Development of a program to determine optimal settings for robotic-assisted rehabilitation of the post-stroke paretic upper extremity: a simulation study

Takashi Takebayashi (✉ takshi77@omu.ac.jp)  
Osaka Metropolitan University

Yuki Uchiyama  
Hyogo College of Medicine

Yuho Okita  
Swinburne University of Technology

Kazuhisa Domen  
Hyogo College of Medicine

---

Research Article

**Keywords:** Occupational therapy, robotics, stroke, upper extremity, robotics, rehabilitation

**Posted Date:** June 8th, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-1713883/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

**Background:** Robot-assisted therapy can effectively treat upper extremity (UE) paralysis in patients who experience stroke. Presently, UE, as a training item, is selected according to the severity of the paralysis, which is based on a clinician's experience. The possibility of objectively selecting robot-assisted training items based on the severity of paralysis was simulated using the item response theory (IRT).

**Methods:** Sample data were generated using Monte Carlo method with 300 random cases. In this simulation, sample data (categorical data with three difficulty values of 0, 1, and 2 [0: too easy, 1: adequate, and 2: too difficult]) with 71 items per case were analyzed. First, the most appropriate method was selected to ensure local independence of the sample data necessary to use IRT. The method was to exclude: items with low response probability (maximum response probability) within a pair in the Quality of Compensatory Movement Score (QCM) 1-point item difficulty curve, items with low item information content within a pair in the QCM 1-point item difficulty curve, and items with low item discrimination. Second, 300 cases were analyzed to determine the most appropriate model (one-parameter or two-parameter item response therapy) to be used and the most favored method to establish local independence. We also examined whether robotic training items could be selected according to the severity of paralysis based on the ability of the person (θ) in the sample data as calculated by IRT.

**Results:** To ensure local independence, excluding items with low response probability (maximum response probability) in a pair in the categorical data 1-point item difficulty curve was effective. Additionally, to ensure local independence, the number of items should be reduced to 61 from 71, indicating that the two-parameter model IRT was an appropriate model. The ability of the person (θ) calculated by IRT suggested that seven training items could be estimated from 300 cases according to severity.

**Conclusion:** From this simulation, it seemed possible to objectively estimate the training items according to the severity of paralysis in a sample of approximately 300 cases using this model.

**Background**

Globally, stroke is a leading cause of physical impairment, with approximately 65% of patients experiencing ongoing post-stroke long-term upper extremity (UE) impairment [1]. Despite some patients regaining independence in daily activities within a few months of onset, the majority tend to rely on using only their non-paralyzed hand in daily life [2]. However, some researchers have highlighted the importance of facilitating paralyzed hand use in daily activities to address concerns regarding decreased quality of life caused by difficulty using the UE [3].

On other hand, several researchers reported that appropriate rehabilitation approaches could help improve UE function even in the chronic phase (more than six months after stroke onset) in patients with stroke [4, 5]. Several approaches to post-stroke UE paralysis rehabilitation are recommended in the current stroke guidelines [6]. Robotic therapy is one of the recommended approaches that allow patients with moderate-to-severe UE paralysis to practice their hand use. This approach can produce outcomes similar to those of conventional rehabilitation programs [7]. Several studies have used robotic approach as adjuvant therapy to one-to-one rehabilitation sessions with a physical therapist. We previously conducted a randomized controlled trial (RCT)
to investigate the effectiveness of the robotic approach as adjuvant therapy in the subacute and post-stroke phases and found a significant difference in the intervention group [8, 9]. However, an RCT from another research group reported no significant UE improvement in patients who had experienced stroke in a subacute setting when compared to either a self-training robot or conventional self-training [10]. Therefore, the effectiveness of robotic therapy as a form of self-training has not been thoroughly clarified. The factors associated with the reported differences in robotic treatment effectiveness in previous studies also remain unclear.

Previous studies have found that several factors are involved in the effectiveness of robot-assisted rehabilitation. It has been suggested that the amount of assistance generated by a robot can affect the intervention outcome [11]. More specifically, when the stroke survivors less actively generate voluntary movement of their paralyzed hand by being given excessive robotic assistance, this could result in a so-called ‘Slack’ state [11]. The study concluded that the group that received no robotic assistance showed more significant improvement in their UE than the group that received excessive robotic assistance [11]. Previous studies have reported that minimal robotic assistance can promote post-stroke rehabilitation in patients with moderate-to-severe UE paralysis [12]. Furthermore, some studies have suggested that robotic assistance should be adjusted according to the severity of the stroke survivor’s affected extremity. Specifically, stroke survivors with severe UE paralysis are encouraged to receive increased robotic assistance, while it is suggested that those with moderate-to-light UE paralysis receive less robotic assistance [13, 14]. These results suggest that delivering a more optimized intervention with various adjusted parameters, including the ideal amount of robotic assistance, could be key to generating a positive outcome. Additionally, previous research investigated two ways of promoting recovery from post-stroke motor paralysis: facilitating the use of voluntary movements of the paralyzed hand and minimizing compensatory movements, including the trunk and other body movements associated with paralyzed hand movement [15]. In clinical practice, the optimal setting for robotic assistance often depends on the therapist’s experience and ability to assess the degree of compensatory movements using the trunk rather than the paralyzed hands. Thus, the optimal setting can vary for each therapist.

This study was planned as an exploratory study to determine whether optimal setting could be selected for different patients with different severities of hand paralysis. The study followed the following steps: 1) each patient tried all the severity-specific settings created by combining the parameters of the robot; 2) the therapist evaluated the voluntary motor output and the degree of compensatory movement as each patient performed all settings and determined the appropriate settings for severity, 3) the item response theory (IRT) was applied to estimate the applicability of the post-stroke UE paralysis to the ability of person (θ) for all settings, and 4) to estimate settings according to the severity of each patient’s paralyzed hand using the ability of person (θ). Using simulation data generated using random numbers, this was a preparatory study to examine whether IRT-based simulation methods can determine the optimal robot-assisted therapy setting and to test several possible analysis methods.

**Methods**

**Data to be analyzed in the simulation**
The system used in this study was a ReoGo-J (Teijin Pharma Limited, Tokyo, Japan). ReoGo-J (Fig. 1) is a robotic system that facilitates functional recovery of the paralyzed hand by allowing survivors of stroke to voluntarily control the robotic arm with the paralyzed hand by following instructions displayed on the monitor. The effectiveness of this robot in UE improvement has already been investigated in two RCTs [8, 9]. The ReoGo-J system set up training tasks according to the severity of the patient's paralysis using the following three parameters: 1) Select the appropriate training tasks from 17 types of tasks (Fig. 2) based on an assessment of the degree of freedom of movement in stroke survivors with severity of their paralysis and 2) adjust the reach range for each training task considering the severity of the patient's paralysis (in this study, the default reach range was set at 100%, 65%, and 30%—smaller reach range for severe case and wide range for light case), 3) for patients with moderate-to-severe paralysis, the level of robotic assistance was adjusted to maximize the patient's voluntary motor output, allowing motorized assisted repetitive movements (Fig. 3) at five levels of assistance. Seventy-one different ReoGo-J settings, combining three parameters, were adopted for the study after consultation with occupational therapy professionals (Supplemental Table 1).

A 3-point method was used to evaluate the quality of the patient's compensatory movements during their practice with the paralyzed hand for each of the 71 different ReoGo-J settings, where Quality of Compensatory Movement Score (QCM) of: 0 points, indicated excessive compensatory movements, including trunk movements and significant difficulty in paralyzed arm training (corresponding to a score of < 3.5 on the Motor Activity Log's Quality of Movement (QOM) scale [16]); 1 point, indicated a setting that is neither too difficult nor too easy, the paralyzed hand can complete practice with some effort (corresponding to a score between 3.5-4.0 on the Motor Activity Log's QOM scale); 2 points, indicated a setting where the subject enters a so-called 'Slack' state, i.e., the effort to generate voluntary movement is greatly reduced due to excessive robotic assistance (corresponding to > 4.0 points on the Motor Activity Log's QOM scale). These data were collected from patients with different degrees of UE paralysis.

According to the guidelines for IRT, the Consensus-based Standards for Selecting Health Measurement Instruments (COSMIN), the two-parameter logistic model (2PML) based on IRT requires over 500 subjects to meet the criteria (with 259–499 subjects described as “doubtful”) [17]. However, given the feasibility of a clinical study of this size, it was anticipated that about 300 cases would be reasonable. Therefore, in this simulation, the sample data was set to 300 cases. The sample data were generated using Monte Carlo methods to generate random numbers using SAS 9.4 for Windows (SAS Institute Inc., Cary, NC, USA) for future clinical trials. The sample data (categorical data with three values [QCM score]: 0, 1, and 2) of 300 cases, with 71 items per case, were analyzed in this simulation. Each of the 71 sample data items were assigned a number, with a higher number indicating a lower QCM score but an increased item difficulty. Additionally, we assigned a number to each of the 300 simulated data cases. We created simulated data such that the value of the categorical data (the higher the ability of the person [θ]) increased as the number value increased. Finally, it was adjusted to eliminate cases rated 0 or 2 points in all 71 items by assuming a scenario where actual patients who experienced stroke used ReoGo-J in this study.

Method of analysis

1. Analysis before addressing Local Independence.
Figure 3 shows the statistical procedures used to address the local independence. The item parameters were estimated using 2PML based on IRT, while the marginal maximum likelihood (MML) estimation method was used to estimate the discriminative power and difficulty level of each item. The ability of a person (θ) in each case was estimated using the MML estimation method, assuming a standard normal distribution. The item parameters and ability of the person (θ) were estimated using the IRT procedure in SAS. Parallel analysis was performed on the poly correlation coefficients to confirm the unidimensionality of each item.

For local independence, we calculated the Q3 statistic (i.e., the correlation coefficient of the difference between measured and expected values) as described by Yen et al. [18] and evaluated whether there were any item pairs for which the absolute value of the Q3 statistic exceeded 0.2. For each item pair exceeding 0.2, one item was excluded from the pair, and the analysis was conducted. The following three strategies were compared to determine the optimal method for item selection, based on the number of pairs exceeding the absolute value of 0.2 in the Q3 statistic (fewer), Cronbach’s alpha coefficient, eigenvalues of the polychoric correlation matrix, and values of Akaike information criterion (AIC) and Bayesian information criterion (BIC) (smaller values): a) exclude items with low response probability (maximum response probability) within a pair in the QCM 1-point item difficulty curve, b) exclude items with low item information content within a pair in the QCM 1-point item difficulty curve, c) exclude items with low item discrimination.

After excluding the relevant items, we also compared the model fit of the statistical findings using the one-parameter logistic model (1PLM) and 2PLM, based on the same criteria as those used for comparing abovementioned strategies.

2. Analysis after addressing local independence.

Figure 3 shows the statistical procedures performed after addressing the local independence. For the simulated data, the cases identified after addressing local independence were removed, and the item parameters were estimated using 2PML based on IRT. The MML estimation method was used to estimate the discriminative power and difficulty level of each item. The ability of a person (θ) in all cases were estimated using the MML estimation method, assuming a standard normal distribution.

Parallel analysis was performed on the poly correlation coefficients to confirm the unidimensionality of each item. For local independence, we calculated the Q3 statistic (correlation coefficient of the difference between the measured and expected values) for the remaining 61 items, as described by Yen et al. [17]. We evaluated whether there were any item pairs for which the absolute value of Q3 exceeded 0.2. A scatter plot was created with the estimated ability of the person (θ) set on the horizontal axis and the mean value of the QCM score on the vertical axis, to discriminate monotonicity. Each graph was created to represent the following: item characteristic curve (ICC), ICC for a single QCM point per item, item information function (IIF), and test information function (TIF).

3. Identification of the optimal item for each ability of the person (θ) in the patient

1) Identification based on the response probability criterion with a 1-point QCT:

For each target item to be analyzed, the probability of 1-point QCT was calculated using $p_{j1}(\theta) = p_{j1}\star(\theta) - p_{j2}\star(\theta)$, $p_{jC}\star(\theta) = 1/(1+e^{p[-Da j(\theta-b_jC\star)]})$. We identified the item with the largest response probability of
1-point QCT along with the ability of the person (θ) by substituting each item parameter into this equation. Additionally, for each ability of the person (θ), we examined the top seven items with the highest probability of 1-point response on the QCT.

2) Identification based on the item information criterion:

The information content of an item is calculated by

\[ I_j(\theta) = D^2 a_j ^2 \sum_{c=0}^{c-1} \frac{(pjc_\theta qjc_\theta) - (pjC+1_\theta qjC+1_\theta)}{pjc_\theta} \]  

We examined the top seven items for each ability of the person (θ) by substituting the parameters estimated from the MML estimator into this equation and changing the value of (θ).

3) Comparison of identification methods based on response probability criteria with the QCM score of 1 point and each item information criterion:

In terms of each criterion, we defined and investigated the top seven items with high values from the response probability criteria with a QCM score of 1 point and item information criterion and item information criterion: severe, in the -2.0 \( \leq \theta < -0.5 \) range; moderate, in the 0.5 \( \leq \theta < 0.5 \) range; and mild, in the 0.5 \( \leq \theta < 2.0 \) range. The analysis for the entire range of -2.0 \( \leq \theta < 2.0 \) was conducted simultaneously. The analysis was performed by incorporating the relevant items into the 2PML model based on the IRT for each condition (severity domain and identification method). The model fit was compared using the AIC and BIC values. All statistical analyses in this study were performed using SAS 9.4 and Microsoft excel 2016 for Windows.

**Results**

1. Simulation results of 71 items before addressing local independence (Analysis before addressing Local Independence)

The descriptive statistics of the QCM data for the 300 simulated cases had a mean value of 0.99, standard deviation of 0.64, minimum value of 0.03, median value of 1.01, and maximum value of 1.99. The average QCM scores for each item (71 items) had average values of 0.99, 0.30, 0.30, 0.48, 1.01, and 1.49, respectively.

With the increasing identification (ID) value of the case numbers (sample data), the frequency of appearance of the 0 points QCM score decreased, two points increased, and one point emerged almost consistently, regardless of the item number. Furthermore, one point appeared at the highest rate when the ID of the case was approximately 150. In the estimation of item parameters, the mean values of item discrimination and item difficulty 1.2 was 2.928, -0.436, and 0.449, respectively. The estimated value of item difficulty ([item difficulty 1 + item difficulty 2]/2) was 0.005. Item difficulty increased with the ID value of the item. The ability of the person (θ) was as follows: mean value − 0.001; standard deviation, 0.984; minimum value − 1.897; median, 0.007; and maximum, 2.064. The ability of the person (θ) increased with a higher ID value for the case number in the target cases.
To confirm the model assumption, the unidimensionality was confirmed by poly correlation analysis, which showed that the first eigenvalue in the poly correlation matrix was 56.094, and the second eigenvalue was 0.75. Since the second eigenvalue was less than 1.0, one-dimensionality was considered established. To confirm local independence, 10 item pairs (item ID: 70–71, 38–49, 1–19, 59–69, 5–19, 12–30, 13–28, 50–70, 8–29, 13–46) had an absolute Q3 value exceeding 0.2. When comparing the three strategies presented in the Methods section, strategy (a) that excluded items with low response probability (maximum response probability) within a pair in the QCM 1-point item difficulty curve, appeared to be the most suitable decision method for establishing local independence. Furthermore, the model fit between the 1PLM and 2PLM, showing that 2PLM was the ideal model for data analysis after the removal of 10 items (Table 1).
### Table 1
Comparison between three strategies, and two models in item response theory, for adopting the best method for selecting deleted items to establish local independence

<table>
<thead>
<tr>
<th>Evaluation method</th>
<th>Two-parameter logistic model</th>
<th>one parameter logistic model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategies for establishing local independence*</td>
<td>None</td>
<td>Strategy 1</td>
</tr>
<tr>
<td>Number of items analyzed †</td>
<td>71</td>
<td>61</td>
</tr>
<tr>
<td>Number of pairs (Q3 statistic &gt;</td>
<td>0.2</td>
<td>)</td>
</tr>
<tr>
<td>Cronbach α</td>
<td>0.991</td>
<td>0.990</td>
</tr>
<tr>
<td>Polychoric correlation matrix (Value of first eigenvalue)</td>
<td>56.1</td>
<td>48.2</td>
</tr>
<tr>
<td>The polychoric correlation matrix (Rate of first eigenvalue)</td>
<td>0.790</td>
<td>0.791</td>
</tr>
<tr>
<td>AIC‡</td>
<td>25930</td>
<td>22484</td>
</tr>
<tr>
<td>BIC‡</td>
<td>26719</td>
<td>23162</td>
</tr>
<tr>
<td>Comparison model fitting</td>
<td>×</td>
<td>†</td>
</tr>
</tbody>
</table>

*Strategy 1; QOM score 1 point: Exclude items with low 1-point response probability (maximum probability) when the QCM score is 1 point in item difficulty; Strategy 2: Exclude items with limited item content information when the QCM score is 1 point in item difficulty; Strategy 3: Exclude items with low item discrimination.

†The reason for the difference in the analyzed items for each strategy is that the target pairs included duplicate items.

‡Smaller AIC and BIC values can be interpreted as higher model fitting.

Abbreviations: AIC, Akaike’s information criterion; BIC, Bayesian information criterion; QCM, Quality of Compensatory Movement Score

### 2. Simulation results of 61 items after addressing local independence (Analysis before addressing Local Independence)

Ten items were deleted from the 71 items to preserve local independence using above strategy (a), resulting in 61 items being analyzed. The descriptive statistics of the QCM data for the 300 simulated cases were as follows: mean value 0.99, standard deviation 0.65, minimum value 0.03, median value 1.00, and maximum 1.98. The QCM score for each item (61 items) had a mean value of 0.99, standard deviation of 0.30, minimum...
value of 0.48, median value of 1.01, and a maximum value of 1.45. Overall, these results were almost identical to the results including the 71 items.

Along with the increased ID value of the case number, the frequency of appearance of the 0 points QCM score decreased, 2 points increased, and 1 point emerged almost consistently regardless of item number. Furthermore, one point appeared to be the highest when the ID of the case was approximately 150.

In the estimation of the item parameters, the mean values of item discrimination and item difficulty 1.2 were 2.931, 0.436, and 0.450, respectively. The estimated value of item difficulty (item difficulty 1 + item difficulty 2)/2 is 0.007. Item difficulty increased with the ID value of the item. The ability of the person (θ) was as follows: mean value, 0.0000; standard deviation, 0.981; minimum value −1.839; median, 0.007; and maximum, 1.998. The ability of a person (θ) increased with a higher ID number in the target cases. Although the distribution of the ability of a person (θ) was originally expected to be normally distributed, the actual distribution is left-skewed, as illustrated in Fig. 5a.

To confirm the model assumption, unidimensionality was confirmed by poly correlation analysis: the first eigenvalue in the poly correlation matrix was 48.244, and the second eigenvalue was 0.64; since the second eigenvalue was less than 1.0, one-dimensionality was considered established, ensuring unidimensionality. Local independence was confirmed as there were no pairs for which the absolute value of the Q3 statistic exceeded 2. Monotonicity was confirmed as the mean value of the QCM increased almost monotonically, corresponding to an increase in the ability of the person (θ) (Fig. 5b).

The apex of the curve of item probability in QCM score 1 was found to shift toward the direction of the higher ability of the person (θ) with higher values of item ID. This finding suggests that the difficulty of the item increased with a higher ID value.

The apex of each item's curve was found to shift toward the direction of the higher ability of the person (θ) with higher item ID values. This result suggests that the ability of the person (θ) to provide the maximum amount of information (optimal measurement accuracy) for an item increases with a higher value of ID. The item information functions for all 61 items are shown in Fig. 5c. This figure shows that the measurement accuracy and test information content of the sample data are maximized when the ability of the person (θ) is approximately zero.

3. Selection of the optimal item ID for the ability of the person (θ) of each subject

1) Discrimination based on the reaction probability criterion for the 1 point QCM score

The top seven items with the highest probability of obtaining a QCM score of one point for each ability of the person (θ) in the cases are shown in Fig. 6a-6c. In the range of the ability of the person (θ) from −2.0 ≤ θ < -0.5 (severe), the response probability of each item varied, but the items with the highest response probability were item ID2, 3, 6, 7, 8, 9, and 10 (Fig. 6a). In the range of the ability of the person (θ) from −0.5 ≤ θ < 0.5
(moderate), the response probability of each item varied with no discernible pattern, but the items with the highest response probability were items ID 30, 32, 33, 36, 39, 43, and 44 (Fig. 6b). In the range of the ability of the person (θ) from 0.5 ≤ θ < 2.0 (mild), the response probability of each item varied, but the items with the highest response probability were items ID 56, 61, 63, 65, 66, 68, and 71 (Fig. 6c).

2) Discrimination by the item information criterion

The top seven items with high item information content for each ability of the subject (θ) are shown in Fig. 6d-6f. In the range of the ability of the person (θ) from −2.0 ≤ θ < −0.5 (severe), the response probability of each item varied significantly, but the items with the highest response probability were items ID 2, 6, 7, 8, 10, 13, and 17 (Fig. 6d). In the range of the ability of the person (θ) from −0.5 ≤ θ < 0.5 (moderate), the response probability of each item was relatively consistent, but the items with the highest response probability were items ID 17, 31, 32, 39, 41, 43, and 51 (Fig. 6e). In the range of the ability of the person (θ) from 0.5 ≤ θ < 2.0 (mild), the response probability of each item varied significantly, but the items with the highest response probability were items ID 60, 62, 63, 66, 67, 68, and 71 (Fig. 6f).

3) Comparison of discrimination methods based on response probability criteria with the QCM score of 1 point and the item information criterion

The results of the 2PML model based on IRT for each condition are shown in Table 2. The following results were found: both conditions were comparable in the −2.0 ≤ θ < −0.5 range; the discrimination method based on the item information criterion in the −0.5 ≤ θ < 0.5 range; and the discrimination method based on the item information criterion was also relatively superior in the 0.5 ≤ θ < 2.0 range. The model fit of the item information criterion was also relatively superior for the comparison of conditions across all areas.
Table 2
Comparison of models based on item response theory

<table>
<thead>
<tr>
<th>Classification of the ability of the person (θ)*</th>
<th>-2.0 ≤ θ&lt; -0.5</th>
<th>-0.5 ≤ θ&lt; -0.5</th>
<th>0.5 ≤ θ &lt; 2.0</th>
<th>-2.0 ≤ θ &lt; 2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Severity)</td>
<td>(Severe)</td>
<td>(Moderate)</td>
<td>(Mild)</td>
<td>(All patients)</td>
</tr>
<tr>
<td>Criterion for selection**</td>
<td>Strategy 1</td>
<td>Strategy 1</td>
<td>Strategy 1</td>
<td>Strategy 1</td>
</tr>
<tr>
<td>Number of items</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Value of items</td>
<td>2, 3, 6, 7, 8, 10</td>
<td>30, 32, 33, 36, 39, 43, 44</td>
<td>17, 31, 32, 39, 41, 43, 51</td>
<td>56, 61, 63, 65, 66, 68, 71</td>
</tr>
<tr>
<td>Evaluation method</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cronbach α</td>
<td>0.922</td>
<td>0.929</td>
<td>0.940</td>
<td>0.915</td>
</tr>
<tr>
<td>Polychoric correlation matrix</td>
<td>5.578</td>
<td>5.706</td>
<td>5.842</td>
<td>5.474</td>
</tr>
<tr>
<td>(Value of first eigenvalue)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polychoric correlation matrix</td>
<td>0.797</td>
<td>0.815</td>
<td>0.835</td>
<td>0.782</td>
</tr>
<tr>
<td>(Rate of first eigenvalue)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC†</td>
<td>2920</td>
<td>2930</td>
<td>3189</td>
<td>3096</td>
</tr>
<tr>
<td>BIC†</td>
<td>2998</td>
<td>3009</td>
<td>3266</td>
<td>3174</td>
</tr>
<tr>
<td>Comparison model fitting</td>
<td>●</td>
<td>●</td>
<td>×</td>
<td>△</td>
</tr>
</tbody>
</table>

*The ability of persons (θ): a measure of ability to be measured by item response theory. It represents the estimated ability of an individual examinee, and its mean value is 0 for all examinees.

**Strategy 1: discrimination based on the reaction probability criterion for the QCM score of 1 point; strategy 2: discrimination by the item information criterion.

†Smaller AIC and BIC values can be interpreted as higher model fitting.

Abbreviations: AIC, Akaike's information criterion; BIC, Bayesian information criterion; QCM, Quality of Compensatory Movement Score.

Discussion
In this study, a simulation method was employed to investigate the feasibility of estimating the item difficulty for each training item setting and the corresponding ability of the person (θ). This study applied IRT and evaluated the QCM scores of 71 different training items in ReoGo-J. Our results showed that ReoGo-J training items could be set at the optimal difficulty level for each subject without being affected by assumed characteristics of the sample data, which was simulated with the most likely variance in stroke patients using Monte Carlo methods. This result could be because the application of the IRT allowed separate modeling of the ability of the person (θ) in the patient and item difficulty, as well as discrimination for each of the 71 items. Furthermore, this indicates that patients with a higher ability (θ) were more likely to adapt to items with higher difficulty.

Regarding the concept of task difficulty in motor learning, Guadagnoli et al. [19] stated that it is necessary to select a task with optimal difficulty for the subject’s ability to facilitate motor learning, even if subjects with different skill levels are given the same training tasks. In a study using machine learning, Wilson et al. [20] reported that the optimal difficulty level for learning a new concept is approximately 85% (neither too easy or difficult) of correct answers (success rate), considering the ability of the subject. Similarly, our study’s findings appear to follow this learning theory, indicating that simulation results possibly explain item difficulty according to the person’s ability.

Particularly, this simulation study confirmed the possibility of setting the optimal difficulty for seven out of the 61 types of training tasks in ReoGo-J, considering the ability of the person (θ) for each subject. This was confirmed based on the response probability criterion with a 1-point QCM score and the discrimination method based on the item information content criterion. Therefore, it is assumed that the optimal training task can be selected in a real clinical setting if the ability of a person (θ) can be evaluated.

Next, this study confirmed the best method to select the optimal seven training tasks that correspond to the ability of the subject (θ). This was completed by comparing two methods: a discrimination method of the item response probability criterion for the QCM score of 1 point and the discrimination method based on the item information criterion. Our results suggest that the discrimination method based on the item information criterion is more likely to help select the optimal training task to set for each subject compared to the discrimination method of the response probability criterion with a QCM score of one point.

Furthermore, in the discrimination method based on the item information criterion, the fit of the model based on the IRT was confirmed in the analysis based on severity (severe; -2.0 ≤ θ < -0.5, moderate; -0.5 ≤ θ < -0.5, mild; 0.5 ≤ θ < 2.0) and all subjects (-2.0 ≤ θ < 2.0). However, since only seven training tasks could be applied to all cases with varying severity of UE paralysis, the selected model may not translate well to a clinical scenario. Based on the results of these simulations, it seems reasonable to use the item information criterion discriminant method, segregated by UE paresis severity using the ability of the person (θ), when conducting analyses using clinical data. However, the results of this simulation study showed that the discriminant method based on item information criteria had a better model fit than the discriminant method based on item response probability criteria with the score 1 QCM. However, we assume that the model fit concept can change depending on the data variance. Therefore, it is necessary to re-examine the goodness-of-fit of the model when applying it to clinical data. Additionally, if the model adaptability of the two discrimination methods appears low when using clinical data, it would be necessary to check the model adaptability of items that can be seen in both discrimination methods.
If the procedure of this simulation study is to be used in clinical practice, it is necessary to assess the ability of the person ($\theta$) after stroke to assess the severity of paralysis when selecting items, although it is difficult to accurately estimate the ability of each stroke patient ($\theta$). However, Woodbury et al. [21] suggested a gold standard outcome measure for functional impairment of UE paralysis, and the Fugl-Meyer Assessment (FMA) for UE function can be used to categorize the severity of impairment (severe: 0–19 points; moderate: 20–46 points; mild: 47 points). Additionally, it may be necessary to evaluate the FMA score for UE in each patient and estimate the relationship between the FMA score for UE and the ability of the person ($\theta$) using a correlation formula to accurately estimate ($\theta$) in future clinical studies.

Finally, the IRT used in this analysis assumes local independence in its theoretical construction. In the simulation, the QCM scores for each item in the 300 cases were generated using random numbers for each subject. This simulation was performed on 300 subjects, and ten items were removed to establish local independence. This means that to establish local independence without eliminating any items, a higher number of subjects would be needed. In clinical data, there may be factors that affect local independence that cannot be detected in this simulation study. Thus, additional items might be eliminated in future clinical studies. Furthermore, 10 items had to be removed from this simulation study to establish local independence; items according to the person's ability value ($\theta$) were selected from the remaining 61 items. Therefore, the sample size required in future clinical trials was estimated to be approximately 300 cases, regardless of the COSMIN recommendations.

**Declarations**

**Ethics approval and consent to participate**

Not applicable

**Consent for publication**

Not applicable

**Availability of data and materials**

The datasets are available from the corresponding author on reasonable request.

**Competing interests**

TT reports receiving personal fees from Teijin Pharma.

OY reports no competing interest.

KD reports non-financial support from Teijin Pharma.

YU reports personal fees and non-financial support from Teijin Pharma.

**Funding**

The work was supported by Teijin Pharma Ltd, Tokyo, Japan.
Authors' contributions

TT and YU contributed equally to the study design and manuscript writing and review. YO contributed to manuscript writing. KD contributed to manuscript review. All authors have read and approved the final manuscript.

Acknowledgments

We would like to thank all parties involved in this study.

References


Figures
Figure 1

Robot-assisted self-training

The figure illustrates robot-assisted self-training using the ReoGo-J robotic upper extremity rehabilitation device.
Training items of the ReoGo-J upper extremity rehabilitation device

The ReoGo-J has 17 training items. The names of each training item are as follows: a) forward reaching, b) rotation reaching, c) abduction reaching, d) radial reaching (two dimensions), e) radial reaching (three dimensions), f) radial reaching (upper direction), g) radial reaching (lower direction), h) reaching in eight directions (counterclockwise rotation), i) reaching in eight directions (clockwise rotation), j) zigzag reaching, k) circle reaching, l) polygonal reaching (1), m) polygonal reaching (2), n) reaching to the body (mouth), o) reaching to the body (shoulder), p) reaching to the body (head), and q) reaching in eight directions (holding upper extremity in the air).

Figure 3

Robotic assistance at the different modes produced by ReoGo-J for voluntary movement

Guided mode: fully dependent on robotic assistance to complete training. Initiated mode: Requires voluntary movement only at the beginning of training. For the remainder of the training, full dependence on robotic assistance is required. Step-Initiated mode: Requires only a few voluntary movements and robot-dependent movements alternately to complete training. Follow-assist mode: Required above a certain level of voluntary movement in training while receiving low level of robotic assistance continuously. Free mode: Uses voluntary movement to complete training without requiring robotic assistance.
Figure 4

Flowchart: statistical analyses

Flowchart illustrating statistical analyses used in this study.
Figure 5

Distribution chart of various data

a) Variance of sample data in this study simulated by Monte Carlo method. b) Confirmation of monotonicity between quality of movement scores and ability of person for each sample data. c) Distribution of item information across 61 items.
Selection of the appropriate items for each ability of the person (θ) in sample data

In the severe (a), moderate (b), or mild group (c), seven items were selected that had a relatively high probability of scoring 1 on the QCM score. In the severe (d), moderate (e), or mild group (f), seven items were selected that had a relatively high item information function.

Supplementary Files

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

- supplementaltable1.docx