = 3.65$, SD = 0.86), $t(257) = -3.04$, P = 0.003, $|d_z| = 0.19$ (Figure 6 and Figure 7). Absolute paired effect magnitudes ranged from $|d_z| = 0.16$ to $|d_z| = 0.28$.

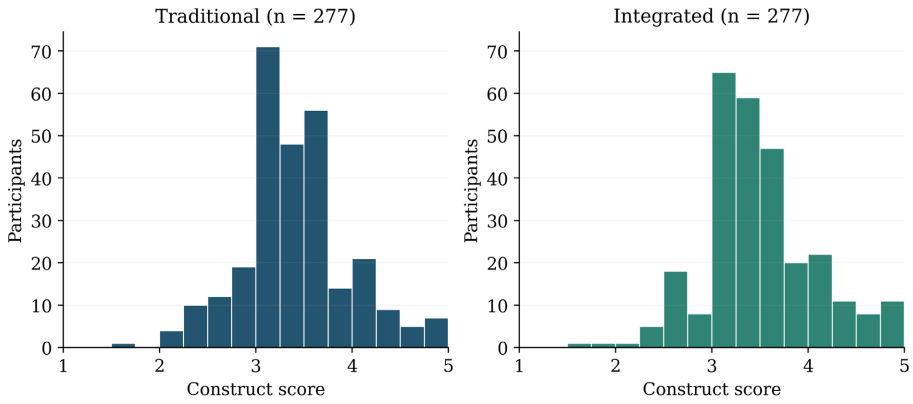

Figure 2. Histograms of system quality scores for the Traditional and Integrated conditions (paired sample: n = 277 per condition).

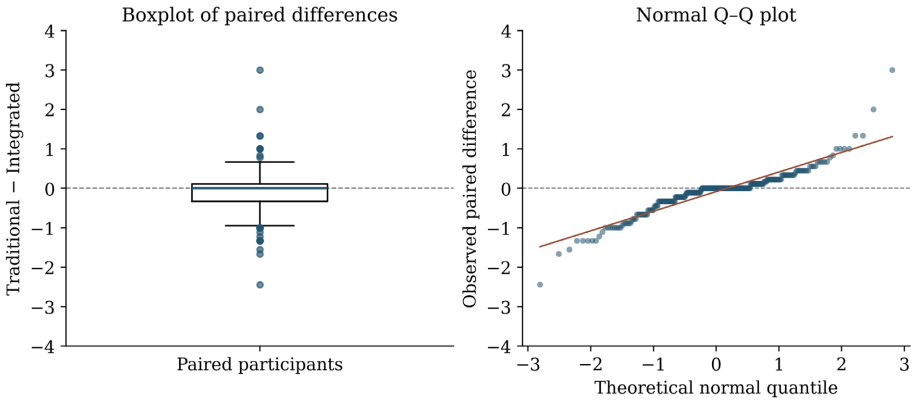

Figure 3. Boxplot and normal Q-Q plot of paired system quality score differences (Traditional minus Integrated; n = 277).

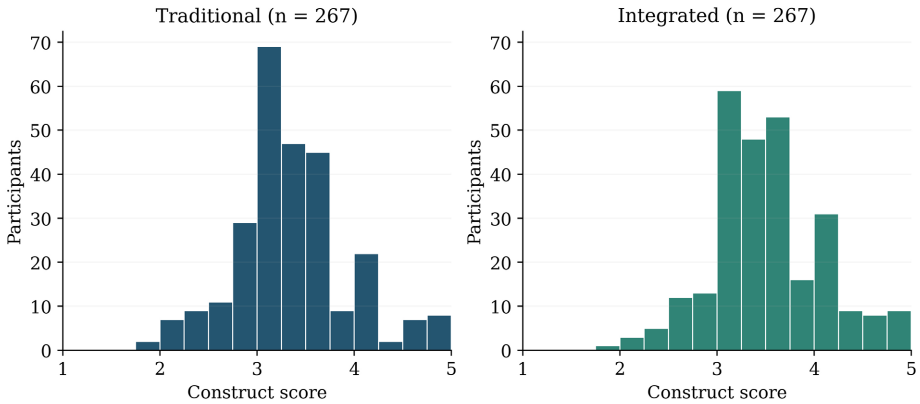

Figure 4. Histograms of information quality scores for the Traditional and Integrated conditions (paired sample: n = 267 per condition).

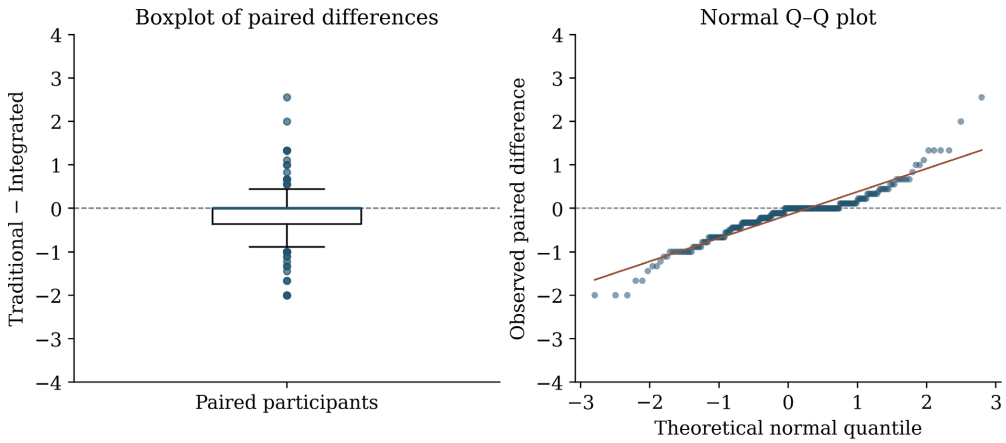

Figure 5. Boxplot and normal Q-Q plot of paired information quality score differences (Traditional minus Integrated; n = 267).

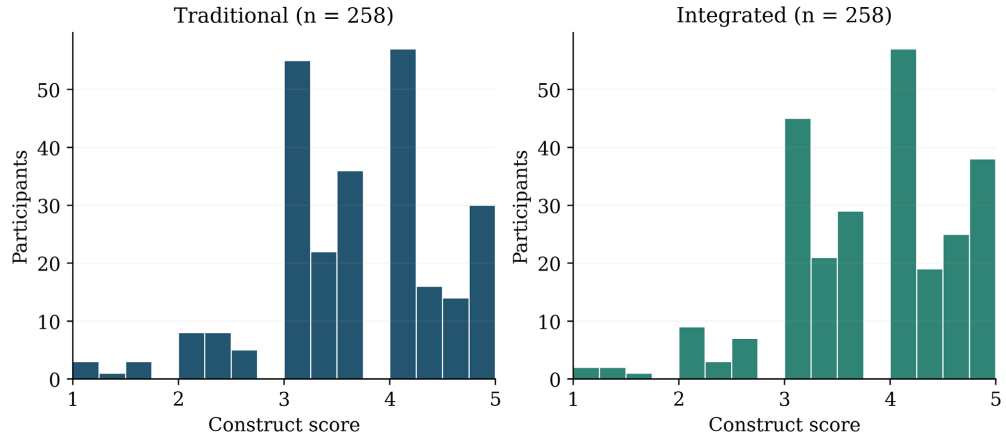

Figure 6. Histograms of perceived transparency and explainability scores for the Traditional and Integrated conditions (paired sample: n = 258 per condition).

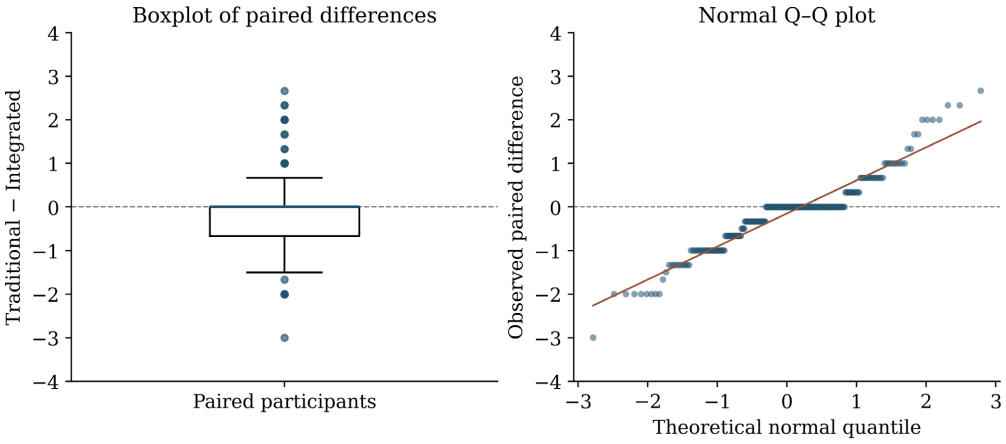

Figure 7. Boxplot and normal Q-Q plot of paired perceived transparency and explainability score differences (Traditional minus Integrated; n = 258).

All three paired-samples *t*-test comparisons remained significant in a post hoc

Holm sensitivity check. The reported P values for information quality ($P < 0.001$), perceived transparency and explainability ($P = 0.003$), and system quality ($P = 0.008$) were below their sequential thresholds of 0.0167, 0.025, and 0.05, respectively.

3.3. Robustness Analyses

Wilcoxon signed-rank tests supported the same significance conclusions as the paired-samples *t* tests (Table 1).

4. Discussion

4.1. Principal Findings

Participants gave higher mean ratings to Integrated than Traditional for all three evaluation constructs. Information quality showed the largest standardized paired difference ($|d_z| = 0.28$), followed by perceived transparency and explainability ($|d_z| = 0.19$) and system quality ($|d_z| = 0.16$). These are magnitudes standardized by the SD of paired differences, not between-subject standardized effects [14]. The findings indicate a perceptual preference for the Integrated framework description.

The information-quality result suggests that coordinated, temporally ordered inputs made anticipated outputs appear more accurate, timely, and complete. This interpretation is compatible with multimodal-health research, which treats data fusion as a means of recovering context that individual streams cannot supply [15]. The design also cannot determine whether ratings reflected timing, multimodal integration, the general sophistication of the Integrated description, or their combination.

The difference in perceived transparency and explainability indicates greater expected explanatory clarity. The System Causability Scale informed a narrow operationalization centered on transparency and explainability [9]. Medical explainability requires understandable outputs as well as faithful links to model behavior and the decision context [16]. Mental health applications add safety concerns when opaque outputs reach users without professional mediation [17]. The study measured whether one framework sounded more understandable; it did not establish explanation fidelity, calibrated trust, or safe reliance.

4.2. Comparison with Prior Work and Implications

The study applies information-systems quality measures to an architectural comparison in digital mental health. The observed differences concern expectations elicited by descriptions, so their relevance to system adoption and implementation requires further testing.

For future prototype studies, the author recommends documenting how data streams are aligned, preserving data provenance, and showing the evidence underlying explanations. These engineering recommendations extend beyond the present perceptual comparisons. Evaluation should combine user ratings with

technical performance measures and assess workflow fit, governance, and maintenance [18].

4.3. Limitations

Several limitations constrain inference. Participants evaluated written scenarios, not operating systems; favorable wording may have produced expectancy or demand effects. Integration and temporal alignment were bundled. The fixed presentation order, with Traditional before Integrated in the definitions and rating matrices, may have introduced order, contrast, or carryover effects that cannot be separated from the observed perceptual differences. The non-probability U.S. panel limited demographic characterization, and lack of a nonresponse-bias assessment limit generalizability. All construct ratings were self-reported in one session, and potentially relevant characteristics such as digital literacy, privacy concern, prior AI experience, and mental health service experience were not modeled. Measures were adapted to a hypothetical context. Although internal consistency was acceptable to good, the study did not establish convergent, discriminant, or criterion validity. The three adapted System Causability Scale items assess anticipated explanation quality rather than directly evaluating explanations from an operating system [9]. They do not measure the broader balance of benefits and harms represented by DeLone and McLean's net-benefits construct [10]. Scores cannot be generalized to clinical outcomes, safety, organizational value, or cost.

4.4. Future Research

Future work should use functioning prototypes and a factorial design that manipulates integration and temporal alignment separately. Condition order should be randomized or counterbalanced, and observable tasks should accompany ratings. Studies should include clinicians, patients, caregivers, and administrators, test measurement validity across groups, and evaluate whether initial perceptions predict comprehension, workflow performance, calibrated trust, and safe use. Longer-term research could examine continuously updated patient-specific models, but such digital-twin approaches require explicit privacy, governance, and clinical-use safeguards [19].

4.5. Conclusions

Participants rated the Integrated framework description more favorably than the Traditional framework description on system quality, information quality, and perceived transparency and explainability. Paired effect magnitudes ranged from $|d_z| = 0.16$ to $|d_z| = 0.28$, with the largest difference in information quality. These findings do not show that temporal alignment caused the difference or that an integrated system performs better clinically or technically. They establish a user-centered signal that should be tested with operating prototypes, separately manipulated architectural features, counterbalanced presentation, and objective measures of information handling, explanation quality, task performance, and safety.

Data Availability

Deidentified participant-level data and supporting materials may be made available by the corresponding author upon reasonable request, subject to review for consistency with participant consent, University of the Cumberlands Institutional Review Board approval, and applicable privacy and ethical requirements.

Generative AI Use

OpenAI ChatGPT was used to check author-original sentences grammatically and for journal-formatting assistance.

Acknowledgements

The author thanks the study participants and Kimberly Lowrey, PhD, Mickey Hendon, PhD, Lori Farr, PhD, and James Webb, PhD, for mentorship and guidance during the doctoral research from which this manuscript was developed.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Mohr, D.C., Weingardt, K.R., Reddy, M. and Schueller, S.M. (2017) Three Problems with Current Digital Mental Health Research. . . and Three Things We Can Do about Them. *Psychiatric Services*, **68**, 427-429. https://doi.org/10.1176/appi.ps.201600541

[2] Onnela, J. and Rauch, S.L. (2016) Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health. *Neuropsychopharmacology*, **41**, 1691-1696. https://doi.org/10.1038/npp.2016.7

[3] Torous, J., Linardon, J., Goldberg, S.B., Sun, S., Bell, I., Nicholas, J., *et al.* (2025) The Evolving Field of Digital Mental Health: Current Evidence and Implementation Issues for Smartphone Apps, Generative Artificial Intelligence, and Virtual Reality. *World Psychiatry*, **24**, 156-174. https://doi.org/10.1002/wps.21299

[4] AlSaad, R., Abd-alrazaq, A., Boughorbel, S., Ahmed, A., Renault, M., Damseh, R., *et al.* (2024) Multimodal Large Language Models in Health Care: Applications, Challenges, and Future Outlook. *Journal of Medical Internet Research*, **26**, e59505. https://doi.org/10.2196/59505

[5] Steidl, M., Felderer, M. and Ramler, R. (2023) The Pipeline for the Continuous Development of Artificial Intelligence Models—Current State of Research and Practice. *Journal of Systems and Software*, **199**, Article 111615. https://doi.org/10.1016/j.jss.2023.111615

[6] Zhang, T., Schoene, A.M., Ji, S. and Ananiadou, S. (2022) Natural Language Processing Applied to Mental Illness Detection: A Narrative Review. *npj Digital Medicine*, **5**, Article No. 46. https://doi.org/10.1038/s41746-022-00589-7

[7] Mohr, D.C., Schueller, S.M., Montague, E., Burns, M.N. and Rashidi, P. (2014) The Behavioral Intervention Technology Model: An Integrated Conceptual and Technological Framework for eHealth and mHealth Interventions. *Journal of Medical Internet Research*, **16**, e146. https://doi.org/10.2196/jmir.3077

[8] Bienefeld, N., Boss, J.M., Lüthy, R., Brodbeck, D., Azzati, J., Blaser, M., *et al.* (2023)

Solving the Explainable AI Conundrum by Bridging Clinicians' Needs and Developers' Goals. *npj Digital Medicine*, **6**, Article No. 94. https://doi.org/10.1038/s41746-023-00837-4

[9] Holzinger, A., Carrington, A. and Müller, H. (2020) Measuring the Quality of Explanations: The System Causability Scale (SCS). *Künstliche Intelligenz*, **34**, 193-198. https://doi.org/10.1007/s13218-020-00636-z

[10] Delone, W.H. and McLean, E.R. (2003) The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. *Journal of Management Information Systems*, **19**, 9-30.

[11] Gable, G.G., Sedera, D. and Chan, T. (2008) Re-Conceptualizing Information System Success: The IS-Impact Measurement Model. *Journal of the Association for Information Systems*, **9**, 377-408. https://doi.org/10.17705/1jais.00164

[12] Faul, F., Erdfelder, E., Lang, A. and Buchner, A. (2007) G*Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences. *Behavior Research Methods*, **39**, 175-191. https://doi.org/10.3758/bf03193146

[13] Cohen, J. (1988) Statistical Power Analysis for the Behavioral Sciences. 2nd Edition, Lawrence Erlbaum Associates.

[14] Lakens, D. (2013) Calculating and Reporting Effect Sizes to Facilitate Cumulative Science: A Practical Primer for T-Tests and ANOVAs. *Frontiers in Psychology*, **4**, Article 863. https://doi.org/10.3389/fpsyg.2013.00863

[15] Kline, A., Wang, H., Li, Y., Dennis, S., Hutch, M., Xu, Z., *et al.* (2022) Multimodal Machine Learning in Precision Health: A Scoping Review. *npj Digital Medicine*, **5**, Article No. 171. https://doi.org/10.1038/s41746-022-00712-8

[16] Tjoa, E. and Guan, C. (2021) A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI. *IEEE Transactions on Neural Networks and Learning Systems*, **32**, 4793-4813. https://doi.org/10.1109/tnnls.2020.3027314

[17] Manríquez Roa, T. and Biller-Andorno, N. (2023) Black Box Algorithms in Mental Health Apps: An Ethical Reflection. *Bioethics*, **37**, 790-797. https://doi.org/10.1111/bioe.13215

[18] Sendak, M.P., D'Arcy, J., Kashyap, S., Gao, M., Nichols, M., *et al.* (2020) A Path for Translation of Machine Learning Products into Healthcare Delivery. *EMJ Innovation*, **2020**, 1-14.

[19] Bruynseels, K., Santoni de Sio, F. and van den Hoven, J. (2018) Digital Twins in Health Care: Ethical Implications of an Emerging Engineering Paradigm. *Frontiers in Genetics*, **9**, Article 31. https://doi.org/10.3389/fgene.2018.00031