

Usability of an Insulin Injection Educational Tool Across Nursing Experience Levels: An Explanatory Sequential Mixed-Methods Study

NAN XUN

The First Affiliated Hospital, Sun Yat-sen University

YI LI

The First Affiliated Hospital, Sun Yat-sen University

QIANQIAN WANG

The First Affiliated Hospital, Sun Yat-sen University

RUI HUANG

The First Affiliated Hospital, Sun Yat-sen University

JINGGUO WU

wujguo@mail.sysu.edu.cn

The First Affiliated Hospital, Sun Yat-sen University

HUANSU LI

The First Affiliated Hospital, Sun Yat-sen University

Research Article

Keywords: System Usability Scale, Nursing experience levels, Mixed-methods research, Nursing education, Insulin injection teaching

Posted Date: September 25th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10684848/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background: Insulin injection teaching is time-consuming for nurses, yet current educational tools may not adequately adapt to users at different experience levels.

Purpose: This study aimed to examine whether the perceived usability of an insulin injection teaching toolkit varies across nursing experience levels.

Methods: Using an explanatory sequential mixed-methods design, we stratified 26 nurses into interns ($n=9$), junior nurses ($n=7$), and senior nurses ($n=10$) based on Benner's framework. Participants completed the System Usability Scale (SUS) and semi-structured interviews.

Results: Median SUS scores were 75.0 for interns, 92.5 for junior nurses, and 87.5 for senior nurses. The intern–junior gap (17.5 points, $r=0.50$) did not remain statistically significant after correction. However, qualitative data revealed that while junior nurses described seamless workflow integration, interns required structured guidance, and senior nurses showed polarized responses. Three usability bottlenecks were identified: lack of guidance for beginners, visual information overload, and physical layout affecting workflow fluency.

Conclusions: Aggregated SUS means may mask experience-level differences. Future SUS evaluations should consider stratified reporting by experience level. Qualitative findings suggest that perceived usability is shaped by person–tool fit rather than experience alone.

Introduction

In endocrinology wards, giving insulin to patients and teaching them to inject themselves consume a substantial proportion of nursing time[1,2]. Minor deviations in injection technique can affect blood sugar levels[3], and poor patient teaching can lead to low blood sugar events at home[4]. Guidelines are continuously updated[5], and procedures get more detailed. Yet gaps persist between evolving best practices and the educational resources available in daily clinical practice[6]. Consequently, structured injection toolkits have been introduced.

The availability of high-quality tools does not guarantee optimal educational outcomes. They can improve technique[7,8], but such tools often assume users are all at the same skill level. In reality, that's not what we see. Senior nurses find the toolkit easy to use and practical. Interns, however, often have to pause mid-teaching to search for specific items, thereby disrupting the instructional workflow. That contrast got us thinking: a 'one-size-fits-all' design probably doesn't work across different experience levels[9,10]. So we made targeted modifications: reorganised the contents into designated sections, cut down the clutter, and made it quicker to find things.

The System Usability Scale (SUS)[11] is a common way to assess how usable people find a product. It's been used for teaching tools too. But a critical limitation remains: most studies just report the aggregate mean score[12]. This approach obscures potential variations between user groups. Early in our study, we

saw exactly that: one nursing intern gave the toolkit a perfect score, while a nurse with 13 years of experience gave it only 75, a 25-point gap. And studies that actually break SUS down by experience level? We couldn't find many[13].

Insulin teaching isn't just about whether nurses can do the procedure right. It also affects whether patients can inject themselves and keep their blood sugar stable[14]. Despite this clinical importance, most existing tools haven't paid much attention to how different experience levels perceive usability[13]. So this study aims to analyse SUS scores by experience level, to see whether the same toolkit feels different to nurses at different career stages, and what that means for designing tools that actually fit their needs.

Method

Study Design

We used an explanatory sequential mixed-methods design[15,16]. Nurses completed the SUS first, then we interviewed them to understand what shaped their scores. To bring the two datasets together, we applied a connecting strategy[17] and displayed the results in a joint display matrix[18]—this allowed us to compare each group's quantitative ratings side by side with their qualitative accounts. We also split the sample at the median SUS score of 86 (high versus low) to help us identify extreme cases.

Participants and Setting

We recruited 26 nurses from the endocrinology ward between March and April 2026. The sample included 9 interns, 9 specialist nurses from other tertiary hospitals who were rotating through our ward, and 8 registered nurses from our own hospital. To be eligible, nurses had to be licensed (or interns), have participated in insulin injection teaching or related training within the past year, and have used the toolkit for at least 7 days before the study started. All participants gave written informed consent. The study was approved by the Ethics Committee of the First Affiliated Hospital of Sun Yat-sen University (Approval No. 2026-054), with full ethics documentation .

Our sample of 26 participants met the 15–20 threshold often cited for thematic saturation in qualitative work[19]. We grouped them based on Benner's framework[20,21] and the Chinese professional title system: interns ($n = 9$, < 6 months, novice/advanced beginner), junior nurses ($n = 7$, 1–7 years, competent), and senior nurses ($n = 10$, > 7 years, proficient/expert). This grouping enabled us to compare how nurses at different stages experienced the same toolkit.

We chose the endocrinology ward because the three groups have comparable responsibilities, which reduces the risk that differences in job roles or settings might confound our results. This real-world setup allowed us to observe how the same tool performs across different experience levels in a clinical context.

Data Collection Instruments

Our insulin injection toolkit is a portable, standardized teaching kit comprising three main components (see Fig. 1). The first component is a visual simulation pad with synthetic skin, which allows learners to practice injections and observe the correct needle angle. The second is a QR code panel linked to WeChat provides access to instructional videos for various insulin types (no personal data is collected via this platform). The third includes supporting materials: an injection site rotation card, a storage kit, and a structured teaching checklist. This checklist follows the Chinese Nursing Association standard (T/CNAS 21-2021) and covers seven steps: equipment preparation, insulin verification, site selection, disinfection, skin pinching and needle insertion, dwell time, and post-injection disposal (see Supplementary Material S1). The entire kit is arranged according to this workflow to ensure consistency across teaching sessions.

We used the Chinese version of the SUS [22,11] to evaluate the usability of this insulin injection toolkit. This version has been validated for Chinese-speaking populations [23]. The scale consists of 10 items rated on a 5-point Likert scale, with total scores ranging from 0 to 100. Scores above 80.3 [24] indicate good usability. The Chinese version has shown good reliability and validity, with Cronbach's α values of 0.81 in previous validation studies [25] and 0.80 in our sample. Although originally designed for digital interfaces, the SUS has since been widely applied to physical healthcare devices [26] and nursing education content [27].

We based our interview guide mainly on the Technology Acceptance Model (TAM) [28], which we used to frame why nurses might be willing to use the toolkit. We also drew on Hornbæk and Hertzum's review [29]. They argue that concrete experiences and affective responses matter in technology acceptance to sure our interviews explored how nurses actually felt when using the toolkit. To meet COREQ standards [30], we had two independent experts review the draft guide: one with a background in qualitative methods, the other in clinical nursing education. They suggested incorporating the critical incident technique [31]—an approach we chose to capture specific usability events, similar to its recent application in nursing contexts [32]. For instance, asking nurses to describe a moment when the toolkit worked particularly well for them, or a moment when it felt awkward or frustrating.

We organised the interview around five broad areas: how nurses moved from initial familiarity to proficiency with the toolkit; how they actually used each module in practice; whether the toolkit was useful in daily work; what advantages they saw compared with existing methods; and what they would want in future versions. Interviews lasted about 30 to 40 minutes each. We had a structured guide, but we didn't follow it rigidly—we adjusted questions and probes as the interview progressed, depending on the nurse's experience level and their responses. The full guide is in Supplementary Material S2.

Nurses used the toolkit for at least a week before the study—we confirmed this both by self-report and by checking with their supervisors. They then filled out the SUS, and we interviewed them afterwards. We collected and analysed data in parallel, and kept going until no new themes emerged from consecutive interviews [33,34]. All interviews were recorded and transcribed word for word."

Data Analysis

Quantitative data were analyzed using SPSS version 26.0. Normality was assessed using the Shapiro-Wilk test. Although overall SUS scores approximated a normal distribution overall, Shapiro-Wilk tests indicated that the subgroup distributions (especially in the intern and senior groups) deviated from normality—likely due to the small subgroup sample sizes ($n = 9, 7, \text{ and } 10$). To maintain consistency in how we handled all continuous variables and to avoid relying on normality assumptions that were not uniformly met, we used the Kruskal-Wallis H test for all three group comparisons (including age, years of experience, and SUS scores). Post-hoc pairwise comparisons were conducted using Dunn's test with Bonferroni correction. Effect sizes were reported as eta-squared (η^2) for the Kruskal-Wallis test and rank-biserial correlation (r) for post-hoc analyses. Overall SUS scores were compared with the benchmark value of 68 using the Wilcoxon signed-rank test. Statistical significance was defined as two-tailed $p < 0.05$.

To explore how SUS scores changed with experience, we conducted two supplementary analyses. We fitted a quadratic regression model. We also applied the Jonckheere–Terpstra test. All tests were two-tailed with $\alpha = 0.05$.

For the qualitative data, we used NVivo 12 and followed Braun and Clarke's framework[35]. We analysed as we went—recruiting and interviewing until three consecutive transcripts turned up nothing new. Two of us coded independently and settled at 85% agreement. We emailed the thematic summaries back to all 26 participants for member checking[36,37]; six got back to us. Five agreed with the themes, and one pointed out something we hadn't quite captured—so we went back and refined the framework. We also checked whether the non-responders looked any different from the responders on key demographics—they didn't. The codebook and GRAMMS checklist are in Supplementary Materials S3 and S4. We reported the study following COREQ[30] and GRAMMS[38].

Results

Participant baseline characteristics and SUS scores

Twenty-six nurses from the endocrinology ward were enrolled and stratified into three experience-based groups (interns, junior, senior; Table 1) according to Benner's framework and the Chinese professional title system. Age differed across groups [$H(2) = 18.43, p < 0.001, \eta^2_H = 0.71$]. Years of experience also showed a group difference [$H(2) = 21.80, p < 0.001, \eta^2_H = 0.86$]. SUS scores likewise varied by experience level [$H(2) = 6.94, p = 0.031, \eta^2_H = 0.23$]. Table 1 provides group-specific medians and interquartile ranges.

Table 1
Participant Characteristics and Comparison of SUS Scores

Variable	Nursing Intern Group (n = 9)	Junior Nurse Group (n = 7)	Senior Nurse Group (n = 10)	Statistic(H)	p-value
Age (years)	22.0 (21.0, 23.0)	29.0 (25.0, 30.0)	39.0 (35.0, 44.0)	H(2) = 18.43	< 0.001
Years of experience	0.4 (0.2, 0.6)	5.0 (5.0, 7.0)	15.0 (10.0, 20.0)	H(2) = 21.80	< 0.001
SUS score	75.00 (72.50–85.00)	92.50 (87.00–100.00)	87.50 (80.00–90.00)	H(2) = 6.94	0.031
Note: Values are median (Q1, Q3); all comparisons used Kruskal-Wallis H tests. Grouping based on Benner's framework and the Chinese professional title system: nursing interns (< 6 months), junior nurses (1–7 years), senior nurses (> 7 years).					

The overall median SUS score was 87.5 (IQR: 77.5–92.5). Table 2 shows the overall SUS score distribution, with 46.1% of participants scoring in the 'excellent' range (≥ 90). A Wilcoxon signed-rank test showed that the overall median SUS score (87.5) was significantly higher than the benchmark value of 68, a threshold indicating acceptable usability as established by Bangor et al. [24] ($V = 351$, $z = 4.22$, $p < 0.001$).

Table 2
Distribution of overall SUS scores (n = 26)

Score range	Frequency	Percentage (%)	SUS rating
90.0–100.0	12	46.1	Excellent
80.0–89.9	6	23.1	Good
70.0–79.9	8	30.8	Acceptable
Total	26	100.0	–
Note: SUS ratings are based on the classification standard of Bangor et al.[24]. The overall median SUS score was 87.5 (IQR: 77.5–92.5).			

Post-hoc Dunn's tests revealed that junior nurses scored significantly higher than interns (median difference = 17.50, 95% CI [2.00, 32.50], adjusted $p = 0.047$, $r = 0.50$), as shown in Fig. 2. No other pairwise comparisons reached statistical significance after Bonferroni correction (both $p > 0.05$).

In the quadratic polynomial regression analysis, the quadratic term for years of experience was not significant ($\beta_2 = -0.002$, 95% CI [-0.041, 0.037], SE = 0.019, $t(23) = -0.11$, $p = 0.924$). The quadratic model did not increase the explanatory power compared to the linear model ($\Delta R^2 < 0.001$, $F(1, 23) = 0.01$, $p = 0.924$; linear model $R^2 = 0.144$, quadratic model $R^2 = 0.145$). Additionally, the Jonckheere-Terpstra test indicated no significant monotonic trend in SUS scores across the three ordered experience groups ($J = 95.5$, $Z = -0.38$, $p = 0.707$). In summary, the observed descriptive pattern across the three groups did not

reach statistical confirmation as a curvilinear or monotonic trend. Thus, the descriptive differences across groups should be interpreted as exploratory rather than definitive.

Individual Score Variability and Identification of Extreme Cases

We split the sample at the median SUS score of 86—high versus low. That gave us 13 in the high group (6 junior nurses, 5 seniors, and 2 interns) and 13 in the low group (7 interns, 5 seniors, and just 1 junior). So junior nurses made up most of the high group, and interns most of the low group. Seniors? They were right down the middle: five and five.

Two cases stood out (Fig. 3). One was N4, an intern who scored a perfect 100—the highest among all interns. The other was N6, a senior nurse who scored 75, the lowest in her group. A full summary of all critical incidents is in Supplementary Material S5. The full SUS raw data, including individual scores for all participants, are provided in Supplementary Material S6.

Qualitative Thematic Differences Between Groups

Thematic differences between groups are summarized in Table 3.

Table 3
Comparison of Usability Perception Themes Between High- and Low-Scoring Groups ($n = 26$)

Dimension	High-scoring group ($SUS \geq 86$ $n = 13$)	Low-scoring group ($SUS < 86$, $n = 13$)
Central theme	Toolbox as helpful scaffolding	Toolbox as confusing and cumbersome
Key focus	Preventing errors & supporting daily workflow	Feeling overwhelmed by excessive information
Common descriptors used ^a	Convenient, comprehensive, easy to follow, systematic, felt reassuring	Too much info, messy layout, time-consuming, needed help, not intuitive enough
Representative quotes	“It’s very convenient; everything is right there.” (N12, junior nurse, 100)	“You need the instructor to guide you through it first before you can use it.” (N18, nursing intern, 72.5)
	The functions are very comprehensive; a single educational toolbox suffices for the entire teaching session.” (N26, junior nurse, 92.5)	There are too many QR codes; it takes time to figure them out.” (N6, senior nurse, 75)
Note: Participants were dichotomized into high- and low-scoring groups based on the sample median SUS score (86). The sample median was 86 (calculated as the average of the 13th and 14th values in the ordered dataset); no participant scored exactly 86, and the split resulted in equal groups of 13 each. High-frequency descriptors (word frequency ≥ 3) were extracted from interview transcripts and validated through independent coding by two researchers, with discrepancies resolved via consensus discussion. A structured codebook was used throughout the analysis to ensure coding reliability. N denotes anonymized participant ID; clinical role and raw SUS score are shown in parentheses. “Core evaluative tendency” refers to participants’ overall stance toward the toolbox; “Key focus” refers to their main cognitive concern during use. SUS: System Usability Scale.		
^a High-frequency descriptors were identified based on word frequency ≥ 3 across interview transcripts.		
N denotes anonymized participant ID; clinical role and raw SUS score are shown in parentheses.		

Nurses with higher scores appreciated the injection kit for its clear instructions and error-prevention features, describing it as a supportive tool that kept them organized and prevented omissions. One intern commented, “You just follow the steps, so there is no need to worry about missing anything” (N4, intern, $SUS = 100$). A junior nurse added, “Everything is neatly categorized, making it very convenient to access” (N26, junior nurse, $SUS = 92.5$).

Those with lower scores found it confusing and cumbersome. One senior nurse noted, “There are too many QR codes, which disrupts the instructional pace” (N6, senior nurse, $SUS = 75$). An intern (N18, $SUS = 70$) described a different barrier: “You can not just pick it up and use it alone—you need someone to show you how first.

Mixed-Methods Integration

The median SUS scores for the three groups were 75.0 for interns, 92.5 for junior nurses, and 87.5 for senior nurses. Formal trend tests did not confirm this pattern as a statistically significant curvilinear or monotonic trend ($p > 0.05$). Table 4 presents each group’s median SUS score alongside their key interview themes.

Table 4
Joint Display Matrix of Quantitative Trajectories and Qualitative Mechanisms

Participant Group & SUS Profile	Qualitative Themes & Quotes	Meta-Inference	Clinical Design Implications
Junior Nurses (n = 7) (1–7 years; median 5.0) SUS = 92.5 (87.0–100.0)	Seamless workflow integration “You just take it and teach the patient directly” (N26, 92.5)	Seems to fit their stage—they can use it without stopping to think.	Keep it fast, keep it simple. They don’t need the hand-holding.
Nursing Interns (n = 9) (Internship < 6 months) SUS = 75.0 (72.5–85.0)	Structured scaffolding reliance “checklists and videos keep them on track” (N4, 100)	Their scores are lower, but that makes sense—they’re still learning the basics. Without clear guidance, they get lost.	Start simple. Show them one step at a time. Pair it with instructor demos
Senior Nurses (n = 10) (> 7 years experience) SUS = 87.5 (80.0–90.0)	Redundancy and efficiency loss “too much stuff in the way”(N6, 75)	They like that it’s all there—but when it’s too cluttered, it slows them down.	Offer streamlined paths, show only core parameters, and let users filter what they see.
Extreme case comparison N4 (intern, 100) vs. N6 (senior nurse, 75)	N4: “Very comprehensive and easy to use.” N6: “There’s so much content; it takes a lot of time to get familiar with the box first.”	One person’s “comprehensive” is another person’s “overwhelming.” How the toolbox is perceived depends on what the user brings to it.	Give users a choice —‘novice mode’ or ‘expert mode.’ Let them set the level.
Note: SUS scores are presented as median (IQR). N denotes anonymised participant codes. QR = Quick Response code. Quotes are annotated with clinical role and individual SUS score. Extreme cases (N4: highest overall SUS; N6: lowest within senior group) were selected to illustrate maximum variation in user experience. The joint display framework is adapted from Fetter, Curry, and Creswell (2013)[47]. This joint display is exploratory in nature and intended to illustrate group-level descriptive patterns and hypothesis generation, rather than to confirm statistical associations.			

Interns reported difficulty using the toolkit independently, noting that it required initial guidance. Junior nurses, by contrast, described it as easy and intuitive. Some senior nurses found that the toolkit contained too much information and disrupted their workflow.

For junior nurses, the highest scores were accompanied by the comment that the toolkit was “a complete system—I can just take it and directly educate the patient, very convenient” (N26, junior nurse, SUS = 92.5). For senior nurses, lower scores reported concerns like “there are QR codes and various things pasted all around; it feels like there’s too much content” (N6, senior nurse, SUS = 75).

Three Identified Usability Gaps

By cross-referencing lower SUS scores with nurses’ interview data, we identified three main usability gaps: that is, discrepancies between the intended design and actual user experience.

The first gap identified was the lack of guidance for beginners, which corresponds to SUS item 4 (“I think that I would need the support of a technical person to be able to use this system”). Interns lacked introductory guidance for independent use: as N18 (nursing intern, SUS = 70) stated, the toolkit “requires instructor guidance before independent use.” The current teaching checklist (c) and QR code panel (f) (Fig. 1) do not differentiate content by user experience level, nor do they provide a simplified introduction for first-time users.

The second gap involved visual clutter (SUS item 6). Participants described the layout as visually overwhelming, and some noted that it took longer to find what they needed during clinical procedures. N6 (senior nurse, SUS = 75) described the layout as having ‘too many QR codes and materials stuck everywhere—it’s visually overwhelming.’ As shown in Fig. 1, the teaching checklist (c), injection site illustration (e), and QR code panel (f) are co-located without color coding or physical separation. Participants associated this arrangement with longer visual search times during clinical procedures.

The third gap was the compact and disorganized internal layout, corresponding to SUS item 8 (“I found the system very cumbersome to use”). N6 (senior nurse, SUS = 75) described it as “a bit too small,” noting that it felt crowded once all items were in place. N22 (junior nurse, SUS = 75) similarly found the variety of pen types made the space feel cluttered. More granular issues included compartment labels that were difficult to locate (N2, senior nurse, SUS = 80) and a needle storage slot regarded as too narrow for convenient access (N26, junior nurse, SUS = 92.5). Nurses reported that the most prominent operational inconvenience occurred when multiple items needed to be retrieved simultaneously: especially when two or three components were taken out at the same time, and that frequent searching and mutual interference significantly increased the operation time.

Discussion

Main findings

The central finding of this study is that aggregated SUS scores can mask real differences across experience levels. In descriptive terms, the three groups’ median SUS scores showed a ‘rise then slight decline’ pattern (75.0 for interns, 92.5 for junior nurses, and 87.5 for senior nurses), although trend tests did not reach statistical significance ($p > 0.05$)—likely a reflection of the small sample size ($n = 9, 7$, and

10 per group). What stood out more than this group-level pattern was the split within the senior nurse group: five scored high, five scored low. That tells us the median of 87.5 doesn't represent a single, shared experience. The contrast was most clearly seen in two extreme cases: N4, an intern, scored the toolkit 100, while N6, a senior nurse with 13 years of experience, gave it only 75—a 25-point gap under identical conditions. To unpack this variability, Table 4 presents a joint display matrix that integrates these quantitative scores with qualitative themes, revealing systematic experiential bottlenecks that aggregate means would otherwise obscure.

Cognitive Stages and Perceived Usability

The SUS score difference between interns and junior nurses can be understood through Benner's stages of clinical competence[20]. Interns are at the "novice" stage. They have not yet developed an automatic mental script for insulin injection teaching. Each step requires deliberate thought and integration. The toolkit presents the full standardized procedure, but for a novice, that means high intrinsic cognitive load[39]. This is why they need instructor guidance to reduce that load before they can use the toolkit independently. One intern commented: "You can't just pick it up and use it alone—you need someone to show you how first" (N18, intern, SUS = 70). Junior nurses, by contrast, have mastered the procedure but are not yet fully automated. They remain dependent on checklists and video prompts. For them, the toolkit offers just the right level of cognitive support: not too simple, not too dense. One junior nurse stated: "Everything is neatly categorized, making it very convenient to access" (N26, junior nurse, SUS = 92.5). What this suggests is that usability is determined not by years of experience per se, but by the fit between the tool's features and the user's cognitive stage. That fit is precisely what is lost when one looks only at the aggregated SUS mean.

These experiential differences were also reflected in the quantitative data, though the pattern was not straightforward. The 17.5-point gap between interns and junior nurses was accompanied by a moderate effect size ($r = 0.50$), but it did not retain statistical significance after Bonferroni correction. With only 26 participants, however, the effect size may be more informative than the p-value in suggesting where to direct attention—it points toward the qualitative accounts and individual cases, rather than confirming a stable group-level difference. What the convergence suggests is not that the difference is definitively real, but that the aggregated mean alone does not tell the full story—a signal worth pursuing in larger samples, not a conclusion about between-group significance.

Expertise Reversal Effect and Within-Group Heterogeneity

The even split of senior nurses' SUS scores between the high- and low-scoring groups can be explained by a combination of Benner's expert stage characteristics and individual differences. Nurses at the expert stage, whether administering insulin injections to patients or teaching patients to self-inject, have already developed stable, internalized operational procedures: actions and verbal explanations flow together seamlessly, without the need for conscious step-by-step thinking. However, when using the toolkit, the mandatory QR code retrieval step requires users to select the matching content from multiple videos, which disrupts their smooth workflow. This burden is not the kind of cognitive load seen in

novices who "do not know what to do next," but rather the specific burden characteristic of experts when automated processes are interrupted by redundant prompts. This is consistent with the predictions of the Expertise Reversal Effect (ERE)[40]: when operations are highly automated, additional explicit prompts increase rather than reduce cognitive load. As N6 noted: "There are too many QR codes, which disrupts the instructional pace" (N6, senior nurse, SUS = 75).

However, the ERE is not a universal experience among all senior nurses. Table 4 also contains contrasting feedback from a senior nurse in the high-scoring group: "Very comprehensive—I can just take it and teach" (N5, senior nurse, SUS = 100). This suggests that even at the same experience level, different individuals respond differently to the same information density. Individual variables such as cognitive style and tolerance for structured materials[41] may be more predictive of this differentiation than years of experience alone.

This variation also highlights a limitation of the framework we employed to group nurses. Benner's model[20], while useful, was developed from qualitative narratives rather than as a quantifiable empirical tool. Grouping nurses by title or years of experience into discrete stages is a practical approximation. It may not fully capture the continuous nature of individual cognitive development[42]. This further supports the idea that tool design should not assume all senior nurses prefer the same information density. We therefore propose an adaptive direction: novices receive structured guidance by default, while experts see a streamlined view, with both able to adjust details as needed (see Supplementary S7). That is, 'senior nurses' are not a single user type.

Comparison with literature

Freeman et al[43] found a mismatch between SUS scores and actual skill performance in a non-nurse population, which is consistent with what we observed in our study. The populations were different, but both studies point in the same direction, suggesting that the person-tool mismatch might not be unique to a specific group. What our study adds is that we used qualitative data to show why that mismatch happens, and we traced it down to three specific usability gaps: lack of guidance for beginners, visual information overload, and physical layout that affects workflow fluency, all of which are addressable.

Kim[13] found that novice nurses rated their emergency information system lower than expert nurses did, the direction of that finding aligns with ours. Where our study adds something new is this: we found that junior nurses, not experts, gave the highest SUS scores for the injection kit. And the senior nurses? Their scores went both ways: some high, some low. That suggests that "senior nurse" is not one single type of user. Their experiences diverge, and we need to pay attention to that split.

Strengths and limitations

Our findings suggest that aggregate SUS means may mask between-group differences. The single-ward design, with a standardized toolkit and patient population, reduced contextual variability across groups, though this inherently limits generalizability. The mixed-methods design enabled us to capture experiential differences that SUS alone did not reflect.

Several limitations should be considered. With only 26 participants, we were underpowered to detect small-to-moderate curvilinear effects. Therefore the absence of statistical confirmation does not rule out a meaningful clinical pattern. Rather, it reflects the exploratory scope of this study. Whether the observed rise-then-slight-decline pattern is stable will need to be tested in larger samples. The single-center design limits generalizability, despite including nine nurses from other hospitals. Years of experience and professional title were tightly linked ($\eta^2 = 0.86$), making it hard to separate their individual contributions. SUS might not be sensitive enough to capture the subtle disruptions that senior nurses experienced during automated tasks [24,44]. We relied on self-reports and preceptor verification, and only a small number of participants completed the member check. Colleagues may have held back criticism out of interpersonal considerations, potentially introducing response bias[45]. Recall bias may also have affected retrospective accounts of toolkit use[46].

Despite these limitations, our findings point to one practical direction. Future iterations of the toolkit may need to consider different user pathways for novice and expert users. Whether such adaptations actually improve usability would need to be tested in future studies. Drawing on participants' accounts of cognitive load and workflow disruptions, we outline experience-stratified design directions for future iterations of the toolkit (see Supplementary S7). These are conceptual only and have not been validated in the current physical toolkit.

Conclusions

We found that the aggregated SUS mean masked the differences in usability across experience levels. In future SUS evaluations, stratified reporting by experience level could be considered to reveal such between-group differences. This suggestion is grounded in the descriptive patterns observed in our sample and warrants further investigation in larger, more diverse populations. Our mixed-methods integration (Table 4) further revealed 'person-tool fit' as a core mechanism underlying usability evaluation. The descriptive trend observed in this study has not been statistically confirmed and should therefore be treated as hypothesis-generating in nature, requiring validation in larger samples. Tool design should take users' experience levels into account, providing different use pathways for novice and expert users.

Declarations

Ethics approval and consent to participate

This study was conducted in accordance with the ethical standards of the Declaration of Helsinki. Ethical approval was obtained from the Ethics Committee of the First Affiliated Hospital of Sun Yat-sen University (Approval No. 2026-054) prior to the commencement of data collection. Written informed consent was obtained from all participants before their inclusion in the study. Participants were informed of the purpose of the study, their right to refuse participation, and their freedom to withdraw at

any time without consequence. To ensure confidentiality, all data were anonymised and stored securely, with access restricted to the research team.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and/or analysed during the current study are not publicly available due to the need to protect participant anonymity and confidentiality, as the qualitative data contain potentially identifiable information. However, anonymised data may be available from the corresponding author upon reasonable request, subject to ethical approval and institutional data access policies.

Competing interests

The authors declare that they have no competing interests

Funding

Not applicable.

Acknowledgements

During the preparation of this manuscript, the authors used DeepSeek for language editing and grammar refinement. The authors reviewed and edited all AI-assisted content and take full responsibility for the accuracy and integrity of the publication.

Author Contribution

X.N. conceived the study, designed the research protocol, and drafted the manuscript. L.Y. and L.H. conducted the investigation and data curation, and contributed to manuscript revision. W.Q. and H.R. performed the formal analysis and validation, and contributed to manuscript revision. W.J. provided supervision and resources, and contributed to manuscript revision. All authors read and approved the final manuscript.

References

1. Martin L, Holt RI, Hill S. The effect of diabetes management on the workload of district and community nursing teams in the UK. *Diabet Med*. 2024;41(10):e15367. doi:10.1111/dme.15367.
2. Baker KM, Magee MF, Smith KM. Understanding Nursing Workflow for Inpatient Education Delivery: Time and Motion Study. *JMIR Nurs*. 2019;2(1):e15658. doi:10.2196/15658.

3. Wei ET, Koh E, Kelly MS, Wright LA, Tylee TS. Patient errors in use of injectable antidiabetic medications: A need for improved clinic-based education. *J Clin Transl Endocrinol*. 2020;20:100224. doi:10.1016/j.jcte.2020.100224.
4. Altunbaş R. Fatal Insulin Overdose in an Illiterate Elderly Patient: A Case Emphasizing the Imperative of Confirmatory Patient Education. *Turkiye Klinikleri J Case Rep*. 2024;32(3):77-80. doi:10.5336/caserep.2024-102384.
5. Klonoff DC, Berard L, Franco DR, et al. Advance Insulin Injection Technique and Education With FITTER Forward Expert Recommendations. *Mayo Clin Proc*. 2025;100(4):682-699. doi:10.1016/j.mayocp.2025.01.004.
6. Yoshida Y, Kikuchi M, Nakamura N, Tomisawa T. Predictors of limited knowledge in insulin injection techniques among non-specialty nurses in Japan. *SAGE Open Med*. 2025;13:20503121251367182. doi:10.1177/20503121251367182.
7. Liao HC, Liang WM, Chu CL, Huang LC. A New Learning Approach to Improve Nurses' Knowledge of and Skills in Diabetes Care. *J Contin Educ Nurs*. 2022;53(2):90-96. doi:10.3928/00220124-20220104-10.
8. Punjot P, Thomas V, Vinodkumar S, Setia MS. A study to assess the effectiveness of a structured teaching programme on knowledge and practice of safe insulin administration among nurses in a tertiary care hospital: a pre-post design. *J Nurs Educ Pract*. 2019;10:72. doi:10.5430/jnep.v10n3p72.
9. Veronesi G, Poerio CS, Braus A, et al. Determinants of nurse satisfaction using insulin pen devices with safety needles: an exploratory factor analysis. *Clin Diabetes Endocrinol*. 2015;1:15. doi:10.1186/s40842-015-0015-3.
10. Shogbon AO, Ngo D, Jacob B, Kimble LP, Ryan G. Nurses' Perceptions and Satisfaction with the Use of Insulin Pen Devices Compared with Insulin Vial and Syringes in an Inpatient Setting. *Diabetes Technol Ther*. 2014;16(11):742-746. doi:10.1089/dia.2014.0072.
11. Brooke J. SUS: a "quick and dirty" usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. *Usability evaluation in industry*. London: Taylor & Francis; 1996. p. 189-194.
12. Marquard JL, Howard J, LeBlanc R. Using Novel Data Visualization Methods to Understand Mobile Health Usability: Exemplar From a Technology-Enabled Sleep Self-monitoring Intervention. *Comput Inform Nurs*. 2023;41(1):1-5. doi:10.1097/CIN.0000000000000970.
13. Kim MS, Shapiro J, Borycki E, Kushniruk A. A pilot study on usability analysis of emergency department information system by nurses. *Appl Clin Inform*. 2012;3(1):135-153. doi:10.4338/ACI-2011-11-RA-0065.
14. Liang W, Chow KM, Ni X, Tola YO, Lo SHS. Effects of a nurse-led diabetes self-management programme for people with type 2 diabetes receiving insulin injection therapy: A randomised controlled trial with qualitative process evaluation. *Int J Nurs Stud*. 2025;171:105200. doi:10.1016/j.ijnurstu.2025.105200.
15. Creswell JW, Plano Clark VL. *Designing and conducting mixed methods research*. 4th ed. Thousand Oaks (CA): SAGE Publications; 2026.

16. Romem A. From two strands to one defensible conclusion: a practical framework for integration in mixed methods nursing research. *SAGE Open Nurs.* 2026;12:1. doi:10.1177/23779608261452444.
17. Younas A, Durante A. Decision tree for identifying pertinent integration procedures and joint displays in mixed methods research. *J Adv Nurs.* 2023;79(7):2754-2769. doi:10.1111/jan.15536.
18. Guetterman TC, Fetters MD, Creswell JW. Integrating quantitative and qualitative results in health science mixed methods research through joint displays. *Ann Fam Med.* 2015;13(6):554-561. doi:10.1370/afm.1865.
19. Guest G, Bunce A, Johnson L. How many interviews are enough? An experiment with data saturation and variability. *Field Methods.* 2006;18(1):59-82. doi:10.1177/1525822X05279903.
20. Benner P. From novice to expert: excellence and power in clinical nursing practice. Menlo Park (CA): Addison-Wesley; 1984.
21. Benner P. Using the Dreyfus model of skill acquisition to describe and interpret skill acquisition and clinical judgment in nursing practice and education. *Bull Sci Technol Soc.* 2004;24(3):188-199. doi:10.1177/0270467604265061.
22. Sheu FR, Fu HJ, Shih M. Pre-testing the Chinese version of the System Usability Scale (C-SUS). In: Hayashi Y, et al., editors. Workshop Proceedings of the 25th International Conference on Computers in Education. New Zealand: Asia-Pacific Society for Computers in Education; 2017. p. 1-7.
23. Gao M, Kortum P, Oswald FL. Multi-Language Toolkit for the System Usability Scale. *Int J Hum Comput Interact.* 2020;36(20):1883-1901. doi:10.1080/10447318.2020.1801173.
24. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. *J Usability Stud.* 2009;4(3):114-123.
25. Wang Y, Lei T, Liu X. Chinese System Usability Scale: translation, revision, psychological measurement. *Int J Hum Comput Interact.* 2020;36(10):953-963. doi:10.1080/10447318.2019.1700644.
26. Kortum P, Peres SC. Evaluation of home health care devices: remote usability assessment. *JMIR Hum Factors.* 2015;2(1):e10. doi:10.2196/humanfactors.4570.
27. Kim HY. Development and Usability Assessment of Virtual Reality- and Haptic Technology-Based Educational Content for Perioperative Nursing Education. *Healthcare (Basel).* 2024;12(19):1947. doi:10.3390/healthcare12191947.
28. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q.* 1989;13(3):319-340. doi:10.2307/249008.
29. Hornbæk K, Hertzum M. Technology acceptance and user experience: a review of the experiential component in HCI. *ACM Trans Comput Hum Interact.* 2017;24(5):33. doi:10.1145/3127358.
30. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care.* 2007;19(6):349-357. doi:10.1093/intqhc/mzm042.
31. Flanagan JC. The critical incident technique. *Psychol Bull.* 1954;51(4):327-358.

32. Hedqvist AT, Holmberg M, Bjurling-Sjöberg P, Ekstedt M. Bracing for the next wave: a critical incident study of frontline decision-making, adaptation and learning in ambulance care during COVID-19. *J Adv Nurs*. 2025;81(9):5442-5457. doi:10.1111/jan.16340.
33. Braun V, Clarke V. To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qual Res Sport Exerc Health*. 2021;13(2):201-216. doi:10.1080/2159676X.2019.1704846.
34. Hennink MM, Kaiser BN, Weber MB. What influences saturation? Estimating sample sizes for qualitative research. *Qual Health Res*. 2019;29(10):1483-1496. doi:10.1177/1049732319864681.
35. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol*. 2006;3(2):77-101. doi:10.1191/1478088706qp063oa.
36. Lincoln YS, Guba EG. *Naturalistic inquiry*. Beverly Hills (CA): Sage Publications; 1985.
37. Urry K, Chur-Hansen A, Scholz B. From member checking to collaborative reflection: a novel way to use a familiar method for engaging participants in qualitative research. *Qual Res Psychol*. 2024;21(3):357-374. doi:10.1080/14780887.2024.2355972.
38. O’Cathain A, Murphy E, Nicholl J. The quality of mixed methods studies in health services research. *J Health Serv Res Policy*. 2008;13(2):92-98. doi:10.1258/jhsrp.2007.007074.
39. Sweller J. Cognitive load during problem solving: Effects on learning. *Cogn Sci*. 1988;12(2):257-285.
40. Kalyuga S, Ayres P, Chandler P, Sweller J. The expertise reversal effect. *Educ Psychol*. 2003;38(1):23-31. doi:10.1207/S15326985EP3801_4.
41. Agarwal R, Karahanna E. Time flies when you’re having fun: cognitive absorption and beliefs about information technology usage. *MIS Q*. 2000;24(4):665-694. doi:10.2307/3250951.
42. Gobet F, Chassy P. Towards an alternative to Benner’s theory of expert intuition in nursing: a discussion paper. *Int J Nurs Stud*. 2008;45(1):129-139. doi:10.1016/j.ijnurstu.2007.01.005.
43. Freeman S, Lecka L, Davis K, Prince G, Carthy K, Seley JJ, et al. Leveraging user-centered design and usability and skills testing for a novel diabetes survival skills toolkit. *J Endocr Soc*. 2025;9(9):bvaf111. doi:10.1210/jendso/bvaf111.
44. Kortum P, Bangor A. Usability ratings for everyday products measured with the System Usability Scale. *Int J Hum Comput Interact*. 2013;29(2):67-76. doi:10.1080/10447318.2012.681221.
45. van de Mortel TF. Faking it: social desirability response bias in self-report research. *Aust J Adv Nurs*. 2008;25(4):40-48. doi:10.37464/2008.254.1817.
46. Hunger M, Schwarzkopf L, Heier M, Peters A, Holle R; KORA Study Group. Official statistics and claims data records indicate non-response and recall bias within survey-based estimates of health care utilization in the older population. *BMC Health Serv Res*. 2013;13:1. doi:10.1186/1472-6963-13-1.
47. Fetzters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs—principles and practices. *Health Serv Res*. 2013;48(6 Pt 2):2134-2156. doi:10.1111/1475-6773.12117.

Figures

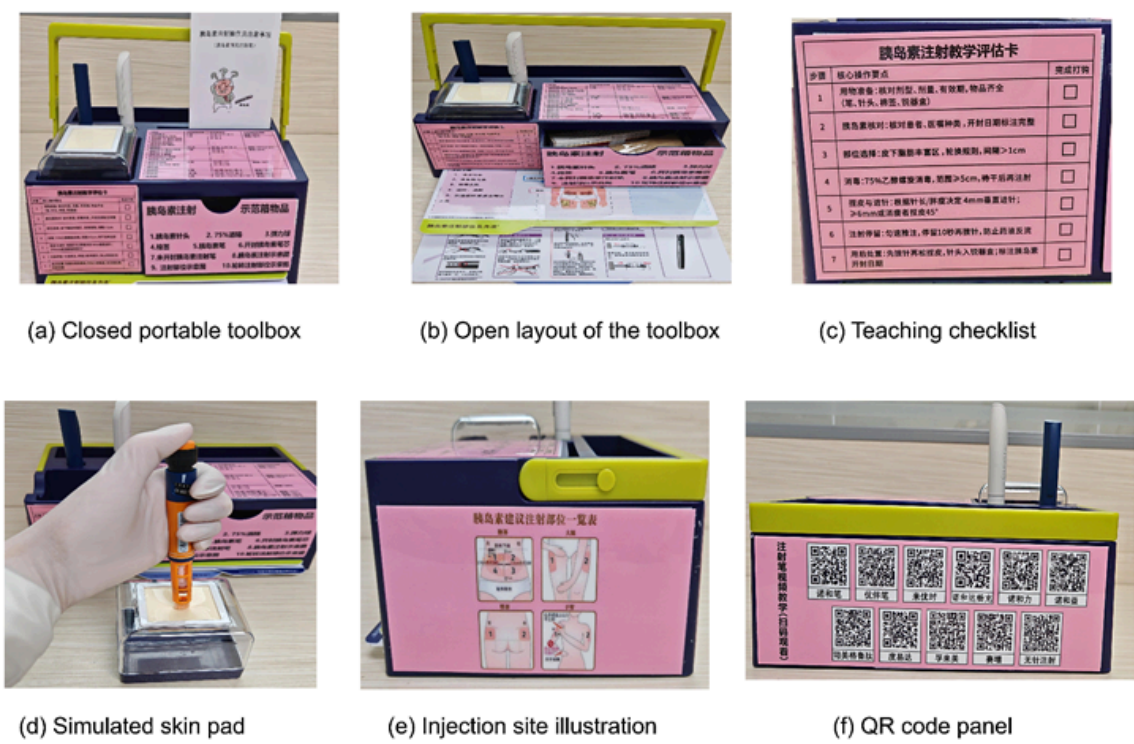


Figure 1

Insulin injection educational toolkit

Note: (a) Closed portable toolkit; (b) Open layout showing internal modules; (c) Teaching checklist (seven-step procedure); (d) Simulated skin pad for injection practice; (e) Injection site illustration for rotation guidance; (f) QR code panel linking to instructional videos.

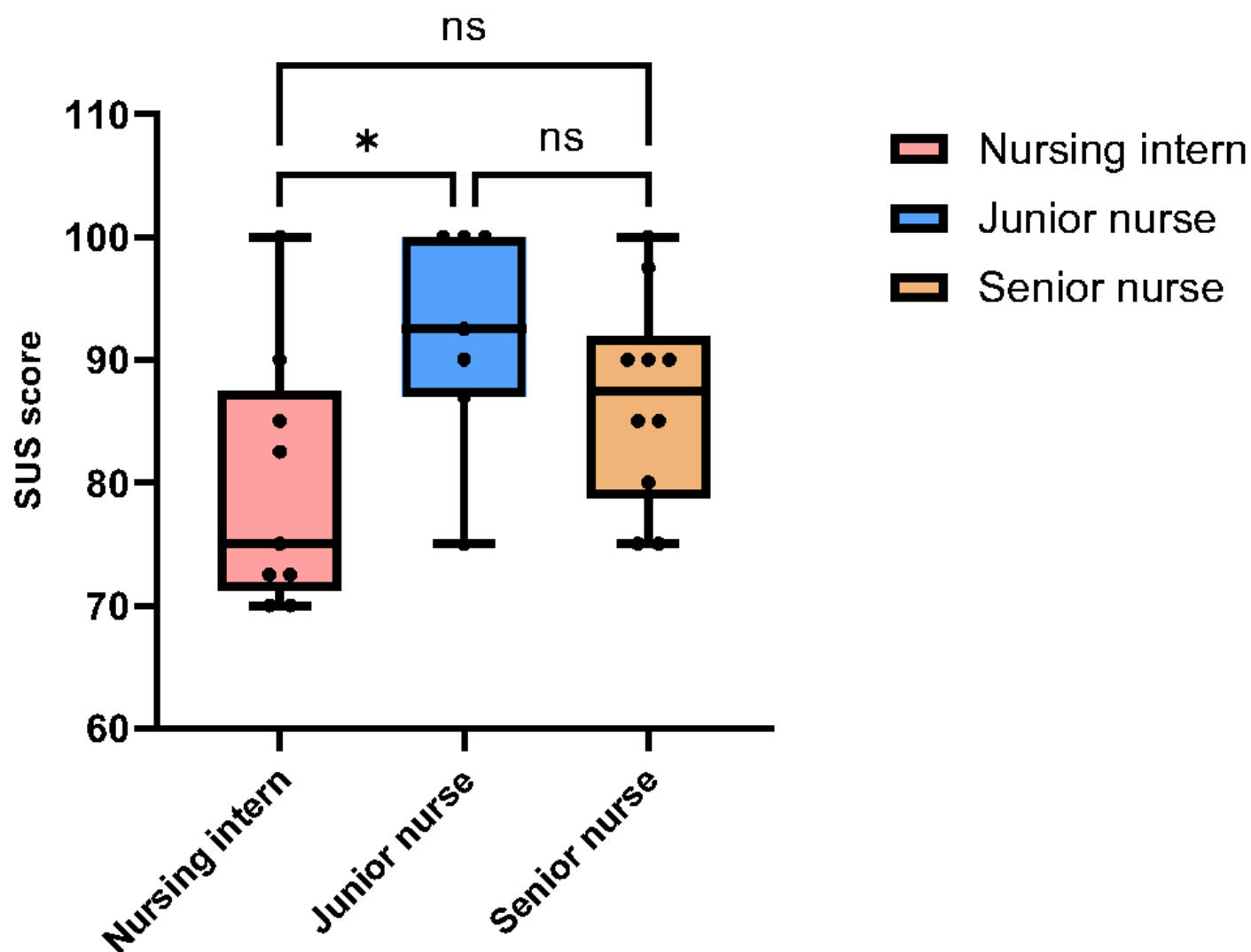


Figure 2

Distribution of System Usability Scale scores by clinical experience level

Note: Box plots show the median (horizontal line), interquartile range (box), and min-max range (whiskers). Individual data points are jittered for clarity. Kruskal-Wallis H test revealed a statistically significant difference between groups ($H(2) = 6.94, p = 0.031, \eta^2_H = 0.23$). Bonferroni post hoc comparisons showed that junior nurses scored significantly higher than nursing interns ($p = 0.047$). ns indicates non-significant difference ($p > 0.05$). SUS, System Usability Scale.

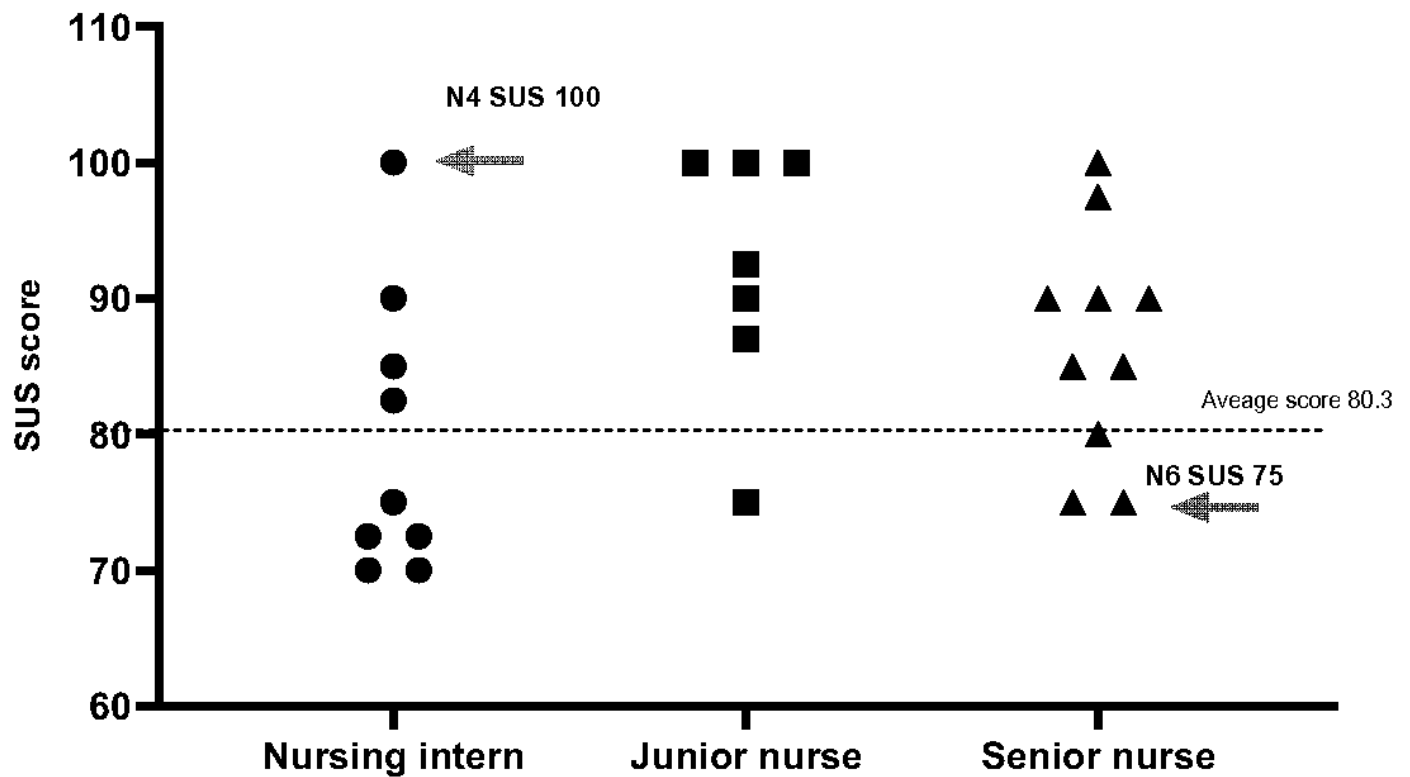


Figure 3

Distribution of individual SUS scores across professional title groups (from left to right: nursing interns, junior nurses, senior nurses)

Note: Arrows indicate the counterintuitive extreme cases: nursing intern N4 (100 points) and senior nurse N6 (75 points). The horizontal dashed line indicates the 'good usability' threshold (>80.3 points; Bangor et al., 2009)[24]. SUS, System Usability Scale.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterialS1S7.docx](#)