# The standardised Malay version of purdue pegboard test: Content validity and test-retest reliability testing

**Yin Sin Poo, Siaw Chui Chai** ⓘ **, Poh Im Goh, Masne Kadar** ⓘ **and Nor Afifi Razaob@Razab** ⓘ

## Abstract

**Introduction:** Purdue Pegboard Test (PPT) is a valid and reliable instrument for measuring hand dexterity among individuals with or without medical conditions. In the Southeast Asia region where Malay is widely spoken, there is a need to have a Malay translation of Purdue Pegboard Test. This study aimed to translate the PPT into the Malay version (PPT-M) and to determine the content validity and test-retest reliability of this translated version.

**Methods:** This study involved: (1) four English teachers (translators) for forward and backward translation procedures; (2) 10 experts in the field of occupational therapy (expert reviewers) for content validity testing; and (3) 60 undergraduate students (participants) for test-retest reliability testing.

**Results:** PPT-M had excellent content validity with Item-Content Validity Index = 0.9–1.0, Scale-Content Validity Index/Average = 0.93–0.95, and a slightly lower Scale-Content Validity Index/Universal Agreement = 0.25–0.75. Test-retest reliability for 3-trial administration ($n = 30$; Intraclass Correlation Coefficients, ICCs = 0.76–0.85; good) was higher compared to 1-trial administration ($n = 30$; ICCs = 0.34–0.46; poor) for all subtests. Both trial administrations were mostly affected by systematic errors, especially practice effect as the retests gave higher scores. Random errors mostly affected Subtest 3 of 1-trial administration, evident by its Minimal Detectable Change Percent values = 30.84% that fell beyond the acceptable range.

**Conclusion:** PPT-M has the potential to be a valuable instrument for measuring hand dexterity among Malay speaking individuals especially when the 3-trial administration is used.

## Introduction

Hand dexterity is routinely being measured by occupational therapists working with individuals with upper limb conditions. It is also a commonly measured parameter during health screening of general populations. When administering a hand dexterity test, the therapist needs to strictly adhere to the test's standardised instructions for facilitation of valid and reliable result attainment, efficient communication, and routinization (Wear, 2015).

Purdue Pegboard Test is a hand dexterity test that was initially developed for use among industrial workers in various job settings (Triffin, 1987; Tiffin & Asher, 1948).

Psychometric studies on individuals with or without health condition supported that PPT is a valid and reliable instrument

Occupational Therapy Programme, Centre for Rehabilitation and Special Needs Studies, Faculty of Health Sciences, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia

**Corresponding author:**
Siaw Chui Chai, Occupational Therapy Programme, Centre for Rehabilitation and Special Needs Studies, Faculty of Health Sciences, Universiti Kebangsaan Malaysia, Jalan Raja Muda Abdul Aziz, Kuala Lumpur 50300, Malaysia.
Email: sc.chai@ukm.edu.my

for testing the hand dexterity (Alderhag, Jonsson, Littorin, & Ansved, 2008; Amirjani et al., 2011; Buddenberg & Davis, 2000; Chen & Ringenbach, 2015; Desrosiers et al., 1995; Doyen & Carlier, 2002; Gallus & Mathiowetz, 2003; Lee et al., 2013; Lindstrom-Hazel & VanderVlies Veenstra, 2015; Mathiowetz et al., 1986; Muller et al., 2011; Proud et al., 2019; Proud et al., 2020; Reddon, Gill, Gauk, & Maerz, 1998; Shahar et al., 1998; Tiffin & Asher, 1948; Wilson, Wilson, Iacoviello, & Risucci, 1982). PPT can be conducted using 1-trial and 3-trial administrations with the latter being a more reliable administering method (Alderhag et al., 2008; Buddenberg & Davis, 2000; Chen & Ringenbach, 2015; Tiffin & Asher, 1948).

Given PPT is constructed in English, its usefulness among non-English speakers is often not fully appreciated. Although spontaneously translating the test into a language that an individual comprehends is a practical solution, violation in terms of the psychometrics warrants careful consideration. Reliability of the spontaneously translated test can be influenced by many factors, e.g., the individual's understanding, the rater's practice effects, and other psychological elements. Fatigue, forgetfulness, and practice effect may also result in selection of preferred or important words, along with unintentional skip of some words (Lim & Chai, 2020). Hence, to improve efficiency and usefulness of the testing process and result, translating the test into different languages in a standardised manner is needed (Sousa & Rojjanasrirat, 2011). Furthermore, given there are 4 subtests that have instructions, chances for translation error and inconsistency to occur is high. The Hong Kong Chinese Version of the Jebsen Hand Function Test is an excellent example of standardised translated performance test (Li-Tsang et al., 2004).

Malay is the national language of Malaysia, Brunei, Indonesia, and Singapore. It is also spoken by Malay individuals residing in other parts of Southeast Asia, including Southern Thailand, the Philippines, Central Eastern Sumatra, Riau Islands, and parts of the coast of Borneo (Razak, Madison, Siow, & Aziz, 2010). In this region, English either serves as a second language or used limitedly in instances related to diplomacy, tourism, and foreign trade (Hashim & Low, 2014). It is this understanding that converged our attention to the need of having a Malay translation to facilitate testing process and result among individuals residing in this region. As an important screening tool for general populations and a measurement tool for clinical usage, this emphasis should be given to PPT. As such, the aims of this study were to translate the PPT from English into the Malay version (PPT-M) and to determine the content validity and test-retest reliability of this translated version.

## Methods

This study was divided into three phases: (1) Translation of the PPT into the Malay version (PPT-M); (2) content validity testing of PPT-M; and (3) test-retest reliability testing of PPT-M. Ethics approval to conduct the study

was granted by Research Ethics Committee, Universiti Kebangsaan Malaysia. All translators, expert reviewers, and participants read the information sheet and signed the informed consent form before participating in the study.

### Instrumentation

The PPT set consists of a User Instruction (Model 32020A), a test board, and score sheets. There are five subtests: Subtest 1 (Dominant hand), Subtest 2 (Non-dominant hand), Subtest 3 (Both hands), Subtest 4 (Dominant hand + Non-dominant hand + Both hands), and Subtest 5 (Assembly). Subtests 1 and 2 involve placing as many pins as possible down the row within 30 seconds. Subtest 3 is a bimanual test that involves placing as many pins as possible with both hands simultaneously down the rows within 30 seconds. Subtests 1, 2, and 3 are scored based on the total numbers of pins placed within the time limit. Subtest 4 is a mathematical calculation by adding the scoring of Subtests 1, 2, and 3. Subtest 5 involves placing pins, washers, and collars with both hands simultaneously in 60 seconds. This subtest is scored based on the total numbers of parts assembled. A higher score signifies better hand dexterity (Triffin, 1987; Tiffin & Asher, 1948).

### Participants

The translation phase involved recruitment of four English language teachers as translators based on the criteria that they must hold a bachelor's degree in Teaching English as a Second Language (TESL) and had no prior exposure to the original English version of PPT. All the translators are of Malay descents and speak fluent Malay (Table 1).

The content validity testing phase involved recruitments of 10 experts (male = 5; female = 5) in the field of occupational therapy as expert reviewers by purposive sampling. They were predominantly aged 30–39 years ($n = 6$); holding a bachelor's degree ($n = 6$); and with 1–9 years of work experience ($n = 5$) (Table 1). All of them are of Malay descent and speak fluent Malay (Table 1).

The test-retest reliability phase involved recruitment of 60 undergraduate students by convenience sampling (1-trial administration, $n = 30$; 3-trial administration, $n = 30$). The decision of using undergraduate students as participants was made upon the basis that PPT was originally constructed for healthy individuals and since then, remains important for dexterity testing among this population. Majority of participants were female ($n = 39$, 65%); Malay ($n = 57$, 95%); right hand dominant ($n = 54$, 90%); studying at Year 3 ($n = 26$, 43.3%); and in the health science field ($n = 42$, 70%) (Table 2). They were selected based on the criteria: no prior exposure to the original English version of PPT; no significant uncorrected vision impairment; and no hand injury. Sample size was determined using the formula for minimum sample size ($n$) estimation using the PASS software (Bujang & Baharum, 2017).

**Table 1.** Demographics of translators and expert reviewers.

| | Translator | | | |
|---|---|---|---|---|
| | Forward Translation (*n* = 2) | Backward Translation (*n* = 2) | Total (*N* = 4) | Expert Reviewer (*N* = 10) |
| Gender | | | | |
| Male | 1 | 0 | 1 | 5 |
| Female | 1 | 2 | 3 | 5 |
| Age (Year) | | | | |
| 20–29 | 1 | 1 | 2 | 2 |
| 30–39 | 1 | 1 | 2 | 6 |
| 40–49 | 0 | 0 | 0 | 1 |
| 50–59 | 0 | 0 | 0 | 1 |
| Academic qualification | | | | |
| Bachelor's degree | 2 | 2 | 4 | 6 |
| Master's degree | 0 | 0 | 0 | 1 |
| Doctoral degree | 0 | 0 | 0 | 3 |
| Working experience (Year) | | | | |
| 1–9 | 1 | 1 | 2 | 5 |
| 10–19 | 1 | 1 | 2 | 3 |
| 20–29 | 0 | 0 | 0 | 2 |

**Table 2.** Demographics of participants.

| | 1-Trial (*n* = 30) | 3-Trial (*n* = 30) | Total (*N* = 60) |
|---|---|---|---|
| Age (mean ± standard deviation) | 21.53 ± 1.14 | 22.37 ± 1.22 | 21.95 ± 1.24 |
| Gender | | | |
| Male | 15 | 6 | 21 (35%) |
| Female | 15 | 24 | 39 (65%) |
| Race | | | |
| Malay | 27 | 30 | 57 (95%) |
| Indian | 2 | 0 | 2 (3.3%) |
| Others | 1 | 0 | 1 (1.7%) |
| Hand dominance | | | |
| Right | 28 | 26 | 54 (90%) |
| Left | 2 | 4 | 6 (10%) |
| Year of study | | | |
| 1 | 10 | 2 | 12 (20%) |
| 2 | 4 | 5 | 9 (15%) |
| 3 | 10 | 16 | 26 (43.3%) |
| 4 | 6 | 7 | 13 (21.7%) |
| Area of study | | | |
| Health sciences | 21 | 21 | 42 (70%) |
| Pharmacy | 7 | 6 | 13 (21.7%) |
| Dentistry | 2 | 3 | 5 (8.3%) |

## Procedures

*Translation.* The study received permission to translate the PPT from English into Malay from the Lafayette Instrument Company, USA. Translation was only performed on the standardised instructions of Subtests 1, 2, 3, and 5 (Model 32020A). General explanation about PPT was not translated.

Translation from English into Malay was done via forward and backward translation procedures that was adapted from the World Health Organization: Process of Translation and Adaptation of Instruments Guideline (2017). The forward translation involved translating the PPT instructions from English into Malay by two translators. After the forward translation, the same translators harmonised the translations to

produce the initial version of PPT-M. The backward translation involved translating the instructions from Malay into English by another two translators who were blinded to the original documents. The backward translations were then compared to the original English version of PPT. With some modifications, the two translations were merged to produce the pre-final version of PPT-M.

*Content validity.* For this content validity testing, each expert reviewer was required to read, understand, and compare the PPT-M to the original English version of PPT. They were required to score and provide feedback on the content validity feedback form given to them. The four criteria of content validity index (CVI), i.e., relevance, clarity, simplicity, ambiguity was used to test the content validity of PPT-M. CVI can be divided into: (1) Item-Content Validity Index (I-CVI), i.e., computed based on each item on an instrument; and (2) Scale-Content Validity Index (S-CVI), i.e., computed for the overall scale (Lynn, 1986; Polit et al., 2007). CVI results and feedback provided by the expert reviewers were discussed and harmonised among the research team members and integrated into the production of the final version of PPT-M.

*Test-retest reliability.* The PPT-M testing, including arrangements of the table, test materials, and timing for the test was administered by strictly adhering to the PPT guidelines. The 1-trial administration was instructed by the first author and the 3-trial administration was instructed by the third author, respectively. For 3-trial administration, the final score was the average score of the three trials. To accommodate extracurricular activities, emergencies, and class schedules of the participants, the retest for 1-trial and 3-trial administrations were completed within 7.47 ± 1.93 and 7.73 ± 2.73 days, respectively. There was no difference in terms of the retest interval between 1-trial and 3-trial administrations ($p = 0.66$).

## Data analysis

CVI result obtained was interpreted as: excellent if I-CVI ≥0.78, Scale-Content Validity Index/Average (S-CVI/Ave) ≥ 0.90, and Scale-Content Validity Index/Universal Agreement (S-CVI/UA) ≥ 0.8 (Polit et al., 2007; Shi et al., 2012).

Test-retest reliability as analysed using a two-way mixed effect model with absolute agreement and expressed using Intraclass Correlation Coefficient (ICC) was interpreted as: poor if ICC <0.50; moderate if ICC = 0.50 – 0.75; good if ICC = 0.75 – 0.90; and excellent if ICC >0.90 (Koo & Li, 2016). Standard Error of Measurement (SEM) as an estimation of precision of measurement was calculated using the square root of the mean square error from the Analysis of Variance with small value indicates an acceptably small range of measurement inconsistency (Denegar & Ball, 1993). Minimal Detectable Change (MDC) as an expression of the minimal threshold beyond the random measurement error was calculated using the formula: SEM × 1.96 × $\sqrt{2}$. Consistently, MDC% as an expression of the relative amount of random measurement error was calculated using the formula: (MDC/Mean) × 100% with a value of <30 to be interpreted as acceptable and <10 as excellent (Smidt et al., 2002). Difference in the initial test and retest scores as analysed using the Paired samples $t$ test with significant level, α = 0.05 was expressed using effect size, $d$, with 0.2 = small, 0.5 = medium, and 0.8 = large (Cohen, 1988).

## Results

### Content validity

PPT-M had excellent content validity with I-CVI = 0.9–1.0 and S-CVI/Ave = 0.93–0.95; although the S-CVI/UA = 0.25–0.75 could be considered as a little lower (Table 3). Most of the expert reviewers suggested changing the translated English phrase of "pick up one pin at a time" in Subtests 1 and 2 to a more grammatically correct Malay phrase so that the PPT-M can be understood by the local people. After discussing with the research team members, the decision was to translate "Pick up one pin at a time" to "*Ambil satu pin pada satu masa*". Therefore, Subtest 1: "Pick up one pin at a time with your right hand from the right-handed cup…" was translated to "*Ambil satu pin pada satu masa dengan tangan kanan anda daripada cawan sebelah kanan…*" and Subtest 2: "Pick up one pin at a time with your left hand from the left-handed cup…" was

**Table 3.** Content validity of the standardised Malay version of purdue pegboard test.

| | I-CVI | | | | S-CVI | |
| --- | --- | --- | --- | --- | --- | --- |
| | Subtest 1 | Subtest 2 | Subtest 3 | Subtest 5 | S-CVI/Ave | S-CVI/UA |
| Relevance | 0.9 | 0.9 | 1.0 | 1.0 | 0.95 | 0.50 |
| Clarity | 0.9 | 0.9 | 0.9 | 1.0 | 0.93 | 0.25 |
| Simplicity | 0.9 | 1.0 | 1.0 | 1.0 | 0.95 | 0.75 |
| Ambiguity | 0.9 | 0.9 | 1.0 | 1.0 | 0.95 | 0.50 |

I-CVI = Item-Content Validity Index; S-CVI/Ave = Scale-Content Validity Index/Average; S-CVI/UA: Scale-Content Validity Index/Universal Agreement; Subtest 1 = Dominant hand (30 s); Subtest 2 = Non-dominant hand (30 s); Subtest 3 = Both hands (30 s); Subtest 5 = Assembly (1 min).

translated to "*Ambil satu pin pada satu masa dengan tangan kiri anda daripada cawan sebelah kiri…*" Given no major changes were required based on CVI results and feedback provided by the expert reviewers, only the research team members were involved in discussion and harmonisation of the production of final PPT-M.

## Test-retest reliability

The 1-trial administration had poor test-retest reliability with ICCs = 0.34–0.46 with respective SEM of 1.40 and 3.25 for all subtests. Subtest 4 scored the highest test-retest reliability and Subtest 2 scored the lowest test-retest reliability. For all subtests, PPT-M had MDC = 3.45–11.69 and MDC% = 21.62%–30.84%, respectively. Only Subtest 3 had MDC% that fell beyond the acceptable range (30.84%). There were significant differences in terms of PPT-M scores between the initial test and retest for all subtests, $t = -5.99$ to $-3.20$, $p < 0.05$ with medium to large effect sizes, $d = 0.59 - 0.93$.

The 3-trial administration had good test-retest reliability with ICCs = 0.76–0.85 with respective SEM of 0.90 and 0.60 for all subtests. Subtest 3 scored the highest test-retest reliability and Subtest 1 scored the lowest test-retest reliability. For all subtests, PPT-M had MDC = 1.65–5.51 and MDC% = 10.61%–16.14%, respectively. There were significant differences in terms of PPT-M scores between the initial test and retest for all subtests, $t = -6.75$ to $-3.67$, $p \leq 0.001$., with medium effect sizes, $d = 0.43 - 0.73$. When comparing descriptively in terms of ICC, SEM, MDC, and MDC%, the 3-trial administration showed higher test-retest reliability than the 1-trial administration (Table 4).

## Discussion

This study highlighted the process of translating the PPT into PPT-M along with its content validity and test-retest reliability testing. The entire process was conducted in a robust manner by adhering to standardised translating guidelines as well as content validity and test-retest reliability procedures. We conducted the test-retest reliability using both 1-trial and 3-trial administrations.

PPT-M had excellent content validity based on the I-CVI and S-CVI/Ave results, signifying that the PPT-M has sufficient quality and validity. The low overall CVI using universal agreement approach (S-CVI/UA) could be due to the use of 10 expert reviewers as the likelihood of obtaining a total agreement reduces with the increased numbers of participating reviewers (Polit & Beck, 2006; Zamanzadeh et al., 2015). In fact, according to Polit et al. (2007), the calculation of S-CVI/UA is always too stringent such that it discounts the risks of both chance and non-chance disagreements. Non-chance disagreements can happen when the expert is biased or has misunderstood specifications of the construct being tested. Therefore, the use of S-CVI/Ave alone is good enough to reflect the content validity of PPT-M as it not only avoids these problems, but also exemplifies information about each item's performance through the averaging feature.

Given PPT-M is a new foreign language version of PPT and the fact that the PPT was originally designed to test healthy individuals, we therefore decided to initiate the PPT-M test-retest reliability testing on undergraduate students. Testing of PPT-M on a healthy population is particularly important to allow its usage as a general health screening tool for Malay speaking individuals. Findings of this study could serve as a foundation to support our subsequent

**Table 4.** Test-retest reliability of the standardised Malay version of purdue pegboard test.

| Subtest | Test M (SD) | Retest M (SD) | t (p) | d | ICC | 95% CI | SEM | MDC (MDC%) |
|---|---|---|---|---|---|---|---|---|
| 1-Trial administration | | | | | | | | |
| DH | 15.10 (1.52) | 16.00 (1.51) | −3.20 (0.003)* | 0.59 | 0.42 | 0.07–0.67 | 1.25 | 3.45 (22.19%) |
| 2 NDH | 13.33 (1.37) | 14.63 (1.54) | −4.71 (<0.001)* | 0.89 | 0.34 | −0.04–0.63 | 1.40 | 3.87 (27.68%) |
| 3 BH | 11.60 (1.71) | 12.63 (1.65) | −3.47 (0.002)* | 0.61 | 0.45 | 0.09–0.70 | 1.35 | 3.74 (30.84%) |
| #4 DH + NDH + BH | 40.03 (3.44) | 43.27 (4.14) | −5.34 (<0.001)* | 0.85 | 0.46 | −0.03–0.74 | 3.25 | 9.00 (21.62%) |
| 5 Assembly | 37.63 (4.69) | 42.07 (4.86) | −5.99 (<.0001)* | 0.93 | 0.45 | −0.06–0.75 | 4.22 | 11.69 (29.33%) |
| 3-Trial administration | | | | | | | | |
| 1 DH | 15.01 (0.34) | 16.00 (0.28) | −4.24 (<0.001)* | 0.59 | 0.76 | 0.28–0.91 | 0.90 | 2.50 (16.14%) |
| 2 NDH | 14.07 (0.26) | 15.01 (0.24) | −5.46 (<0.001)* | 0.69 | 0.77 | 0.08–0.92 | 0.67 | 1.85 (12.75%) |
| 3 BH | 11.92 (0.24) | 12.49 (0.24) | −3.67 (0.001)* | 0.43 | 0.85 | 0.54–0.94 | 0.60 | 1.65 (13.56%) |
| #4 DH + NDH + BH | 41.07 (0.77) | 43.50 (0.69) | −5.82 (<0.001)* | 0.61 | 0.83 | 0.14–0.95 | 1.62 | 4.49 (10.61%) |
| 5 Assembly | 38.40 (0.86) | 41.87 (0.88) | −6.75 (<0.001)* | 0.73 | 0.79 | −0.06–0.94 | 1.99 | 5.51 (13.73%) |

M = Mean; SD = standard deviation; t = t-value; *p = significant at 0.05; d = effect size; ICC = intra-class correlation coefficient; CI = confidence interval; SEM = standard error of measurement; MDC = minimal detectable change; DH = dominant hand: NDH = non-dominant hand; BH = both hands; # = not an actual test.

studies on a larger sample by involving both Malay speaking healthy individuals and individuals with different health conditions. We conducted the test-retest reliability testing using both 1-trial and 3-trial administrations, aligned with published studies (Alderhag et al., 2008; Buddenberg & Davis, 2000; Chen & Ringenbach, 2015; Tiffin & Asher, 1948). It is important to have test-retest reliability established using both 1-trial and 3-trial administrations. Typically, the 1-trial administration is useful when treatment time is limited or when assessing individuals with disabilities, including multiple sclerosis; whilst the 3-trial administration can be used to test healthy individuals or when highest possible test-retest reliability is needed (Gallus & Mathiowetz, 2003).

With higher ICCs and lower SEMs, the 3-trial administration of PPT-M therefore had higher test-retest reliability compared to the 1-trial administration for all subtests. Although lower ICCs were noted in certain subtests, attention needs to be paid on the SEM analyses, that this consistency of test-retest scores generally happened in a small range and therefore could be accepted. Overall, comparing to the 1-trial administration, the 3-trial administration has greater extent of consistency in yielding the same results when it is being used in the same situation on repeated occasions (Heale & Twycross, 2015) as well as greater degree of correlation and agreement between occasions (Portney & Watkins, 2000). Published studies using individuals with healthy wellbeing (Buddenberg & Davis, 2000; Tiffin & Asher, 1948), muscular dystrophies (Alderhag et al., 2008), and Down syndrome (Chen & Ringenbach, 2015) also shared similar results that 3-trial administration has higher test-retest reliability. Our reliabilities for both 1-trial and 3-trial administration studies are also either comparable to or only slightly lower than those obtained by using healthy children, younger and older adults (Buddenberg & Davis, 2000; Desrosiers et al., 1995; Doyen & Carlier, 2002; Muller et al., 2011; Reddon et al., 1998; Tiffin & Asher, 1948; Wilson et al., 1982), carpal tunnel syndrome (Amirjani et al., 2011), Down syndrome (Chen & Ringenbach, 2015), muscular dystrophies (Alderhag et al., 2008), multiple sclerosis (Gallus & Mathiowetz, 2003), and Parkinson's disease (Proud et al., 2019). This comparability supports the value of PPT-M as a reliable translated version of PPT.

The higher retest scores in both 1-trial and 3-trial administrations with respective effect sizes of medium to large and medium were likely contributed by both systematic and random errors. Our MDC and $t$ test analyses showed that systematic errors affected the test greater compared to random errors. As a performance test, we believe that this systematic error occurred mainly via practice effect as the initial test is often served as a practice for the retest (Weir, 2005). Given most retests of this study were administered approximately after a week, therefore, we speculate that practice effect will not become an issue if the test-retest interval is scheduled longer apart. In fact, in clinical settings, practice effect can be seen as negligent since reassessment is often scheduled after several sessions or several weeks of treatment. A common practice of using many testing instruments in a random order as an assessment strategy could also help to counterbalance practice effects in clinical settings. In our study, this counterbalancing method could not be used because the participants were only assessed using PPT-M. The contribution of random errors, conversely, was minimal as the MDC% values of all subtests fell within the acceptable range except Subtest 3 of 1-trial administration.

Our study is limited to homogeneity of participants and small sample size. The use of university students limits its generalisability to wider populations, especially individuals with various health conditions. As discussed earlier, our subsequent studies will focus on testing the PPT-M on a larger sample by involving both healthy individuals and individuals with different health conditions. Not having the inter-rater reliability tested is another limitation as the test can be administered on the same patient by different therapists or during general health screening, by different raters on different occasions. We recommend performing PPT-M inter-rater reliability testing in future studies and compare the result with the one established for the original PPT (ICC = 0.99) (Lindstrom-Hazel & VanderVlies Veenstra, 2015). To enhance the cross-cultural adaptability, it would be useful to conduct cognitive interviews on PPT-M among various ethnic groups, including the Malay, Chinese, Indian, and Indigenous people so that all the words used are relevant and comprehensible.

## Conclusion

The PPT-M had content validity and test-retest reliability and could potentially be used to test the hand dexterity among Malay speaking individuals in clinical settings or during general health screening. For more reliable testing, it is recommended to use the 3-trial administration procedure.

### Declaration of authorship contribution

All authors have contributed significantly to this paper.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

## ORCID iDs

Siaw Chui Chai https://orcid.org/0000-0003-3228-5795
Masne Kadar https://orcid.org/0000-0002-1497-0548
Nor Afifi Razaob@Razab https://orcid.org/0000-0002-3774-0286

## References

Aldehag, A. S., Jonsson, H., Littorin, S., & Ansved, T. (2008). Reliability of hand function testing instruments in patients with muscular dystrophies. *International Journal of Therapy and Rehabilitation*, 15(5), 211–219. https://doi.org/10.12968/ijtr.2008.15.5.29232

Amirjani, N., Ashworth, N. L., Olson, J. L., Morhart, M., & Chan, K. M. (2011). Validity and reliability of the Purdue Pegboard Test in carpal tunnel syndrome. *Muscle and Nerve*, 43(2), 171–177. https://doi.org/10.1002/mus.21856

Buddenberg, L. A., & Davis, C. (2000). Test–retest reliability of the Purdue pegboard test. *American Journal of Occupational Therapy*, 54(5), 555–558. https://doi.org/10.5014/ajot.54.5.555

Bujang, M. A., & Baharum, N. (2017). A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: A review. *Archives of Orofacial Science*, 12(1), 1–11. Retrived from http://www.dental.usm.my/aos/docs/Vol_12/aos-article-0246.pdf

Chen, C.-C., & Ringenbach, S. D. (2015). A pilot study for test–retest reliability of the Purdue Pegboard Test in adolescents and young adults with down syndrome. *Journal of Motor Learning and Development*, 3(2), 151–157. https://doi.org/10.1123/jmld.2015-0004

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Denegar, C. R., & Ball, D. W. (1993). Assessing reliability and precision of measurement: An introduction to intraclass correlation and standard error of measurement. *Journal of Sport Rehabilitation*, 2(1), 35–42. https://doi.org/10.1123/jsr.2.1.35

Desrosiers, J., Hebert, R., Bravo, G., & Dutil, E. (1995). The Purdue pegboard test: Normative data for people aged 60 and over. *Disability and Rehabilitation*, 17(5), 217–224. https://doi.org/10.3109/09638289509166638

Doyen, A.-L., & Carlier, M. (2002). Measuring handedness: A validation study of bishop's reaching card test. *Laterality*, 7(2), 115–130. https://doi.org/10.1080/13576500143000140

Gallus, J., & Mathiowetz, V. (2003). Test–retest reliability of the Purdue Pegboard for persons with multiple sclerosis. *American Journal of Occupational Therapy*, 57(1), 108–111. https://doi:10.5014/ajot.57.1.108

Hashim, A., & Low, E. L. (2014). Introduction: English in Southeast Asia. *World Englishes*, 33(4), 423–425. https://doi.org/10.1111/weng.12104

Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence-based Nursing*, 18(3), 66–67. https://doi.org/10.1136/eb-2015-102129, Retrieved from https://ebn.bmj.com/content/ebnurs/18/3/66.full.pdf

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Lee, P., Liu, C.-H., Fan, C.-W., Lu, C.-P., Lu, W.-S., & Hsieh, C.-L. (2013). The test–retest reliability and the minimal detectable change of the Purdue Pegboard Test in schizophrenia. *Journal of the Formosan Medical Association*, 112(6), 332–337. https://doi:10.1016/j.jfma.2012.02.023

Lim, Y. X., & Chai, S. C. (2020). Standardized translated instruction versus spontaneously translated instruction: Test-retest and inter-rater reliability of a hand function test. *Journal of Hand Therapy*, 33(4), 553–561. https://doi.org/10.1016/j.jht.2019.09.001

Lindstrom-Hazel, D. K., & VanderVlies Veenstra, N. (2015). Examining the Purdue pegboard test for occupational therapy practice. *The Open Journal of Occupational Therapy*, 3(3). https://doi.org/10.15453/2168-6408.1178

Li-Tsang, C. W. P., Chan, S. C. C., Chan, S. Y. Y., & Soo, A. K. W. (2004). The Hong Kong Chinese version of the jebsen hand function test: Inter-rater and test-retest reliabilities. *Hong Kong Journal of Occupational Therapy*, 14(1), 12–20. Available at: https://doi.org/10.1016/S1569-1861(09)70024-5

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382–386. https://doi.org/10.1097/00006199-198611000-00017

Mathiowetz, V., Rogers, S. L., Dowe-Keval, M., Donahoe, L., & Rennells, C. (1986). The Purdue pegboard: Norms for 14-to 19-year-olds. *American Journal of Occupational Therapy*, 40(3), 174–179. https://doi:10.5014/ajot.40.3.174

Muller, M. D., Ryan, E. J., Kim, C.-H., Muller, S. M., & Glickman, E. L. (2011). Test–retest reliability of Purdue Pegboard performance in thermoneutral and cold ambient conditions. *Ergonomics*, 54(11), 1081–1087. https://doi:10.1080/00140139.2011.620178

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing and Health*, 29(5), 489–497. https://doi.org/10.1002/nur.20147

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing and Health*, 30(4), 459–467. https://doi.org/10.1002/nur.20199

Portney, L. G., & Watkins, M. P. (2000). *Foundations of clinical research: Applications to practice* (2nd ed.). Prentice Hall Health.

Proud, E. L., Bilney, B., Miller, K. J., Morris, M. E., & McGinley, J. L. (2019). Measuring hand dexterity in people with Parkinson's disease: Reliability of pegboard tests. *The American Journal of Occupational Therapy*, *73*(4), 7304205050. https://doi.org/10.5014/ajot.2019.031112

Proud, E. L., Miller, K. J., Bilney, B., Morris, M. E., & McGinley, J. L. (2020). Construct validity of the 9-Hole Peg test and Purdue Pegboard Test in people with mild to moderately severe Parkinson's disease. *Physiotherapy*, *107*(4), 202–208. https://doi:10.1016/j.physio.2019.12.002

Razak, R. A., Madison, C. L., Siow, Y. K., & Aziz, M. A. A. (2010). Preliminary content validity and reliability of a newly developed malay preschool language assessment tool. *Asia Pacific Journal of Speech Language and Hearing*, *13*(4), 217–234. https://doi.org/10.1179/136132810805334985

Reddon, J. R., Gill, D. M., Gauk, S. E., & Maerz, M. D. (1988). Purdue pegboard: Test-retest estimates. *Perceptual and Motor Skills*, *66*(2), 503–506. https://doi:10.2466/pms.1988.66.2.503

Shahar, R. B., Kizony, R., & Nota, A. (1998). Validity of the Purdue Pegboard Test in assessing patients after traumatic hand injury. *Work*, *11*(3), 315–320. https://doi.org/10.3233/WOR-1998-11308

Shi, J., Mo, X., & Sun, Z. (2012). Content validity index in scale development. *Zhong Nan Da Xue Xue Bao. Yi Xue Ban = Journal of Central South University. Medical Sciences*. *37*(2), 152–155. https://doi.org/10.3969/j.issn.1672-7347.2012.02.007 (Article in Chinese).

Smidt, N., van der Windt, D. A., Assendelft, W. J., Mourits, A. J., Devillé, W. L., de Winter, A. F., & Bouter, L. M. (2002). Interobserver reproducibility of the assessment of severity of complaints, grip strength, and pressure pain threshold in patients with lateral epicondylitis. *Archives of Physical Medicine and Rehabilitation*, *83*(8), 1145–1150. https://doi.org/10.1053/apmr.2002.33728

Sousa, V. D., & Rojjanasrirat, W. (2011). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user friendly guideline. *Journal of Evaluation in Clinical Practice*, *17*(2), 268–274. https://doi.org/10.1111/j.1365-2753.2010.01434.x

Tiffin, J. (1987). *Purdue pegboard, examiner's manual*. Science Research Associates: Inc.

Tiffin, J., & Asher, E. J. (1948). The Purdue Pegboard: Norms and studies of reliability and validity. *The Journal of Applied Psychology*, *32*(3), 234–247. https://doi.org/10.1037/h0061266

Wears, R. L. (2015). Standardisation and its discontents. *Cognition, Technology and Work*, *17*(1), 89–94. https://doi.org/10.1007/s10111-014-0299-6

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research*, *19*(1), 231–240. https://doi:10.1519/15184.1

Wilson, B. C., Iacoviello, J. M., Wilson, J. J., & Risucci, D. (1982). Purdue Pegboard performance of normal preschool children. *Journal of Clinical Neuropsychology*, *4*(1), 19–26. https://doi.org/10.1080/01688638208401113

World Health Organization (WHO) (2017). *Process of translation and adaptation of instruments*. Available from: http://www.who.int/substance_abuse/research_tools/translation/en/

Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., & Nikanfar, A.-R. (2015). Design and implementation content validity study: Development of an instrument for measuring patient-centered communication. *Journal of Caring Sciences*, *4*(2), 165–178. https://doi.org/10.15171/jcs.2015.017