

# Reliability and validity of a physical capacity evaluation used to assess individuals with intellectual disabilities and mental illness

Yuh Jang<sup>a</sup>, Tzyh-Chyang Chang<sup>b</sup> and Keh-Chung Lin<sup>a</sup>

Physical capacity evaluations (PCEs) are important and frequently offered services in work practice. This study to investigate the reliability and validity of the National Taiwan University Hospital Physical Capacity Evaluation (NTUH PCE) on a sample of 149 participants consisted of three groups: 45 intellectual disability (ID), 56 mental illness (MI), and 48 no disability (ND). Intraclass correlation coefficient (ICC), Cronbach  $\alpha$ , and one-way ANOVA were calculated where appropriate. Results indicated that most subtests of the NTUH PCE had good to excellent test-retest reliability (ICC, 0.75–0.95), with the exception of subtests of the position tolerance section for ID and MI groups and the hand coordination section for the ND group. Good internal consistency was shown on strength, mobility, and hand coordination sections ( $\alpha$ , 0.96, 0.90, and 0.95, respectively), with the exception of the position tolerance section ( $\alpha=0.37$ ). The known group validity of the NTUH PCE in people with disabilities was satisfactory. In conclusion, although some improvements are needed in some subtests, the NTUH PCE is suitable for measuring physical work-related abilities of people with disabilities.

Physical capacity evaluation-tests (PCEs) sind wichtige und häufig angebotene Dienstleistungen zur Beurteilung von motorischen Funktionsbereichen in der beruflichen Praxis. Die vorliegende Studie untersucht die Reliabilität und Gültigkeit der Physical Capacity Evaluation-Tests des National Taiwan University Hospital (NTUH PCE) anhand einer Stichprobe von 149 Studienteilnehmern, die drei Kategorien angehören: 45 Teilnehmer mit geistiger Behinderung (ID), 56 mit psychischer Erkrankung (MI) und 48 ohne Behinderung (ND). Der Intraklass-Korrelationskoeffizient (IKK), Cronbach-Alpha und die einfaktorielle Varianzanalyse (ANOVA) wurden je nach Sachlage berechnet. Den Ergebnissen zufolge wiesen die meisten Subtests der NTUH PCE mit Ausnahme der Subtests zur Lagetoleranz bei den ID- und MI-Gruppen und zur Handkoordination bei der ND-Gruppe eine gute bis ausgezeichnete Test-Wiederholungsreliabilität auf (ICC, 0.75–0.95). Eine gute interne Konsistenz lag bei den Bereichen Stärke, Mobilität und Handkoordination vor (IKK=jeweils 0,96, 0,90 und 0,95), mit Ausnahme der Lagetoleranz ( $\alpha=0.37$ ). Die bekannte Gruppenvalidierung der NTUH PCE bei Menschen mit Behinderung war zufrieden stellend. Daraus lässt sich schließen, dass die NTUH PCE trotz Verbesserungsbedarfs in manchen Subtests für die Messung der körperlichen

berufsrelevanten Fähigkeiten von Menschen mit Behinderung geeignet ist.

Las valoraciones de la capacidad física (VCF) son servicios importantes y frecuentemente brindados en la práctica laboral. Este estudio investiga la fiabilidad y validez de la Valoración de la Capacidad Física aplicada en el Hospital Nacional Docente de Taiwán (VCF HNDT). Para ello utilizamos una muestra de 149 participantes, divididos en tres grupos: 45 con discapacidad intelectual (DI), 56 con trastornos psíquicos (TP) y 48 sin discapacidad (SD). En los casos pertinentes, se calcularon el coeficiente de correlación intraclass (ICC), el  $\alpha$  de Cronbach y el ANOVA de una sola vía. Los resultados indicaron que la mayoría de las subpruebas de la VCF HNDT tuvieron una fiabilidad de buena a excelente con relación a las pruebas repetidas (ICC, 0.75–0.95), excepto en el caso de las subpruebas de la sección sobre la tolerancia de la posición en los grupos con DI y con TP y en la valoración de la coordinación de la mano en el grupo SD. Se halló una buena consistencia interna en las secciones relativas a la fuerza, la movilidad y la coordinación de la mano ( $\alpha$ , 0.96, 0.90 y 0.95, respectivamente), excepto en el caso de la sección sobre la tolerancia de la posición ( $\alpha$ , 0.37). La validez grupal conocida de la VCF HNDT en personas con discapacidades fue satisfactoria. En conclusión, aunque algunas subpruebas necesitan ser mejoradas, la VCF HNDT resulta apropiada para medir las capacidades físicas relativas al trabajo en personas con discapacidades.

Les évaluations de capacité physique (ECP) constituent des services importants et fréquemment proposés dans le cadre professionnel. La présente étude examinait la fiabilité et la validité de l'évaluation de capacité physique de l'hôpital universitaire national de Taïwan (NTUH PCE) sur un échantillon de 149 participants répartis en trois groupes: 45 incapacité intellectuelle (II), 56 maladie mentale (MM), et 48 aucune invalidité (AI). Le coefficient de corrélation Intraclass (ICC), l'indice Cronbach  $\alpha$  et la mesure ANOVA unilatérale ont été calculés ad hoc. Les résultats indiquent que la plupart des sous-essais de l'évaluation NTUH PCE présentent une excellente fiabilité de test-test secondaire (ICC, 0.75–0.95), à l'exception des sous-essais de la section de tolérance de position pour les groupes II et MM et de la section de coordination manuelle pour le groupe de AI. Good internal consistency was shown on strength, mobility, and hand coordination sections ( $\alpha$ , 0.96, 0.90, and 0.95, respectively), with the exception of

the position tolerance section ( $\alpha=0.37$ ). Une bonne homogénéité interne a été démontrée pour les sections concernant la force, la mobilité et la coordination manuelle (ICC=0.96, 0.90, et 0.95 respectivement), à l'exception de la section tolérance de position ( $\alpha=0.37$ ). La validité de groupe connue de l'évaluation NTUH PCE pour les personnes handicapées était satisfaisante. Pour conclure, bien que certaines améliorations soient nécessaires dans quelques-uns des sous-essais, l'évaluation NTUH PCE est adéquate pour mesurer les capacités physiques liées au travail des personnes handicapées. **Mots clefs :** Évaluation de capacités physiques, fiabilité de test-test secondaire, validité de groupe connue, incapacité intellectuelle, maladie mentale. *International Journal of Rehabilitation*

## Introduction

For most people, work is one of the major daily activities in their life. To fulfill the specific demands of work, individuals need to possess specific work behaviors, skills, aptitudes, and physical capacities (Jacobs, 1991). Physical capacity is considered as second only to cognitive capacity as a barrier to the execution of skilled tasks, especially for people with disabilities (Menard and Hoens, 1994; Silverstein *et al.*, 1991). To assess an individual's physical capacity is important for determining the factors that limit work performance and the types and regimens of intervention that most appropriately address each limitation. For this purpose, physical capacity evaluations (PCEs) have become the most important tools in medical, occupational, and legal settings and the most frequently offered services in the work practice (Deen *et al.*, 2002; Jundt and King, 1999). A PCE is a systematic and objective tool to measure an individual's physical ability to perform work-related activities based on the physical demand factors of the *Dictionary of Occupational Titles* (DOT), a publication of the United States Department of Labor (United States Department of Labor, 1993).

Since the enactment of Taiwan's Protecting Physically and Mentally Disabled People Act in 1997 and its Occupational Accident Labor Protection Law in 2002, use of PCEs in vocational and occupational rehabilitation has accelerated in Taiwan. Three major areas for the use of PCEs in Taiwan are present: vocational rehabilitation; occupational rehabilitation; and litigation. Within vocational rehabilitation, PCEs are used to define the maximal physical work-related abilities of people with disabilities and determine barriers of vocational rehabilitation or job placement services by comparing results with known demands of occupations. Within occupational rehabilitation, there are two reasons for doing a PCE: job-specific evaluation before and during rehabilitation to formulate and examine the progression and the effectiveness of the occupational rehabilitation plan; and disability determination evaluation to assist the physician

**Research 32:77–84 © 2009 Wolters Kluwer Health | Lippincott Williams & Wilkins.**

International Journal of Rehabilitation Research 2009, 32:77–84

**Keywords:** intellectual disability, known group validity, mental illness, physical capacity evaluation, test-retest reliability

<sup>a</sup>School of Occupational Therapy, College of Medicine, National Taiwan University, Taipei and <sup>b</sup>Department of Occupational Therapy, Chang Gung University, Taoyuan, Taiwan.

Correspondence to Yuh Jang, Assistant Professor of School of Occupational Therapy, College of Medicine, National Taiwan University, Taipei 100, Taiwan  
Tel: +886 23 322 8164; e-mail: yuhj@ntu.edu.tw

Received 24 March 2008 Accepted 7 July 2008

in making a return to work decision for the injured worker. Within litigation, the PCEs are used to assist physicians in the measurement of the extent of injury and loss of work ability in personal injury and insurance liability cases.

The increasing costs of workers' compensation and vocational rehabilitation have created a substantial need for a PCE that is an effective and practical tool for use in Taiwan. Although several PCEs are available on the market, they are expensive and lack studies of reliability and validity for people with disabilities (Innes and Straker, 1999a, 1999b; Jones and Kumar, 2003; King *et al.*, 1998). Most reliability and validity studies focus on healthy people and people with physical disabilities, especially back problems (Brouwer *et al.*, 2003; Durand *et al.*, 2004; Gross and Battie, 2002; Lygren *et al.*, 2005; Tuckwell *et al.*, 2002). As the PCE also needs to be used in vocational rehabilitation for people with intellectual disability (ID) and people with mental illness (MI) in Taiwan, there are no reliability and validity studies reported. Thus, since the year 1995, a new PCE has been developed, called the National Taiwan University Hospital Physical Capacity Evaluation (NTUH PCE), to fulfill the needs of occupational and vocational rehabilitation services in Taiwan.

If a PCE is to be considered a useful tool, reliability and validity must be demonstrated (Gross, 2004; Innes and Straker, 1999a, b; Task Force on Standards for Measurement in Physical Therapy, 1991). Thus, the purpose of this study was to investigate the test-retest reliability, internal consistency, and known group validity of the NTUH PCE in participants with ID, MI, and no disability (ND).

## Methods

### Participants

Participants selected met three eligibility criteria: being 17 years of age or older, not having any medical condition

that would restrict them from performing maximally, and ability to comply with safety rules and to follow instructions during the testing period. The sample size was estimated based on a power analysis for discriminating performance on the NTUH PCE by the study groups. For a moderate effect size  $f$  of 0.30, at least 36 participants per group were needed to achieve a power of 0.80 to detect a significant group effect ( $\alpha = 0.05$ ) (Cohen, 1988). The study and consent forms were approved by the Ethics Committee of NTUH, Taiwan and all participants or their guardians provided written informed consent. Characteristics of participants are presented in Table 1.

Participants consisted of three groups: ID, MI, and ND. The ID group participants were recruited from a special school for individuals with ID. The 45 participants in the ID group consisted of 21 (47%) with mild ID, 18 (40%) with moderate ID, and six (13%) with severe ID based on disability identification cards that were issued by Taiwan government. The MI group was recruited from two psychiatric hospitals and the diagnoses included schizophrenia ( $n = 34$ ), affective psychosis ( $n = 20$ ), depressive disorder ( $n = 1$ ), and bipolar disorder ( $n = 1$ ). The ND group was recruited from the internet ( $n = 30$ ), advertisement in the Bulletin Board of the NTUH ( $n = 13$ ), and by recommendation of a friend ( $n = 5$ ).

A total of 149 participants were included in the study, consisting of 76 males and 73 females. Mean age was 25 years ( $SD = 7$ ), mean weight was 61 kg ( $SD = 12$ ), and mean height was 164 cm ( $SD = 8$ ). There were no statistically significant differences in sex ( $\chi^2$  test = 3.428,  $P = 0.180$ ), height ( $F = 0.06$ ,  $P = 0.938$ ), nor weight ( $F = 2.00$ ,  $P = 0.140$ ) among groups, but there was a statistically significant difference in age ( $F = 48.36$ ,  $P < 0.001$ ). Participants in the MI group were older than those in the ID and ND group, and participants in the ND group were older than those in the ID group.

## Measurements

On the basis of physical demand factors described by the DOT, several existing PCE tools and experts were consulted, and the NTUH PCE was developed in the

year 1995. The NTUH PCE is a battery of tests created to assist a therapist in determining an individual's maximum work-related physical abilities. It consists of 28 subtests classified into four sections: strength, mobility, hand coordination, and position tolerance. These subtests cover the nine out of 20 physical demand factors described by the DOT (United States Department of Labor, 1993), including strength, climbing, balancing, stooping, kneeling, crouching, crawling, handling, and fingering. The strength section evaluates the participant's ability to carry, push, pull, lift to different levels, and power grip/lateral (key) pinch force. The mobility section evaluates the participant's ability to climb stairs and stepladder, maintain balance while walking over a balance beam, crawl, and perform repetitive crouching/stooping. The hand coordination section evaluates the participant's ability to seize, turn, grasp, pick, and pinch an object. The position tolerance section evaluates the participant's ability to assume and maintain crouching, stooping, kneeling, sitting, and standing positions. Appendix 1 presents the description of the subtests of the NTUH PCE.

The NTUH PCE test battery requires 1.5–2 h to administer and was designed to be administered by an occupational or physical therapist. An additional 1–2 h are required for the therapist to summarize the test results, describe the participant's physical ability to perform work-related activities; determine an overall work level for the participant; and measure the participant's sensory function, range of motion of joints, and muscle power tested by manual muscle testing procedures. Equipment includes shelving at adjustable height intervals of 5 cm ranging from 35 to 180 cm; a 35 × 35 × 35 cm wooden box for weights with cut-out handles for easy gripping; weights at 2.5, 5, and 10 kg; a weight sled, 61 × 102 cm with three push/pull poles at 87, 97, and 107 cm in height; a stepladder; a crawling tunnel; a Purdue Pegboard (Lafayette Instrument, Lafayette, Indiana, USA); a Complete Minnesota Dexterity Test (CMDT) (Lafayette Instrument); a hand dynamometer (NexGen Ergonomics, Inc., Montreal, Quebec, Canada); a pinch gauge (B & L Engineering, California, USA); a balance beam (11 × 300 and 33 cm high); and a heart rate monitor.

## Data collection procedures

The study was completed at the Occupational Therapy Department of the NTUH. Participants or their guardians were introduced to the general testing procedures of the NTUH PCE and then signed informed consent before enrollment in this study. Participants or their guardians were free to stop testing or withdraw at any time. Before each subtest, participants were instructed verbally on the required tasks. The evaluator then demonstrated each subtest. In this way, 28 subtests were performed by each participant. Participants were asked to

**Table 1** Demographic characteristics of participants

| Category                                 | ID                 | MI                | ND                |
|------------------------------------------|--------------------|-------------------|-------------------|
| Number                                   | 45                 | 56                | 48                |
| Age (years) <sup>a</sup>                 | 20 ± 3 (17–29)     | 30 ± 7 (18–43)    | 24 ± 5 (18–36)    |
| Male                                     | 28 (62.2%)         | 27 (48.2%)        | 21 (43.8%)        |
| Female                                   | 17 (37.8%)         | 29 (51.8%)        | 27 (56.3%)        |
| Height (cm) <sup>a</sup>                 | 163 ± 10 (135–180) | 164 ± 8 (150–181) | 164 ± 8 (146–178) |
| Weight (kg) <sup>a</sup>                 | 60 ± 14 (37–104)   | 64 ± 12 (40–95)   | 59 ± 10 (42–87)   |
| Test–retest interval (days) <sup>a</sup> | 9 ± 3 (7–16)       | 10 ± 5 (7–21)     | 9 ± 3 (7–16)      |

ID, intellectual disability; MI, mental illness; ND, no disability.

<sup>a</sup>All values are given as a mean ± SD (range).

perform to their maximal abilities. The evaluator recorded the results directly after each test.

For test–retest reliability, two testing sessions were held within a 1–3-week interval by the same evaluator (Table 1). The testing interval was selected to minimize functional change and testing effect (i.e. to prevent the therapist remembering participant's performance in the first session, and learning effect of the participants). Time of day of testing was kept constant. At the end of first testing, the evaluator sealed the ratings and gave them to the primary researcher (first author). Participants were not informed of their testing results until all retesting was completed.

Five occupational/physical therapists were trained to conduct the test and score participant performance by attending a 3-day workshop on the NTUH PCE, reading the procedure manual, and completing the testing of three healthy volunteers. Each of the five therapists who participated in the project had more than 1 year professional experience and all had equivalent experience administering the NTUH PCE.

#### Data analysis

Intraclass correlation coefficients (ICCs, model one-way random, single measure) were calculated for test–retest reliability. Agreement strengths for ICC values have been classified as follows: < 0, poor; 0–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial, and 0.81–1.00, almost perfect (Landis and Koch, 1977). An ICC of 0.75 or more was considered acceptable reliability (Portney and Watkin, 2000; Reneman *et al.*, 2004).

#### Internal consistency

The Cronbach  $\alpha$  coefficient was used to examine the internal consistency of the subtests of each section in the NTUH PCE. This coefficient estimates the reliability of an instrument according to the extent to which items are internally consistent taking into consideration the number of items in the instrument (Cronbach, 1951). Cronbach  $\alpha$  values greater than 0.70 were considered acceptable for internal consistency (Fayers and MaChin, 2000).

#### Known group validity

Known group validity tests are used to examine differences among groups completing the same task (Fayers and MaChin, 2000; Innes and Straker, 1999b). For the purposes of this study, the ability of the subtest scores to discriminate among the ID, MI, and ND groups was tested by comparing the mean scores in the three groups, using one-way analysis of variance (ANOVA) tests. Tukey's HSD post-hoc analysis was conducted for determining significance of the difference between the groups if a significant *F*-test was found by ANOVA. All

analyses were performed using the SPSS (SPSS Inc., Chicago, Illinois, USA, version 11.0 for Windows).

## Results

Table 2 presents the results of the test–retest reliability of the NTUH PCE. Within the NTUH PCE, 25 out of 28 subtests had acceptable test–retest reliability for all participants, with the exception of standing, stooping, and kneeling where values ranged from 0.12 to 0.49). For the ID group, 21 out of 28 subtests had acceptable test–retest reliability. The ICC values of pushing and left hand and both hand portions of the Purdue Pegboard subtests were, however, fair to acceptable. Standing subtest had moderate agreement. Only stooping, crouching, and kneeling had unacceptable test–retest reliability.

For the MI group, 24 out of 28 subtests had acceptable test–retest reliability. ICC values of the pulling and walking subtests were fair to acceptable. Stooping subtest had moderate agreement. Only kneeling had unacceptable test–retest reliability.

For the ND group, 19 out of 28 subtests had acceptable test–retest reliability. ICC values for pushing, pulling, walking, balancing, and repetitive crouching subtests were fair to acceptable. The turning test of the CMDT and the three out of four of the Purdue Pegboard tests had fair-to-moderate agreement.

Table 2 presents the internal consistency of the sections of the NTUH PCE. Acceptable internal consistency (Cronbach  $\alpha$  values > 0.70) was found for strength, mobility, and hand coordination sections (ranging from 0.90 to 0.96), with the exception of the position tolerance section ( $\alpha$  = 0.37).

Table 3 shows the results of the known group validity by ANOVA. Significant mean differences were found among ID, MI, and ND groups for almost all subtests of the NTUH PCE, with the exception of the sitting subtest (*F* test = 1.16, *P* = 0.317). People without disabilities got better scores on all subtests of the NTUH PCE, especially in hand coordination and mobility sections. When comparing the ND with the ID group, results indicated significant differences in all subtests of the NTUH PCE, with the exception of the sitting subtest (*P* = 0.193). Significant differences between the ND group and MI group in 24 out of 28 subtests were present, with the exception of the sitting (*P* = 0.999), standing (*P* = 0.638), crouching (*P* = 0.096), lateral pinch-left (*P* = 0.396), and lateral pinch-right (*P* = 0.375) subtests. When comparing the MI with the ID group, we found that people with MI got significantly better scores in 15 out of 28 subtests, with the exception of the lifting, carrying, and push/pull subtests and position tolerance sections.

**Table 2** Internal consistency (Cronbach  $\alpha$  coefficient) and test–retest reliability (ICC and 95% CI) of the NTUH PCE

| Subtest                      | $\alpha$    | All ( $n=149$ )   | ID ( $n=45$ )     | MI ( $n=56$ )      | ND ( $n=48$ )    |
|------------------------------|-------------|-------------------|-------------------|--------------------|------------------|
| <b>Strength</b>              | <b>0.96</b> |                   |                   |                    |                  |
| Floor to knuckle lifting     |             | 0.87 (0.82–0.90)  | 0.91 (0.84–0.95)  | 0.88 (0.81–0.93)   | 0.81 (0.68–0.89) |
| Knuckle to shoulder lifting  |             | 0.93 (0.90–0.95)  | 0.94 (0.89–0.97)  | 0.93 (0.88–0.96)   | 0.90 (0.84–0.94) |
| Shoulder to overhead lifting |             | 0.90 (0.87–0.93)  | 0.88 (0.78–0.93)  | 0.91 (0.85–0.95)   | 0.88 (0.79–0.93) |
| Bilateral carrying           |             | 0.88 (0.83–0.91)  | 0.76 (0.61–0.86)  | 0.92 (0.87–0.95)   | 0.84 (0.73–0.91) |
| Pushing                      |             | 0.78 (0.71–0.83)  | 0.69 (0.50–0.82)  | 0.86 (0.77–0.91)   | 0.69 (0.51–0.81) |
| Pulling                      |             | 0.81 (0.74–0.86)  | 0.87 (0.79–0.92)  | 0.73 (0.56–0.84)   | 0.73 (0.56–0.82) |
| Power grip left              |             | 0.94 (0.92–0.96)  | 0.90 (0.83–0.95)  | 0.94 (0.90–0.97)   | 0.94 (0.89–0.96) |
| Power grip right             |             | 0.93 (0.91–0.95)  | 0.88 (0.79–0.93)  | 0.93 (0.89–0.96)   | 0.93 (0.88–0.96) |
| Lateral pinch left           |             | 0.90 (0.86–0.93)  | 0.92 (0.86–0.96)  | 0.85 (0.76–0.91)   | 0.90 (0.84–0.95) |
| Lateral pinch right          |             | 0.92 (0.89–0.94)  | 0.93 (0.87–0.96)  | 0.92 (0.87–0.95)   | 0.88 (0.80–0.93) |
| <b>Mobility</b>              | <b>0.90</b> |                   |                   |                    |                  |
| Walking                      |             | 0.75 (0.67–0.81)  | 0.75 (0.59–0.85)  | 0.74 (0.59–0.84)   | 0.70 (0.53–0.82) |
| Ladder climbing              |             | 0.92 (0.89–0.94)  | 0.89 (0.80–0.94)  | 0.87 (0.79–0.92)   | 0.77 (0.62–0.86) |
| Stair climbing               |             | 0.85 (0.80–0.89)  | 0.79 (0.65–0.88)  | 0.82 (0.71–0.89)   | 0.87 (0.77–0.92) |
| Balancing                    |             | 0.90 (0.86–0.92)  | 0.91 (0.84–0.95)  | 0.77 (0.63–0.86)   | 0.63 (0.42–0.77) |
| Crawling                     |             | 0.86 (0.81–0.90)  | 0.80 (0.67–0.89)  | 0.82 (0.71–0.89)   | 0.79 (0.66–0.88) |
| Repetitive stooping          |             | 0.92 (0.89–0.94)  | 0.86 (0.77–0.92)  | 0.89 (0.82–0.93)   | 0.76 (0.61–0.86) |
| Repetitive crouching         |             | 0.89 (0.85–0.92)  | 0.81 (0.68–0.89)  | 0.88 (0.80–0.93)   | 0.74 (0.59–0.85) |
| <b>Hand coordination</b>     | <b>0.95</b> |                   |                   |                    |                  |
| CMDT placing                 |             | 0.95 (0.94–0.97)  | 0.94 (0.89–0.96)  | 0.92 (0.87–0.95)   | 0.76 (0.61–0.86) |
| CMDT turning                 |             | 0.94 (0.92–0.96)  | 0.93 (0.87–0.96)  | 0.86 (0.78–0.92)   | 0.57 (0.34–0.73) |
| Purdue right                 |             | 0.86 (0.82–0.90)  | 0.80 (0.67–0.89)  | 0.83 (0.73–0.90)   | 0.36 (0.09–0.58) |
| Purdue left                  |             | 0.85 (0.80–0.89)  | 0.72 (0.54–0.84)  | 0.82 (0.71–0.89)   | 0.55 (0.32–0.72) |
| Purdue both                  |             | 0.85 (0.80–0.89)  | 0.73 (0.56–0.84)  | 0.81 (0.69–0.88)   | 0.52 (0.28–0.70) |
| Purdue assembly              |             | 0.91 (0.88–0.93)  | 0.83 (0.71–0.90)  | 0.79 (0.67–0.87)   | 0.79 (0.65–0.87) |
| <b>Position tolerance</b>    | <b>0.37</b> |                   |                   |                    |                  |
| Sitting                      |             | 1.0 (1.0–1.0)     | 1.0 (1.0–1.0)     | a                  | a                |
| Standing                     |             | 0.49 (0.36–0.60)  | 0.54 (0.30–0.72)  | a                  | a                |
| Stooping                     |             | 0.37 (0.23–0.50)  | 0.13 (–0.16–0.41) | 0.46 (0.23–0.64)   | 1.0 (1.0–1.0)    |
| Crouching                    |             | 0.80 (0.73–0.85)  | 0.44 (0.17–0.65)  | 0.91 (0.85–0.94)   | a                |
| Kneeling                     |             | 0.12 (–0.04–0.27) | 0.25 (–0.04–0.50) | –0.10 (–0.35–0.16) | a                |

ICC, intraclass correlation coefficient; ID, intellectual disability group; MI, mental illness group; ND, no disability group; NTUH PCE, National Taiwan University Hospital Physical Capacity Evaluation.

<sup>a</sup>ICCs cannot be computed because scale or part of scale has zero variance.

## Discussion

Although several PCEs are commercially available, none have been studied for reliability and validity in people with ID and MI. For the needs of vocational rehabilitation, the NTUH PCE was developed. Generally, the results supported reliability of the NTUH PCE for people with ID and MI and acceptable validity; however, a few modifications, discussed below, are needed.

### Reliability

Lifting, carrying, pushing, and pulling subtests of the NTUH PCE adopted a progressive, isoinertial approach. This approach places the evaluator in control, and tasks are stopped when biomechanical signs of maximal effort are observed, such as accessory muscle usage and counterbalancing (i.e. altered biomechanics judged as being unsafe) (Mayer *et al.*, 1988; Isernhagen, 1992). Although conducting these subtests requires the evaluator's judgment regarding safety and maximal effort, some variance is expected between evaluators with repeated measures. In addition, especially for people

with ID or MI, variations in participant performance because of motivation, wellness on the day of the evaluation, or interactions between the participant and evaluator may influence results. Thus, test–retest reliability has been viewed as the most important indicator of reliability in these subtests (King *et al.*, 1998; Lechner *et al.*, 1991; Velozo, 1993; Wickstrom, 1990). Substantial levels of test–retest reliability in assessing safety and effort level are reported for people without disability or musculoskeletal disorders (Durand *et al.*, 2004; Gross and Battie, 2002; Isernhagen *et al.*, 1999; Lechner *et al.*, 1991; Lygren *et al.*, 2005; Mayer *et al.*, 1988; Reneman *et al.*, 2004; Reneman *et al.*, 2002; Tuckwell *et al.*, 2002). In addition to revealing similar results for people without disabilities, this study indicates test–retest reliability for people with ID and MI. The ICC values of lifting, carrying, pushing, and pulling subtests of the NTUH PCE ranged from 0.69 to 0.94. It seems that the reliability of safe, maximum effort determinations seems to be acceptable for clinical use for all populations when an operational definition is provided to the rater.

**Table 3 Known group validity of the NTUH PCE by ANOVA**

| Subtest                      | ID (1)       | MI (2)       | ND (3)       | F        | Comparison   |
|------------------------------|--------------|--------------|--------------|----------|--------------|
| <b>Strength</b>              |              |              |              |          |              |
| Floor to knuckle lifting     | 21.4 ± 7.0   | 23.9 ± 9.5   | 29.4 ± 10.8  | 9.06***  | 1 < 3, 2 < 3 |
| Knuckle to shoulder lifting  | 17.2 ± 5.7   | 18.7 ± 7.4   | 23.4 ± 7.7   | 10.02*** | 1 < 3, 2 < 3 |
| Shoulder to overhead lifting | 12.4 ± 4.2   | 13.9 ± 6.1   | 17.8 ± 6.7   | 10.76*** | 1 < 3, 2 < 3 |
| Bilateral carrying           | 21.8 ± 6.1   | 23.9 ± 8.4   | 30.7 ± 10.4  | 13.97*** | 1 < 3, 2 < 3 |
| Pushing                      | 3.2 ± 0.9    | 3.3 ± 0.8    | 3.7 ± 0.5    | 7.29**   | 1 < 3, 2 < 3 |
| Pulling                      | 2.8 ± 0.8    | 3.1 ± 0.8    | 3.5 ± 0.6    | 8.54***  | 1 < 3, 2 < 3 |
| Power grip left              | 21.3 ± 7.5   | 25.7 ± 9.6   | 29.9 ± 8.2   | 11.76*** | 1 < 2 < 3    |
| Power grip right             | 22.4 ± 7.5   | 27.9 ± 9.8   | 32.2 ± 9.5   | 13.63*** | 1 < 2 < 3    |
| Lateral pinch left           | 6.0 ± 1.8    | 6.9 ± 1.8    | 7.2 ± 1.8    | 5.42**   | 1 < 2, 1 < 3 |
| Lateral pinch right          | 6.4 ± 1.9    | 7.4 ± 2.0    | 7.7 ± 1.9    | 5.52**   | 1 < 2, 1 < 3 |
| <b>Mobility</b>              |              |              |              |          |              |
| Walking                      | 66.7 ± 10.3  | 70.5 ± 10.3  | 74.8 ± 9.0   | 7.72**   | 1 < 3, 2 < 3 |
| Ladder climbing              | 2.2 ± 0.8    | 2.8 ± 0.8    | 3.5 ± 0.6    | 37.81*** | 1 < 2 < 3    |
| Stair climbing               | 94.6 ± 17.0  | 102.0 ± 20.1 | 115.4 ± 22.3 | 13.13*** | 1 < 3, 2 < 3 |
| Balancing                    | 20.3 ± 16.0  | 12.3 ± 6.5   | 7.9 ± 1.8    | 19.02*** | 1 > 2 > 3    |
| Crawling                     | 21.0 ± 10.9  | 13.9 ± 6.9   | 9.9 ± 3.4    | 25.46*** | 1 < 2 < 3    |
| Repetitive stooping          | 112.6 ± 31.2 | 89.7 ± 22.6  | 66.7 ± 9.9   | 47.46*** | 1 < 2 < 3    |
| Repetitive crouching         | 111.2 ± 36.3 | 81.3 ± 18.0  | 61.3 ± 9.7   | 47.46*** | 1 < 2 < 3    |
| <b>Hand coordination</b>     |              |              |              |          |              |
| CMDT placing                 | 202.3 ± 54.9 | 150.2 ± 30.3 | 128.3 ± 15.8 | 50.28*** | 1 > 2 > 3    |
| CMDT turning                 | 209.0 ± 95.5 | 131.8 ± 40.3 | 103.0 ± 13.5 | 40.70*** | 1 > 2 > 3    |
| Purdue right                 | 10.8 ± 2.9   | 13.7 ± 3.0   | 16.3 ± 1.6   | 50.90*** | 1 < 2 < 3    |
| Purdue left                  | 9.8 ± 2.7    | 13.1 ± 2.8   | 15.2 ± 1.9   | 53.77*** | 1 < 2 < 3    |
| Purdue both                  | 8.4 ± 2.4    | 10.9 ± 2.3   | 13.0 ± 1.4   | 57.82*** | 1 < 2 < 3    |
| Purdue assembly              | 17.9 ± 5.7   | 28.4 ± 8.8   | 38.8 ± 6.6   | 94.91*** | 1 < 2 < 3    |
| <b>Position tolerance</b>    |              |              |              |          |              |
| Sitting                      | 4.0 ± 0.1    | 4.0 ± 0.0    | 4.0 ± 0.0    | 1.16     |              |
| Standing                     | 3.9 ± 0.3    | 4.0 ± 0.1    | 4.0 ± 0.0    | 4.46*    | 1 < 2, 1 < 3 |
| Stooping                     | 3.5 ± 0.7    | 3.8 ± 0.5    | 4.0 ± 0.1    | 11.63*** | 1 < 3, 2 < 3 |
| Crouching                    | 3.7 ± 0.5    | 3.8 ± 0.7    | 4.0 ± 0.0    | 4.10*    | 1 < 3        |
| Kneeling                     | 3.8 ± 0.4    | 3.8 ± 0.4    | 4.0 ± 0.0    | 5.25**   | 1 < 3, 2 < 3 |

ANOVA, analysis of variance; ID, intellectual disability; MI, mental illness; ND, no disability; NTUH PCE, National Taiwan University Hospital Physical Capacity Evaluation.

\* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ .

Few studies report the test–retest reliability on other subtests in a PCE battery. A study of test–retest reliability on nine tasks of the Physical Work Performance Evaluation revealed that sitting, stair climbing, and walking showed poor agreement, and that standing, kneeling, and repetitive squatting showed fair-to-good agreement (Tuckwell *et al.*, 2002). Reneman *et al.* (2004) reported test–retest reliability of the Isernhagen Work Systems Functional Capacity Evaluation in healthy adults and results showed acceptable reliability on most subtests with the exception of the shuttle walk and forward bend subtests (Reneman *et al.*, 2004). This study reported similar results. Crude analyses of the NTUH PCE reveal acceptable test–retest reliability of most, but not all, subtests. The test–retest reliability of the strength and mobility sections is good to excellent. These subtests are safe and reliable for all studied populations to evaluate an individual's physical ability. Some problematic subtests (with ICC values  $< 0.60$ ) were found in the hand coordination section for the ND group and in the position tolerance section for all participants in ID and MI groups.

Detailed analyses of these problematic subtests in each group indicated that considerable within participant variances between two testing sessions were found.

Mean performance of hand coordination for the ND group was significantly better in the second testing session ( $t = 3.03$ – $7.71$ ,  $P < 0.001$ ). The differences likely affected the test–retest reliability of these subtests. Although these subtests were standardized tests, to shorten the administration time of the NTUH PCE, the number of test trials was limited in these subtests. As mentioned in the manual (Lafayette Instrument Company, 1998, 1999), an increased number of test trials increased the test–retest reliability of the CMDT and the Purdue Pegboard test. Changes in future testing methods were made after the testing procedure presented in the manual, including increasing test trials from one trial to two trials for the Purdue Pegboard test, and from two trials to three trials for the CMDT. In addition, the retest criteria were set for each test session. For instance, if the difference between two trials of the subtest for the Purdue Pegboard test exceeds 3, the subtest may be performed again after resting.

With the exception of sitting, all subtests within the position tolerance section had unacceptable test–retest reliability. For some subtests, such as sitting and standing (MI and ND) and crouching and kneeling (ND), ICC values could not be calculated because there were no

differences between two testing sessions. Therefore, only crouching of the MI group had acceptable reliability. This indicates that most subtests in the position tolerance section are unreliable for use in the ID and MI groups. Variations in a participant's performance between days may have been because of lack of motivation, boredom, or sense of meaninglessness. The testing procedure needs to be changed to maintain the attention of participants with disability.

### Internal consistency

The NTUH PCE showed internal consistency in three sections, but not within the position tolerance section ( $\alpha = 0.37$ ). Detailed analyses of the position tolerance section found that subtests measured different dimensions of body motion, including sitting, standing, stooping, crouching, and kneeling. This lower value was not unexpected. The four sections seemed consistent with theoretical expectations of the construct of the physical capacity. All subtests were maintained in the original sections for content validity and for the ability to draw inferences from results despite a low internal correlation factor in the position tolerance section.

### Known group validity of the National Taiwan University Hospital Physical Capacity Evaluation

The NTUH PCE was able to differentiate among the ID, MI, and the ND groups in all subtests except sitting. Participants with ID or MI have no obvious physical problems based on appearance alone. The literature, however, shows that because of side effects of medications, psychomotor retardation because of illness, lack of motivation, or sedentary lifestyle, persons with ID or MI have poor physical fitness in general and motor performance deficits in speed, balance, hand coordination, and strength (Bracegirdle, 2002; Kioumourtzoglou *et al.*, 1994; Stanish *et al.*, 2006; Wykoff, 1993). Results of the present study indicate that performance of ID and MI participants in strength, mobility, hand coordination, and position tolerance tests was significantly poorer than participants in the ND group, especially in hand coordination and mobility sections. Physical capacity is considered to be a major barrier for people with disabilities to gain or maintain employment (Menard and Hoens, 1994; Silverstein *et al.*, 1991). Results of this study indicate that it is important to assess physical capacity in determining work ability and limitations for people with ID or MI.

### Study limitations

In this study, the evaluator was not blinded to the results of the first session of testing. Theoretically, this could affect the evaluator's judgment and the participant's test results. The interval between testing sessions was, however, 7–21 days, and evaluators performed two or more PCEs per day as well as routine clinical work. It was assumed that the interval between two testing sessions

would be long enough to enable evaluators and participants to forget testing results from the first session. It was noted that variations in a participant's performance between two testing sessions may have been because of issues of motivation, such as a sense of boredom or meaninglessness of the subtests. Another potential contribution to the observed variability is a learning effect in people participating in two testing sessions. Comparison of mean scores between sessions revealed an increase in second session scores for most subtests, indicating that a learning effect likely did occur. Despite this effect, test-retest ICC values were acceptable, with the exception of the subtests in the hand coordination section for the ND group.

### Conclusion

This study describes the initial testing of the NTUH PCE in persons with ID and MI. The results indicate acceptable test-retest reliability of most subtests of the NTUH PCE, with the exception of the standing, stooping, crouching, and kneeling subtests. Some changes in testing procedures for the hand coordination and position tolerance sections are needed. With evidence for internal consistency, known group validity, and reliability, the NTUH PCE provides additional support for the applicability of this instrument in occupational and vocational rehabilitation research and daily practice and is a good instrument for assessing an individual's physical abilities. Future research on the psychometric properties of the NTUH PCE may include concurrent validity, predictive validity, and responsiveness.

### Acknowledgement

This study was supported by the National Science Council (grant number NSC 89-2614-B-002-014-M47 and NSC 95-2314-B-002-223-MY3).

### References

- Bracegirdle H (2002). Developing physical fitness to promote mental health. In: Creek J, editor. *Occupational therapy and mental health*. 3rd ed. New York: Churchill Livingstone. pp. 209–225.
- Brouwer S, Reneman MF, Dijkstra PU, Groothoff JW, Schellekens JMH, Goeken LNH (2003). Test-retest reliability of the Isernhagen Work Systems Functional Capacity Evaluation in patients with chronic low back pain. *J Occup Rehabil* 13:207–218.
- Cohen J (1988). *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Cronbach LJ (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* 16:297–334.
- Deen M, Gibson L, Strong J (2002). A survey of occupational therapy in Australian work practice. *Work* 19:219–230.
- Durand MJ, Loisel P, Poitras S, Mercier R, Stock SR, Lemaire J (2004). The interrater reliability of a functional capacity evaluation: the Physical Work Performance Evaluation. *J Occup Rehabil* 14:119–129.
- Fayers PM, MaChin D (2000). *Quality of life: assessment, analysis and interpretation*. New York: John & Wiley. pp. 45–116.
- Gross DP (2004). Measurement properties of performance-based assessment of functional capacity. *J Occup Rehabil* 14:165–174.
- Gross DP, Battie MC (2002). Reliability of safe maximum lifting determinations of a functional capacity evaluation. *Phys Ther* 82:364–371.
- Innes E, Straker L (1999a). Reliability of work-related assessments. *Work* 13:107–124.

- Innes E, Straker L (1999b). Validity of work-related assessments. *Work* **13**: 125–152.
- Isernhagen SJ (1992). Functional capacity evaluation: rationale, procedure, utility of the kinesiophysical approach. *J Occup Rehabil* **2**:157–168.
- Isernhagen SJ, Hart DL, Matheson LM (1999). Reliability of independent observer judgments of level of lift effort in a kinesiophysical physical functional capacity evaluation. *Work* **12**:145–150.
- Jacobs K (1991). *Occupational therapy: Work-related programs and assessments*. 2nd ed. Boston, Massachusetts: Little, Brown and Company; pp. 17.
- Jones T, Kumar S (2003). Functional capacity evaluation of manual materials handlers: a review. *Disabil Rehabil* **25**:179–191.
- Jundt J, King PM (1999). Work rehabilitation programs: a 1997 survey. *Work* **12**:139–144.
- King PM, Tuckwell N, Barrett TE (1998). A critical review of functional capacity evaluations. *Phys Ther* **78**:852–866.
- Kioumourtoglou E, Batsiou S, Theodorakis Y, Mauromatis G (1994). Selected motor skills of mentally retarded and nonretarded individuals. *Percept Mot Skills* **78**:1011–1015.
- Lafayette Instrument Company (1998). *The Complete Minnesota Dexterity Test: Examiner's manual*. Lafayette, Indiana: Lafayette Instrument Company.
- Lafayette Instrument Company (1999). *The Purdue Pegboard Test: Examiner's manual*. Lafayette, Indiana: Lafayette Instrument Company.
- Landis JR, Koch GG (1977). The measurement of observer agreement for categorical data. *Biometrics* **33**:159–174.
- Lechner DE, Roth D, Straaton K (1991). Functional capacity evaluation in work disability. *Work* **1**:37–47.
- Lygren H, Dragesund T, Joensen J, Ask T, Moe-Nilssen R (2005). Test-retest reliability of the progressive isoinertial lifting evaluation (PILE). *Spine* **30**:1070–1074.
- Mayer TG, Barnes D, Kishino ND, Nichols G, Gatchel RJ, Mayer H, Mooney V (1988). Progressive isoinertial lifting evaluation. I. a standardized protocol and normative database. *Spine* **13**:993–997.
- Menard MR, Hoens AM (1994). Objective evaluation of functional capacity: medical, occupational, and legal settings. *J Orthop Sports Phys Ther* **19**:249–260.
- Portney LG, Watkin MP (2000). *Foundations of clinical research: applications to practice*. 2nd ed. New Jersey, Pennsylvania: Prentice Hall.
- Reneman MF, Jaegers SMHJ, Westmaas M, Göeken LNH (2002). The reliability of determining effort level of lifting and carrying in a functional capacity evaluation. *Work* **18**:23–27.
- Reneman MF, Brouwer S, Meinema A, Dijkstra PU, Geertzen JHB, Groothoff JW (2004). Test-retest reliability of the Isernhagen Work Systems Functional Capacity Evaluation in healthy adults. *J Occup Rehabil* **14**:295–305.
- Silverstein B, Kilgore KM, Fisher WP, Harley JP, Harvey RF (1991). Applying psychometric criteria to functional assessment in medical rehabilitation: I exploring unidimensionality. *Arch Phys Med Rehabil* **72**:631–637.
- Stanish HI, Temple VA, Frey GC (2006). Health-promoting physical activity of adults with mental retardation. *Ment Retard Dev Disabil Res Rev* **12**:13–21.
- Task force on standards for measurement in physical therapy (1991). Standards for tests and measurements in physical therapy practice. *Phys Ther* **71**:589–622.
- Tuckwell NL, Straker L, Barrett TE (2002). Test-retest reliability on nine tasks of the Physical Work Performance Evaluation. *Work* **19**:243–253.
- United States Department of Labor, Employment and Training Administration (1993). *Selected characteristics of occupations defined in the revised Dictionary of Occupational Titles*. Baton Rouge, Louisiana: U.S. Government Printing Office. pp. C1–C3.
- Velozo CA (1993). Work evaluations: Critique of the state of the art of functional assessment of work. *Am J Occup Ther* **47**:203–209.
- Wickstrom R (1990). Functional capacity testing. In: Scheer SJ, editor. *Multidisciplinary perspectives in vocational assessment of impaired workers*. Gaithersburg, Maryland: Aspen Publishers Inc. pp. 73–88.
- Wykoff W (1993). The psychological effects of exercise on non-clinical and clinical populations of adult women: a critical review of the literature. *Occup Ther Ment Health* **12**:69–106.