

## RESEARCH ARTICLE

# Validation of the Chinese Non-pharmacological Therapy Experience Scale in persons with intellectual disability

Lily Yuen Wah Ho<sup>1</sup>  | Kenny Chi Wing Chin<sup>2</sup> | Connie Yuen Yee Fung<sup>3</sup> | Claudia Kam Yuk Lai<sup>1</sup> 

<sup>1</sup>School of Nursing, The Hong Kong Polytechnic University, Kowloon, Hong Kong

<sup>2</sup>Stat Solutions Co, Kowloon, Hong Kong

<sup>3</sup>Hong Kong Association of Therapeutic Horticulture, Tsuen Wan, Hong Kong

## Correspondence

Lily Yuen Wah Ho, School of Nursing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong.  
Email: lily.yw.ho@polyu.edu.hk

## Abstract

**Aim:** To validate the 4- and 7-point Chinese Non-pharmacological Therapy Experience Scales and test the psychometric properties of the scales on persons with intellectual disability.

**Design:** A validation study.

**Methods:** Sixty-seven persons with intellectual disability were recruited from six hostels or centres for persons with intellectual disability in Hong Kong. A total of 1,163 and 1,161 observations were collected by the trained observers with the 4-point and 7-point scales, respectively. The floor and ceiling effects, inter-rater reliability, internal consistency, responsiveness of both scales and the scale equivalence were examined.

**Results:** The Cronbach's  $\alpha$  of the 4- and 7-point scales was .762 and .797, respectively. The correlation between the two scales was 0.906. The inter-rater reliability of the 4- and 7-point scales was 0.774 and 0.835, respectively. Neither scale had the floor or ceiling effects. The effect size of the 7-point scale was consistently higher than that of the 4-point scale.

## KEYWORDS

horticultural therapy, instrument development, instrument responsiveness, intellectual disability, process evaluation, reliability and validity

## 1 | INTRODUCTION

According to the Medical Research Council, process evaluation is a complement of outcome evaluation, with the emphasis being on how complex interventions are implemented, the causal mechanisms of impact and contextual factors associated with the outcomes (Moore et al., 2015). A process evaluation gives insights into the failure or success of the interventions (Craig et al., 2008), the barriers and facilitators in their implementation and the experience of the participants (Hulscher, Laurant, & Grol, 2003). It helps to explain why an intervention may or may not be effective by examining its internal dynamics and actual operations. Most

importantly, process evaluation helps to evaluate the effectiveness which is concerned by the funders. Good evaluation practices need to include both process evaluation and outcome evaluation (Munro & Bloor, 2010), as the outcomes of an intervention may be affected by the process of implementing the intervention, rather than by the intervention itself. Moreover, process evaluation helps to inform the future development of interventions. A process evaluation can be conducted using either a quantitative or a qualitative approach, or both, for example, using self-reported questionnaires, structured observations and interviews and involving different parties such as intervention participants and implementers (Moore et al., 2015).

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2019 The Authors. *Nursing Open* published by John Wiley & Sons Ltd.

## 2 | BACKGROUND

There is an increasing emphasis on evaluating the process of interventions. Researchers are aiming to better understand the key components in an intervention process and the fidelity in the implementation of that intervention (i.e. whether the intervention was delivered as planned) (Arends, Bode, Taal, & Van de Laar, 2017; Trigwell et al., 2015; Van Olmen et al., 2015). In this way, they can improve the process of implementing the intervention (Driediger et al., 2018; Liu et al., 2018). More importantly, they will be better able to identify the value and reach of that intervention (Milliron et al., 2017). This will allow the targeted outcomes, such as behaviours (Lin et al., 2018), physical activity levels (Driediger et al., 2018) and a healthy lifestyle (Heim, Stang, & Ireland, 2009) to be promoted with the intervention.

Horticultural therapy (HT) is a non-pharmacological therapy where people are encouraged to engage in horticultural activities to promote their physical, cognitive, psychological, spiritual and social health (Porchey, 2007). HT has become more commonly used in various healthcare settings such as in nursing homes (Tse, 2010), communities (Van Den Berg & Custers, 2011) and outpatient clinics (Verra et al., 2012). During the therapy, plants provide visual, auditory, tactile, gustatory and olfactory stimulation (Edwards, McDonnell, & Merl, 2012).

HT has been used on persons with intellectual disability (ID), and its effects have been explored in the literature. HT was found to have improved the sociality of elementary school children with ID (Kim, Park, Song, & Son, 2012). A recent study also showed that HT provided a pleasurable experience to adults with ID and improved their social self-efficacy (Lai, Ho, Kwan, Fung, & Mak, 2016). One advantage of HT is that the activities can be adjusted to suit the capabilities of different people. Some reviews of non-pharmacological therapies on persons with ID have been conducted, but the focuses of the reviews were on the outcomes rather than the process of implementing the therapies (Unwin, Tsimopoulou, Kroese, & Azmi, 2016; Vereenoghe & Langdon, 2013). The experience of the process is best to be evaluated from the perspective of a participant. However, persons with ID may have difficulty in expressing their thoughts and needs (Centers for Disease Control & Prevention, 2015). They usually express themselves with some non-verbal behaviours and facial expressions, which require the persons around them to guess their needs (Boardman, Bernal, & Hollins, 2014). Therefore, it may be highly challenging for the participants with ID, especially those with deficits or impairments in communication, to express their opinions. Observing their behaviours in the process is a possible way to evaluate the process of implementation.

Thus, an instrument that can be employed throughout the intervention to examine the implementation, mechanisms and context is important for evaluating the process of implementing HT. Yet, few instruments serve such a purpose for persons with ID in Chinese communities. The 4-point Non-pharmacological Therapy Experience Scale (NPT-ES) (Muñiz, Olazarán, Poveda, Lago, &

Peña-Casanova, 2011), which is in English, is a possible choice for use in a process evaluation of an intervention involving persons with ID, as it can be employed to assess the immediate affective and social effects during an intervention from the perspective of the observer. Nevertheless, in our experience, the 4-point NPT-ES does not easily capture qualitative differences in the level of engagement or behaviours of participants with ID in HT sessions. Thus, the NPT-ES score may not be able to reflect actual changes if the changes are not large.

Studies have suggested that a good instrument is one that is reliable, valid and responsive (Cummins & Gullone, 2000). However, systematic reviews have shown that the measurement of the responsiveness of instruments is always ignored (Gagné, Boulet, Pérez, & Moisan, 2018; Lamarche, Tejpal, & Mangin, 2018; Redlich-Amirav, Ansell, Harrison, Norrena, & Armijo-Olivo, 2018), although responsiveness is an important psychometric property. Responsiveness is defined as the ability of an instrument to detect clinically meaningful differences (Hays & Hadorn, 1992). It has been pointed out that, apart from reliability and validity, sensitivity is a basic property of a scale since small deviations are highly meaningful in subjective instruments (Cummins & Gullone, 2000). Therefore, the scale should be responsive to change. An instrument with poor response will show little change in score even with a significant improvement, while a highly responsive instrument will show some change even with only minimal improvement (van Bennekom, Jelles, Lankhorst, & Bouter, 1996).

Preston and Colman (2000) examined the optimal number of response categories in rating scales. They found that 2- to 4-point scales performed relatively poorly and that the test-retest reliability was more likely to decrease for instruments with more than 10 responses. They suggested that 7-10 points be used for rating. Lozano, García-Cueto, and Muñiz (2008) studied the effect of the number of response categories on the psychometric properties of rating scales and reported that the optimum number of scale points was between 4-7. Since both Preston and Colman (2000) and Lozano et al. (2008) support the use of 7 points and because the range of points can have a profound effect on the outcome measurement in rater-graded scales, a 7-point scale that can detect small, clinically significant differences in interventions will further be developed in this study. The purpose of this study was to develop and test the psychometric properties of the 4- and 7-point Chinese NPT-ESs (CNPT-ESs). The objectives were to translate the 4-point scale into Chinese, expand the Chinese 4-point scale into 7 points, validate both the 4- and 7-point CNPT-ESs and compare them in an HT intervention involving persons with ID.

## 3 | METHODS

### 3.1 | Design

This study involved 2 phases. In phase 1, the 4-point CNPT-ES was translated into Chinese and expanded into 7 points. In phase 2, the psychometric properties were evaluated in 67 persons with ID after each session of HT intervention. Data were collected from June 2017 to January 2018.

### 3.2 | Phase 1: translating the instrument into Chinese and expanding it into a 7-point scale

The NPT-ES assesses the overall experience of participants during an intervention process. It was developed based on several types of non-pharmacological interventions, including cognitive stimulation, massage, psychomotor exercises, the use of music and training in the activities of daily living (Muñiz et al., 2011). This scale assesses five items: participation, pleasure, relationship with others, displeasure and rejection. Each item is rated on a 4-point scale ranging from 0–3 (0 = never, 1 = sometimes, 2 = frequently, 3 = always). The total score ranges from 0–15, with a higher total score representing a more positive experience. The internal consistency and the inter-rater reliability were estimated with observers watching videos of the sessions only, or taking on different roles, such as that of non-participant observer or therapist. The internal consistency in terms of the Cronbach's  $\alpha$  of the 4-point scale was .86–.88 when the non-participant observers rated the participants during the sessions, while the inter-rater reliability in terms of the intra-class correlation coefficient (ICC) was 0.83 (Muñiz et al., 2011).

To apply the scale across cultures, the scale was translated according to the procedures recommended by Brislin (1986). The original 4-point scale was translated into Chinese by a senior member of the research staff and also by a technical writer at the University, both proficient in Chinese and English. The two translated Chinese versions were then compared and appraised by the last author, and a final Chinese version was derived. This translated Chinese version was then back-translated into English by a research assistant whose first language is English but who is also proficient in Chinese. The back-translated version was compared with the original English version and again verified by the research team as having a high level of equivalence and in no need of further changes in wording. The equivalence of the original English and the Chinese versions of the 4-point scale was therefore confirmed.

To improve the sensitivity of the scale, the 4-point CNPT-ES was expanded into 7 points, ranging from 0–6 (0 = never, 1 = seldom, 2 = sometimes, 3 = half the time, 4 = quite frequently, 5 = very frequently, 6 = always). The total score of the 7-point scale ranges from 0–30, with a higher total score indicating a more positive experience with the non-pharmacological therapy.

### 3.3 | Phase 2: psychometric evaluation

#### 3.3.1 | Setting and sample

Using convenience sampling, participants were recruited from those who had enrolled in HT at six hostels or centres, which are funded by the government for persons with ID in Hong Kong. In Hong Kong, ID is diagnosed by a physician, following the definition in the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (the fourth edition). The grades of severity are classified as mild, moderate, severe and profound, with the intelligence quotient level ranging from 50–55 to approximately 70, 35–40 to 50–55,

20–25 to 35–40 and below 20–25, respectively (Labour & Welfare Bureau, 2007). The inclusion criteria for this study were as follows: those age 18 or above, able to understand Cantonese and with a diagnosis of mild to severe grade of ID. The staff incharges of the participating venues were asked to screen the potential participants, following the inclusion criteria. They also helped to invite the eligible members to join this study.

The inter-rater reliability in terms of the ICC was 0.82 in our previous pilot test (unpublished data) and 0.83 in the original 4-point scale (Muñiz et al., 2011). The sample size was estimated according to the method developed by Bonett (2002). At a significance level of 0.05, eight observation time points and a 95% confidence interval at a width of 0.20, 70 participants would be needed to achieve 80% statistical power. During the study period, there were 83 eligible participants. Among them, 67 consented to join.

#### 3.3.2 | Intervention

The programme was a group-based intervention, with eight weekly sessions. Each group had 6–10 persons with ID. Each session lasted for 60 min. The aim of the programme was to foster a sense of connection between participants and plants. The programme included standard contents, such as seed sowing, plant tending, decoration of flowerpots, plant artwork, harvesting and snack making (using the plants grew by the participants). The continuity of the HT activities was emphasized in the HT programme. The programme was delivered by the HT interns who had completed the HT training. The implementation of the programme was supervised by the registered horticultural therapist to ensure that the intervention protocol was closely followed. Other than the HT interns who led the sessions, three to four helpers were assigned to each group to assist the participants to manage different activities of the programme.

#### 3.3.3 | Rater training

Because numerous study sites were involved, 20 raters were trained to use both the 4- and 7-point CNPT-ESs through hands-on practice and discussions. Among them, 17 were HT interns, one was a registered horticultural therapist and two were research staff. Three of the raters were male. Two of the HT interns had completed elementary and intermediate levels of HT training, while the remaining 15 had already completed elementary, intermediate and advanced levels of HT training offered by the Hong Kong Association of Therapeutic Horticulture. The two research staff members did not have any knowledge of HT.

#### 3.3.4 | Data collection

The demographic data of the participants were collected with help from their guardians or the staff of the hostels or centres. In each session, the overall experience of each participant during the intervention process was rated by two, or sometimes three trained raters using both the 4- and 7-point CNPT-ESs. Each rater observed an average of not more than five participants to ensure

accuracy in the appraisal. This study followed the STROBE checklist (Appendix S1).

### 3.3.5 | Ethical considerations

This study followed the requirements of the Declaration of Helsinki. Research Ethics Committee approval was obtained from the first author's University. All potential recruits and their guardians were informed of the aims of the study. They were assured that there would be no penalties, regardless of when they withdrew from the study. The presence of the raters to collect data during the HT sessions was also explained. Participants with a mild degree of ID signed a written informed consent form. For participants with a moderate or severe degree of ID, written informed consent from their guardians was obtained. We also sought procedural, verbal and informed consent from all of the participants prior to the HT sessions.

### 3.3.6 | Statistical analysis

The data were analysed with IBM SPSS 24. Descriptive statistics were used to describe the participants' characteristics and the scale. Internal consistency reliability estimates were first conducted to determine whether there were any differences among the items of the Chinese 4-point and 7-point scales. A test of inter-rater reliability was also conducted. ICC was then used to examine the agreement between raters. The ratings of individual raters were then combined by stacking them on top of the others for both scales. The total scores were then compiled based on the five items in each of the scales. A Cronbach's  $\alpha$  coefficient was used to evaluate the internal consistency of both scales. According to Polit and Beck (2008), a Cronbach's  $\alpha$  value of  $>.7$  was satisfactory. The correlation between the two scales was then analysed by regression analysis as suggested by Colman, Morris, and Preston (1997). As the aim in this study was to develop a responsive instrument, effect size was used to assess the responsiveness of the instrument (van Bennekom et al., 1996). A total of three sessions, namely the 1st, 4th and 8th sessions, were used to test the responsiveness of the scale. These three sessions were chosen because the investigators wanted to observe the differences between the two scales at baseline, midway and at the end of the 8-week intervention programme. The differences in score between the 4th and 1st session, the 8th and 4th session and the 8th and 1st session therefore represented the early, late and overall responses, respectively. Participants who were present in the 1st, 4th and 8th sessions would be included in this pairwise comparison. Effect sizes of 0.20, 0.50 and 0.80 were regarded as small, moderate and large, respectively (Cohen, 1992). Other than the effect size, repeated-measures ANOVA with one within factor was also conducted to examine the responsiveness of the two scales using the 1st, 4th and 8th sessions. Test-retest reliability was not conducted, as each session in the HT programme differed.

## 4 | RESULTS

### 4.1 | Characteristics of the participants

In this study, the mean age of the 67 participants was 45.7% and 55.2% of them were male. All were single and 31.3%, 53.7% and 14.9% of them had mild, moderate and severe ID, respectively. The mean score in the Mini-Mental State Examination (Chiu, Lee, Chung, & Kwong, 1994) was 10.8. More than half of the participants did not have any experience in HT. Table 1 summarizes the demographic data of the participants.

### 4.2 | Mean CNPT-ES score of the sample

In total, there should have been 1,257 observations (counting all the programme cycles that were conducted) had all of the participants been present. Some participants were absent on 38 occasions due to, for example, attending follow-up appointments or competing activities. Two participants joined less than half of a session and those two occasions were not counted. In addition, the scoring of the 7-point scale was not reported on two occasions. Therefore, 1,163 and 1,161 observations, rated with the Chinese 4- and 7-point scales, respectively, were used in the analysis. The mean score of the observations of the entire sample using the 4-point scale was 10.26 (*SD* 2.13), with a score range of 0–15. For the 7-point scale, the mean score was 21.26 (*SD* 4.65), with the scores ranging from 0–30.

### 4.3 | Inter-rater reliability

The reliability of the measurements of the 20 raters was also compared. The results show that the mean value of the ICC across rater

**TABLE 1** Demographic data of 67 participants

Variables	N (%)
Gender	
Male	37 (55.2)
Female	30 (44.8)
Marital status	
Single	67 (100)
Degree of intellectual disability	
Mild	21 (31.3)
Moderate	36 (53.7)
Severe	10 (14.9)
Previous horticultural experience	
Yes	30 (44.8)
No	37 (55.2)
	Mean $\pm$ <i>SD</i>
Mini-Mental State Examination (scores)	10.8 $\pm$ 8.7 (range 0–27)
Age (years)	45.7 $\pm$ 12.4 (range 20–71)

**TABLE 2** Floor and ceiling effects of the 4-point scale and the 7-point scale

	Scale range	Mean (SD)	Definition of floor effect <sup>a</sup>	Definition of ceiling effect <sup>b</sup>	Floor effect (%)	Ceiling effect (%)
4-point scale sum score (N = 1,163)	0–15	10.26 (2.13)	≤7.2	≥11.9	8.6	7.8
7-point scale sum score (N = 1,161)	0–30	21.26 (4.65)	≤15.2	≥25.9	9.2	9.2

<sup>a</sup>Floor effect is equivalent to the 10% worst possible results of the scale (Bennett et al., 2002).

<sup>b</sup>Ceiling effect is equivalent to the 10% best possible results of the scale (Bennett et al., 2002).

pairs for the 4-point scale was 0.774 (SD 0.146), whereas the mean ICC of the rater pairs for the 7-point scale was 0.835 (SD 0.105). The ICC values mean that about 77.5% of the observed variance among the raters in the 4-point scale was similar, while the corresponding figure in the 7-point scale was 83.5%.

#### 4.4 | Internal consistency

The Cronbach's  $\alpha$  of the 4-point CNPT-ES using all observed data from the 20 raters was .734 (range 0.163–0.951) and that of the 7-point CNPT-ES was 0.766 (range 0.299–0.937). When looking at the data of individual raters, one rater was found to have a Cronbach's  $\alpha$  of <.3 on both scales. When that particular rater was excluded, the mean value of the Cronbach's  $\alpha$  of the 4-point CNPT-ES became .740 (range 0.494–0.951) and that of the 7-point CNPT-ES became 0.773 (range 0.530–0.937). The Cronbach's  $\alpha$  of the 7-point CNPT-ES had slightly better internal consistency.

Internal consistency reliability estimates were further conducted on the items of stacked data on both scales. The Cronbach's  $\alpha$  for the 4-point and 7-point CNPT-ESs was .762 and .797, respectively. Even for the analysis of stacked items, the performance of the 7-point scale was found to exceed that of the 4-point scale.

#### 4.5 | Estimation of scale equivalence

The correlation between the total scores for the 4-point and 7-point CNPT-ESs was high ( $r = .906$ ,  $p < .001$ ). To compare the equivalence and estimation of both scales, the ratings were further analysed using least squares regressions to determine the best fit for linear, quadratic and power function equations. These are the simplest equations that might reasonably be expected to explain the relationship between the 4-point and 7-point ratings. Of the three methods,

the results show that the quadratic equation is the best fit with the data. The coefficient of determination was  $R^2 = 0.908$  when the rating of the 4-point scale was used to estimate the 7-point scale, whereas it was  $R^2 = 0.907$  when the rating of the 7-point scale was used as an independent variable to estimate the 4-point scale.

#### 4.6 | Floor and ceiling effect

The floor effect and the ceiling effect were 8.6% and 7.8%, respectively, for the 4-point scale, and both the floor and ceiling effects were 9.2% for the 7-point scale (Table 2).

#### 4.7 | Responsiveness

As mentioned earlier, the difference in the means and the SDs was calculated in the three response periods. A positive score difference indicates improvement, whereas a negative difference indicates deterioration. Only 59 participants were present for all three sessions and were included in this pairwise comparison. The results show that effect sizes of the early response are much higher than those of the late response on both scales (Table 3). The effect size of the 7-point scale is consistently higher than that of the 4-point scale, indicating that the 7-point scale is more responsive than the 4-point scale. A repeated-measures ANOVA was conducted to assess whether there were differences between the three sessions with regard to the 4-point and 7-point scales, respectively. With the Huynh–Feldt correction, the 7-point CNPT-ES showed a significant difference in the 4th and 8th sessions compared with that of the 1st session (Tables 4 and 5). In support of this, polynomial contrasts indicated a significant linear trend,  $F(1,53) = 8.271$  ( $p < .01$ ) and a significant quadratic trend,  $F(1,53) = 11.973$  ( $p < .001$ ). With the assumption of sphericity, the 4-point CNPT-ES showed that there was no significant difference

**TABLE 3** Mean change scores and standard deviation of the 4-point scale and the 7-point scale (N = 59 participants)

	Early response <sup>a</sup>		Late response <sup>b</sup>		Overall response <sup>c</sup>	
	Mean change (SD)	Effect size <sup>d</sup>	Mean change (SD)	Effect size <sup>d</sup>	Mean change (SD)	Effect size <sup>d</sup>
4-point scale sum score	3.15 (12.10)	0.260	−0.97 (9.50)	0.102	2.34 (10.36)	0.226
7-point scale sum score	7.01 (13.27)	0.528	−2.56 (9.85)	0.260	4.22 (11.44)	0.369

Note: All scores in the 4- and 7-point scales have been standardized to a scale of 0–100 for computation.

<sup>a</sup>Early response = Middle Session – First Session.

<sup>b</sup>Late response = Last Session – Middle Session.

<sup>c</sup>Overall response = Last Session – First Session.

<sup>d</sup>Cohen's *d*.

**TABLE 4** Results of repeated-measures ANOVA with one within factor

Source	SS	df	MS	F
4-point scale				
Within subjects	6.406	2	3.203	2.394
Error	141.798	106	1.338	
7-point scale				
Within subjects	126.330	1.84	68.486	10.198*
Error	656.559	97.76	6.716	

\* $p < .001$ .

among the sessions (Tables 4 and 5). It is evident that the 7-point scale detects changes more effectively than the 4-point scale.

## 5 | DISCUSSION

This was the first study to validate a Chinese version of a process evaluation instrument to examine the experiences with a non-pharmacological therapy that people with ID might have. This client group was chosen for this validation study because they might have been unable to articulate their experiences and give comments about the implementation process, making the process evaluation challenging to conduct. The CNPT-ESs can fill this gap that exists in the process evaluation of interventions.

The ICC was good on both CNPT-ESs, indicating a high degree of agreement among the raters (Hallgren, 2012). The ICC of the 7-point scale was slightly better than that of the 4-point scale. This is consistent with the result of another study showing that the inter-rater reliability increased steadily from a 2-point to a 7-point scale (Cicchetti, Shoinralter, & Tyrer, 1985) as more points allowed behavioural expressions to be captured more fully (Preston & Colman, 2000).

In terms of internal consistency, one rater had a Cronbach's  $\alpha$  of  $< .3$ . One possible explanation for this was that the rater needed further training. Since most raters had a Cronbach's  $\alpha$  that was not low, both 4- and 7-point CNPT-ESs are reliable for use, with the 7-point scale having a higher Cronbach's  $\alpha$ . In the stacked data, the Cronbach's  $\alpha$  of the 7-point scale was still better than that of the 4-point scale. These results are consistent with the findings reported in previous studies on Likert scales, which showed that

psychometric properties improved with more response categories (Lozano et al., 2008; Preston & Colman, 2000). Moreover, scales with 4 or fewer points have a lower discriminative power than those with 6 or more points (Preston & Colman, 2000). Since small deviations can be highly meaningful in statistical analyses, a scale would become more sensitive following an expansion in the number of points (Cummins & Gullone, 2000), as in the case of the 7-point CNPT-ES.

Floor and ceiling effects are among the criteria used to assess the properties of an instrument (Terwee et al., 2007). According to Bennett et al. (2002), there were no floor and ceiling effects for both CNPT-ESs. This implies that extreme items are not missed at both ends of the scales and that differences in the responses of the participants can be discerned (Terwee et al., 2007). Thus, both scales can capture clinically meaningful data.

Regarding the sensitivity of the scales to changes, the 7-point scale shows a consistently greater effect size than the 4-point scale in all responses. In the early response, the 7-point scale reflects an obvious difference by giving a moderate effect instead of the small effect found in the 4-point scale. In late and overall responses, although the effect size of both scales is small, the 7-point scale still consistently performed slightly better than the 4-point scale. The repeated-measures ANOVA further showed that the 7-point scale detected changes more sensitively than did the 4-point scale. These findings suggest that, with the expansion in points, the 7-point scale is better at identifying clinically significant changes in the experience of the participants. It is important to have a scale that is responsive at detecting small changes, since this will help healthcare professionals to identify the active therapeutic component in an intervention.

In our study, the change in the mean CNPT-ES score in the early, late and overall responses provides some clinical information. The positive change in the mean score in the early response indicated that the participants showed an increase in interest in the HT programme during the first half of the programme. However, there was a negative change in the mean score in the late response. The overall positive change in outcome was relatively small compared with the early response. It is difficult to know the reasons for such differences in response at different stages of an intervention. But this is exactly why researchers and clinicians need to further explore what makes an intervention work or how long an intervention should last. All of these differences in the intervention effect at different time points could be related to many factors. For example, the rearrangement of

**TABLE 5** Means and standard deviations of the 4-point and 7-point scales in three sessions

Session	4-point scale			7-point scale		
	Mean	SD	Mean difference with 1st session	Mean	SD	Mean difference with 1st session
1st	10.06	2.11		20.26	4.67	
4th	10.53	1.89	0.472	22.40	3.89	2.139*
8th	10.40	1.65	0.340	21.60	3.77	1.349**

\* $p < .001$ .\*\* $p < .05$ .



the seating plan in some sessions might have positively or negatively influenced the social interactions of the participants with ID with the participants sitting next to them. In addition, tiredness caused by the side effects of some medications might have reduced the participants' enjoyment of the HT session. The postponing of sessions due to frequent typhoons in Hong Kong in 2017 might also have affected the continuity of the HT programme.

## 5.1 | Limitations

There are some limitations in this study. The sample might not be reflective of the target population, as convenience sampling was used. In addition, although training was provided to the 20 raters and they were assessed prior to the validation study to ensure that they had mastered the skills of observation and rating, there might still have been variations in the performance of a fairly large group of raters. Nonetheless, the good inter-rater reliability and internal consistency reveal that CNPT-ES is a user-friendly scale for clinical use.

## 6 | CONCLUSION

In conclusion, the findings of our study indicate that the 7-point scale has no floor and ceiling effects and is highly equivalent to the original 4-point scale. It has slightly better inter-rater reliability and internal consistency. It is a more responsive instrument that reflected changes in the behaviour of the participants with ID when they experienced the HT intervention. With more response categories, meaningful clinical changes can more easily be captured. Further validation of the 7-point scale is recommended using a larger client population or various client populations such as those who have cognitive impairment and with different interventions.

## ACKNOWLEDGEMENTS

We are grateful to Ms. June Lau for coordinating the raters and for providing other support for this study. We also thank Mr. Hugo Lam for his kind help in data collection. We acknowledge the support and facilitation of the participating organizations. We sincerely thank all of the participants and the project team of the Hong Kong Association of Therapeutic Horticulture for their participation, and the interns involved in running the horticultural therapy sessions.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## AUTHOR CONTRIBUTIONS

LYWH and CKYL conceived or designed the work. LYWH, CYFF and CKYL acquired the data. LYWH, KCWC and CKYL analysed and interpreted the data. LYWH, KCWC, CYFF and CKYL drafted

the article or revised it critically for important intellectual content. LYWH, KCWC, CYFF and CKYL approved the version to be published.

## ORCID

Lily Yuen Wah Ho  <https://orcid.org/0000-0001-8967-8763>

Claudia Kam Yuk Lai  <https://orcid.org/0000-0001-8503-5429>

## REFERENCES

- Arends, R. Y., Bode, C., Taal, E., & Van de Laar, M. A. (2017). A mixed-methods process evaluation of a goal management intervention for patients with polyarthritis. *Psychology & Health, 32*(1), 38–60. <https://doi.org/10.1080/08870446.2016.1240173>
- Bennett, S. J., Oldridge, N. B., Eckert, G. J., Embree, J. L., Browning, S., Hou, N., ... Murray, M. D. (2002). Discriminant properties of commonly used quality of life measures in heart failure. *Quality of Life Research, 11*(4), 349–359.
- Boardman, L., Bernal, J., & Hollins, S. (2014). Communicating with people with intellectual disabilities: A guide for general psychiatrists. *Advances in Psychiatric Treatment, 20*, 27–36. <https://doi.org/10.1192/apt.bp.110.008664>
- Bonett, D. G. (2002). Sample size requirements for estimating intraclass correlations with desired precision. *Statistics in Medicine, 21*(9), 1331–1335. <https://doi.org/10.1002/sim.1108>
- Brislin, R. W. (1986). The wording and translation of research instruments. In W. J. Lonner, & J. W. Berry (Eds.), *Field methods in cross cultural research* (pp. 137–164). Beverly Hills, CA: Sage Publications.
- Centers for Disease Control and Prevention (2015). *Facts about intellectual disability*. Retrieved from [https://www.cdc.gov/ncbddd/actearly/pdf/parents\\_pdfs/IntellectualDisability.pdf](https://www.cdc.gov/ncbddd/actearly/pdf/parents_pdfs/IntellectualDisability.pdf)
- Chiu, H. F. K., Lee, H. C., Chung, W. S., & Kwong, P. K. (1994). Reliability and validity of the Cantonese version of Mini-Mental State Examination: A preliminary study. *Journal of the Hong Kong College of Psychiatrists, 4*(suppl 2), 25–28.
- Cicchetti, D. V., Shoinralter, D., & Tyrer, P. J. (1985). The effect of number of rating scale categories on levels of interrater reliability: A Monte Carlo investigation. *Applied Psychological Measurement, 9*(1), 31–36. <https://doi.org/10.1177/014662168500900103>
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>
- Colman, A. M., Morris, C. E., & Preston, C. C. (1997). Comparing rating scales of different lengths: Equivalence of scores from 5-point and 7-point scales. *Psychological Reports, 80*(2), 355–362. <https://doi.org/10.2466/pr0.1997.80.2.355>
- Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: The new Medical Research Council guidance. *British Medical Journal, 337*, a1655. <https://doi.org/10.1136/bmj.a1655>
- Cummins, R. A., & Gullone, E. (2000). *Why we should not use 5-point Likert scales: The case for subjective quality of life measurement*. Paper presented at the Proceedings, Second International Conference on Quality of Life in Cities, Singapore.
- Driediger, M., Vanderloo, L. M., Burke, S. M., Irwin, J. D., Gaston, A., Timmons, B. W., ... Tucker, P. (2018). The implementation and feasibility of the Supporting Physical Activity in the Childcare Environment (SPACE) Intervention: A process evaluation. *Health Education and Behavior, 45*(6), 935–944. <https://doi.org/10.1177/1090198118775489>
- Edwards, C. A., McDonnell, C., & Merl, H. (2012). An evaluation of a therapeutic garden's influence on the quality of life of aged care residents with dementia. *Dementia, 12*(4), 494–510.

- Gagné, M., Boulet, L.-P., Pérez, N., & Moisan, J. (2018). Patient-reported outcome instruments that evaluate adherence behaviours in adults with asthma: A systematic review of measurement properties. *British Journal of Clinical Pharmacology*, 84(9), 1928–1940. <https://doi.org/10.1111/bcp.13623>
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23–34. <https://doi.org/10.20982/tqmp.08.1.p023>
- Hays, R., & Hadorn, D. (1992). Responsiveness to change: An aspect of validity, not a separate dimension. *Quality of Life Research*, 1(1), 73–75. <https://doi.org/10.1007/BF00435438>
- Heim, S., Stang, J., & Ireland, M. (2009). A garden pilot project enhances fruit and vegetable consumption among children. *Journal of the American Dietetic Association*, 109(7), 1220–1226. <https://doi.org/10.1016/j.jada.2009.04.009>
- Hulscher, M., Laurant, M., & Grol, R. (2003). Process evaluation on quality improvement interventions. *Quality and Safety in Health Care*, 12(1), 40–46. <https://doi.org/10.1136/qhc.12.1.40>
- Kim, B. Y., Park, S. A., Song, J. E., & Son, K. C. (2012). Horticultural therapy program for the improvement of attention and sociality in children with intellectual disabilities. *Horttechnology*, 22(3), 320–324.
- Labour and Welfare Bureau (2007). *Hong Kong Rehabilitation Programme Plan*. Retrieved from [https://www.lwb.gov.hk/eng/advisory/rac/rpp\\_report.htm](https://www.lwb.gov.hk/eng/advisory/rac/rpp_report.htm)
- Lai, C. K. Y., Ho, L. Y. W., Kwan, R. Y. C., Fung, C. Y. Y., & Mak, Y. W. (2016). An exploratory study on the effect of horticultural therapy for adults with intellectual disabilities. *Journal of Therapeutic Horticulture*, 26(2), 3–15.
- Lamarche, L., Tejpal, A., & Mangin, D. (2018). Self-efficacy for medication management: A systematic review of instruments. *Patient Preference and Adherence*, 12, 1279–1287. <https://doi.org/10.2147/PPA.S165749>
- Lin, S.-F., Binggeli-Vallarta, A., Cervantes, G., Angulo, J., Moody, J. S., McKenzie, T. L., ... Ayala, G. X. (2018). Process evaluation of an early care and education intervention: The California Childhood Obesity Research Demonstration Study (CA-CORD). *Health Promotion Practice*. <https://doi.org/10.1177/1524839918786953>
- Liu, H., Lindley, R., Alim, M., Felix, C., Gandhi, D. B. C., Verma, S. J., ... Jan, S. (2018). Family-led rehabilitation in India (ATTEND): Findings from the process evaluation of a randomized controlled trial. *International Journal of Stroke*, 14(1), 53–60. <https://doi.org/10.1177/1747493018790076>
- Lozano, L. M., García-Cueto, E., & Muñoz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology*, 4(2), 73–79. <https://doi.org/10.1027/1614-2241.4.2.73>
- Milliron, B.-J., Vitolins, M. Z., Gamble, E., Jones, R., Chenault, M. C., & Tooze, J. A. (2017). Process evaluation of a community garden at an urban outpatient clinic. *Journal of Community Health*, 42(4), 639–648. <https://doi.org/10.1007/s10900-016-0299-y>
- Moore, G. F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., ... Baird, J. (2015). Process evaluation of complex interventions: Medical Research Council guidance. *British Medical Journal*, 350, h1258. <https://doi.org/10.1136/bmj.h1258>
- Muñoz, R., Olazarán, J., Poveda, S., Lago, P., & Peña-Casanova, J. (2011). NPT-ES: A measure of the experience of people with dementia during non-pharmacological interventions. *Non-Pharmacological Therapies in Dementia*, 1(3), 1–11.
- Munro, A., & Bloor, M. (2010). Process evaluation: The new miracle ingredient in public health research? *Qualitative Research*, 10(6), 699–713. <https://doi.org/10.1177/1468794110380522>
- Polit, D. F., & Beck, C. T. (2008). *Nursing research: Generating and assessing evidence for nursing practice* (8th ed.). Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins.
- Porchey, P. (2007). Horticultural therapy: How can it make a difference in your everyday life. *Proceedings of the Florida State Horticultural Society*, 120, 351–352.
- Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power and respondent preferences. *Acta Psychologica*, 104(1), 1–15. [https://doi.org/10.1016/S0001-6918\(99\)00050-5](https://doi.org/10.1016/S0001-6918(99)00050-5)
- Redlich-Amirav, D., Ansell, L. J., Harrison, M., Norrena, K. L., & Armijo-Olivo, S. (2018). Psychometric properties of Hope Scales: A systematic review. *International Journal of Clinical Practice*, 72, e13213. <https://doi.org/10.1111/ijcp.13213>
- Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., ... de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, 60(1), 34–42. <https://doi.org/10.1016/j.jclinepi.2006.03.012>
- Trigwell, J., McGee, C. E., Murphy, R. C., Porcellato, L. A., Ussher, M., Garnham-Lee, K., ... Foweather, L. (2015). Process evaluation of a sport-for-health intervention to prevent smoking amongst primary school children: SmokeFree Sports. *BMC Public Health*, 15(1), 347. <https://doi.org/10.1186/s12889-015-1645-1>
- Tse, M. M. Y. (2010). Therapeutic effects of an indoor gardening programme for older people living in nursing homes. *Journal of Clinical Nursing*, 19(7–8), 949–958. <https://doi.org/10.1111/j.1365-2702.2009.02803.x>
- Unwin, G., Tsimopoulou, I., Kroese, B. S., & Azmi, S. (2016). Effectiveness of cognitive behavioural therapy (CBT) programmes for anxiety or depression in adults with intellectual disabilities: A review of the literature. *Research in Developmental Disabilities*, 51–52, 60–75. <https://doi.org/10.1016/j.ridd.2015.12.010>
- van Bennekom, C. A., Jelles, F., Lankhorst, G. J., & Bouter, L. M. (1996). Responsiveness of the rehabilitation activities profile and the Barthel index. *Journal of Clinical Epidemiology*, 49(1), 39–44. [https://doi.org/10.1016/0895-4356\(95\)00559-5](https://doi.org/10.1016/0895-4356(95)00559-5)
- Van den Berg, A. E., & Custers, M. H. G. (2011). Gardening promotes neuroendocrine and affective restoration from stress. *Journal of Health Psychology*, 16(1), 3–11. <https://doi.org/10.1177/1359105310365577>
- Van Olmen, J., Van Pelt, M., Malombo, B., Ku, G. M., Kanda, D., Heang, H., ... Schellevis, F. (2015). Process evaluation of a mobile health intervention for people with diabetes in low income countries: The implementation of the TEXT4DSM study. *Journal of Telemedicine and Telecare*, 23(1), 96–105. <https://doi.org/10.1177/1357633X15617885>
- Vereenoghe, L., & Langdon, P. E. (2013). Psychological therapies for people with intellectual disabilities: A systematic review and meta-analysis. *Research in Developmental Disabilities*, 34(11), 4085–4102. <https://doi.org/10.1016/j.ridd.2013.08.030>
- Verra, M. L., Angst, F., Beck, T., Lehmann, S., Brioschi, R., Schreiner, R., & Aeschlimann, A. (2012). Horticultural therapy for patients with chronic musculoskeletal pain: Results of a pilot study. *Alternative Therapies in Health and Medicine*, 18(2), 44–50.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Ho LYW, Chin KCW, Fung CYY, Lai CKY. Validation of the Chinese Non-pharmacological Therapy Experience Scale in persons with intellectual disability. *Nursing Open*. 2020;7:390–397. <https://doi.org/10.1002/nop2.398>