Cross-cultural adaptation and psychometric validation of the Indonesian version of the Oxford Hip Score

J. Silitonga, Y. P. Djaja, I. H. Dilogo, L. A. P. Pontoh
From Fatmawati General Hospital, Jakarta, Indonesia

Aims
The aim of this study was to perform a cross-cultural adaptation of Oxford Hip Score (OHS) to Indonesian, and to evaluate its psychometric properties.

Methods
We performed a cross-cultural adaptation of Oxford Hip Score into Indonesian language (OHS-ID) and determined its internal consistency, test-retest reliability, measurement error, floor-ceiling effect, responsiveness, and construct validity by hypotheses testing of its correlation with Harris Hip Score (HHS), visual analogue scale (VAS), and Short Form-36 (SF-36). Adults (>17 years old) with chronic hip pain (osteoarthritis or osteonecrosis) were included.

Results
A total of 125 patients were included, including 50 total hip arthroplasty (THA) patients with six months follow-up. The OHS questionnaire was translated into Indonesian and showed good internal consistency (Cronbach’s alpha = 0.89) and good reliability (intraclass correlation = 0.98). The standard error of measurement value of 2.11 resulted in minimal detectable change score of 5.8. Ten out of ten (100%) a priori hypotheses were met, confirming the construct validity. A strong correlation was found with two subscales of SF-36 (pain and physical function), HHS (0.94), and VAS (-0.83). OHS-ID also showed good responsiveness for post-THA series. Floor and ceiling effect was not found.

Conclusion
The Indonesian version of OHS showed similar reliability and validity with the original OHS. This questionnaire will be suitable to assess chronic hip pain in Indonesian-speaking patients.

Cite this article: Bone Jt Open 2021;2-9:765–772.

Keywords: Hip pain, Oxford Hip Score, Cross-cultural adaptation, Indonesian, Hip osteoarthritis

Introduction
Hip osteoarthritis (OA) is one of a major cause of disability worldwide especially for the elderly. One in four people who live to age 85 may develop hip OA in their lifetime, and 10% lifetime risk of having total hip arthroplasty (THA) surgery.1,2 The traditional approach to measure the disability caused by hip OA has been by assessing the clinical signs and symptoms. However, in recent years, there has been an increasing trend on the use of patient-reported outcome measures (PROMs) to provide a better description of the patient’s perspective.3

The Oxford Hip Score (OHS)4,5 is a 12-item, hip-specific, self-administered questionnaire. It has been widely used as one of the primary outcome measures for hip arthritis and THA due to its reliability, validity, and responsiveness.4-6 However, due to its self-reported nature, its validity is questionable when applied in non-English-speaking countries.7-10 Therefore, a cross-cultural adaptation is required before using it in different languages or cultures.7

Indonesian language is one of the most spoken languages in the world as Indonesia’s current population exceeds 250 million.11 However, no Indonesian language version
of OHS is available or validated. Our study aimed to perform a cross-cultural adaptation of OHS in Indonesian language and to assess its psychometric properties.

**Methods**

The original UK English version of the OHS was used in this study. Approval for the adaptation was obtained from Oxford University Innovation. The Institutional Review Board approved the trial prior to the study (including ethical approval). Adaptation was performed based on Beaton's cross-cultural guidelines.7

**Phase 1: forward translation.** Two native Indonesian speakers independently translated the original questionnaire to Indonesian (YPD, LU). The first translator was a certified English-Indonesian translator who was not familiar with the concept of OHS questionnaire. The second translator (YPD) was a clinician who was familiar with OHS. Both forward translations were compared. The different/ambiguous terms were documented and resolved after a discussion between the two translators resulting in combined forward translations.

**Phase 2: back translation.** Two native English speakers with a medical background, fluent in Indonesian and blinded to the English version of OHS, separately translated back the combined forward translation into English (ND, AD). Both back translations then were compared with the original version of the questionnaire to validate if the translated version reflected the same content as the original version.

**Phase 3: expert committee.** The back translation was reviewed by the four translators, and principal investigators team. This review process aimed to highlight any discrepancies in meaning or terminology used and to obtain the best possible translation, which was the pre-final version. Each issue, rationale, and decision during the discussions was documented.

**Phase 4: pilot test of the pre-final version.** The comprehensiveness of the pre-final questionnaire was tested in 30 hip patients to ensure the adapted version was understandable. After completing the questionnaire, the subjects were interviewed to explore their understanding of each question and response. This result of this test was then re-evaluated by the committee and the final form of a questionnaire (OHS-ID) was then established (Supplementary Material).

**Phase 5: test of the final version.** The questionnaire (OHS-ID) was field-tested in 125 patients to assure the validity and other psychometric properties remained intact. Consecutive sampling was conducted in the outpatient clinic in a tertiary referral hospital from January 2019 to December 2020. Sample size (125 patients) was predetermined based on the subject to item ratio ≥ 10.12 The inclusion criteria were: chronic hip pain ≥ three months; diagnosed as hip OA or osteonecrosis; scheduled for hip arthroplasty surgery; adult (older than 17 years); and able to read and write in Indonesian fluently. The exclusion criteria were acute hip pain (trauma, such as fracture or dislocation); presence of neurological deficit; and had an incidental event (including surgery) during the observation period that might increase/reduce pain significantly. All patients were informed of the nature of the study and informed consent was obtained before these procedures.

The study initially included 150 respondents; 14 dropped out due to inability to return for the second test; meanwhile, 11 underwent changes in their condition (from the transition question), leaving 125 final respondents who corresponded with the predetermined minimum sample. Their mean age was 52.9 years (standard deviation (SD) 14.3) and 44 respondents were male (35.2%).

Each respondent filled two booklets (one consisted of the OHS-ID, the other consisted of the VAS and Indonesian Short-Form 36 questionnaire13) twice within a one-week interval. The respondents were asked to fill the questionnaire by themselves or with the help of the primary caregivers. If they could not understand some questions, they were allowed to skip them. The response rate and time required to complete the first booklet was recorded. Interview and physical examination to evaluate Harris Hip Score (HHS)14 were performed at the first meeting. On the second meeting, the respondents were also given a transition question to determine whether their conditions were stable during the one-week period. Follow-up evaluation was performed after six months postoperatively in 50 patients who underwent THA surgery during the study period.

The scoring of OHS-ID was performed based on the 2007 OHS update (total score range: 0 to 48).3 The assessment of psychometric properties (internal consistency, test-retest reliability, measurement errors, responsiveness, and construct validity) was performed and presented based on COndensus-based Standards for the selection of health status Measurement INstruments guidelines.15

**Statistical analysis.** Analysis was performed using the IBM SPSS software v. 22.0 (SPSS, USA). Internal consistency of OHS-ID was evaluated by calculating Cronbach's alpha (CA) of the baseline questionnaires.16 It indicates the correlation between all items within the test instrument and the correlation between each item and the whole test instrument. CA values were considered high if they ranged from 0.70 to 0.90. If CA is too high (> 0.90), it may suggest that some items are redundant as they are testing the same question.17 Previous studies reported CA of OHS ranges from 0.84 to 0.93.18 Correlation between each item and the whole instrument (OHS-ID) were calculated using Spearman’s rank correlation.

Test-retest reliability was determined using intraindividual correlation coefficient (ICC) and standard error of measurement (SEM) on each component of the questionnaire between the first and second tests. The ICC used was a single measurement, absolute agreement,
two-way mixed-effects model with 95% confidence intervals. An ICC of > 0.70 was regarded as good reliability.19

**Measurement error.** SEM was determined from the error variance of the analysis of variance (ANOVA) associated with determination of the ICC.20 Minimum detectable change (MDC95%) was calculated by multiplying SEM by 2.77, where 2.77 was obtained from Z-value for the 95% confidence interval (CI) (1.96) × the variance of two measurements (\(\sqrt{2}\)).20,21 The distribution of floor-and-ceiling effect (percentage of sample achieving the worst and best possible scores, respectively) was also determined for both the baseline test and the follow-up test. Test instruments should exhibit minimal floor-and-ceiling effect (less than 15% of the respondents) to be considered reliable.22

**Construct validity: hypothesis testing.** A total of ten hypotheses (Table I) were tested to evaluate the construct validity of OHS-ID, using the standard hypothesis testing methodology.23 Spearman’s rank correlation coefficient was used to assess the association of baseline OHS-ID with HHS, VAS for pain, and the eight subscales of SF-36 questionnaire; VAS, visual analogue scale.

### Table I. Hypotheses testing for construct validity of the Indonesian Oxford Hip Score.

| Scale                        | Hypothesis                   |
|------------------------------|------------------------------|
| OHS-ID score vs VAS          | Negative strong association  |
| OHS-ID score vs Harris Hip Score | Positive strong association |
| OHS-ID score vs SF-36        | Positive strong association  |
| SF-36 physical function      | Positive moderate to strong association |
| SF-36 role limitation physical | Positive moderate to strong association |
| SF-36 role limitation emotional | Positive moderate to strong association |
| SF-36 mental health          | Positive moderate to strong association |
| SF-36 bodily pain            | Positive moderate to strong association |
| SF-36 general health perceptions | Positive moderate to strong association |
| SF-36 vitality (energy/fatigue) | Positive moderate to strong association |
| SF-36 social functioning     | Positive moderate to strong association |

OHS-ID, Indonesian Oxford Hip Score; SF-36, Short-Form 36 questionnaire; VAS, visual analogue scale.

### Table II. Characteristics of baseline Indonesian Oxford Hip Score and its internal consistency.

| Item | Mean scale when the item is removed | Corrected item-total correlation* | Cronbach’s α when the item is removed |
|------|--------------------------------------|-----------------------------------|---------------------------------------|
| OHS 1 | 14.57                                | 0.671                             | 0.88                                  |
| OHS 2 | 13.81                                | 0.542                             | 0.88                                  |
| OHS 3 | 13.84                                | 0.712                             | 0.87                                  |
| OHS 4 | 13.91                                | 0.560                             | 0.88                                  |
| OHS 5 | 14.00                                | 0.692                             | 0.87                                  |
| OHS 6 | 13.63                                | 0.600                             | 0.88                                  |
| OHS 7 | 13.93                                | 0.742                             | 0.87                                  |
| OHS 8 | 13.77                                | 0.536                             | 0.88                                  |
| OHS 9 | 14.21                                | 0.614                             | 0.89                                  |
| OHS 10| 13.62                                | 0.576                             | 0.88                                  |
| OHS 11| 13.96                                | 0.540                             | 0.88                                  |
| OHS 12| 13.86                                | 0.603                             | 0.88                                  |

*Significant correlation (95% Critical value of the Pearson correlation coefficient for 12 items = 0.532)

### Results

During the cross-cultural adaptation process, several noteworthy issues arose during the translation phase and were solved in an expert meeting.

Translation of “most nights”: the direct translation of the phrase is “sebagian besar malam” which was an uncommon phrase in the Indonesian language. Therefore, we changed the term to “hampir setiap malam”, which means “almost every night” in English.

Translation of “extreme difficulty”: the translation of this phrase to Indonesian is “kesulitan ekstrim”, which is not a common phrase used in daily living/speaking in Indonesia. During the pilot test for the pre-final version, most of the respondents complained about this phrase. Therefore, we decided to change the phrase to “sangat kesulitan”, which means “very difficult” in English.

Back translation of “all over”: the phrase “all over” that followed “washing and drying yourself” was self-explanatory. It synonymized with “completely” or “everywhere”. Direct translation of the term to Indonesian is “seluruhnya”. However, the back translation always resulted in “completely”. Due to the same meaning of both words, we decided to keep this translation.

Back translation of “put on”: similar to the above, the back translation of the phrase “put on” or “memakai” in Indonesian always resulted in its synonym “wear”. The same decision was made since the definition is similar.

Back translation of “household shopping”: the phrase was translated as “belanja kebutuhan rumah tangga” in Indonesian, but in the back translation, the result was “shop household needs”. The reason for this back translation was the grammatical difference. Indonesian language has simpler grammar, and therefore the back translation of a longer phrase may result in simple English.

### Patient demographics and scores distribution.

The response rate was 100%. The mean time required to complete OHS-ID was 4.5 minutes (SD 1.6; 2.5 to 8.0) and the
The mean interval between first and second questionnaire was 6.9 (SD 0.7).

The mean OHS-ID score on the first and second administrations was 15.2 (SD 6.99) and 15.4 (SD 7.01), respectively. The mean VAS was 5.9 (SD 1.5), indicating moderate pain. The mean score of HHS was 32.1 (SD 5.2) and the mean total SF-36 score was 46.3 (SD 10.9). These data were not normally distributed, therefore Spearman’s test was used for further analysis.

**Internal consistency.** Cronbach’s α index for OHS-ID was 0.89, which suggested good internal consistency. A significant correlation between item and total scores was found in every item. (Table II)

**Test-retest reliability and measurement error.** Table III describes the test-retest reliability between the first and second tests of OHS-ID expressed by ICC with its 95% CI. Intraclass correlation for OHS-ID was 0.98 (95% CI 0.97 to 0.99), showing good reliability. SEM was 2.11. The MDC was 5.8. Floor-and-ceiling effects were not found.

**Construct validity.** OHS-ID has statistically significant correlations with HHS, VAS, and all SF-36 subscales (Table IV). We found all of the a priori hypotheses were met, confirming the construct validity of OHS-ID. It correlated strongly with VAS, HHS, SF-36 physical function, and SF-36 bodily pain, and moderately with the other SF-36 subscales.

**Responsiveness.** The mean OHS-ID improved to 42.6 (SD 3.17) after THA surgery (p < 0.001, paired t-test). The ES and SRM values were 4.56 and 3.23, respectively.

**Discussion**

At present, there are no hip-specific instruments that have been translated and cross-culturally adapted into Indonesian. The current main instrument used for evaluating THA patients in Indonesia is the English HHS, which is a clinician-based outcome measure. However, the outcome of a clinical intervention obtained by PROMs are regarded to be more important and reliable than any other outcome measure, including those which are clinician-based.25

The adaptation process was carried out in accordance with the established guideline for cross-cultural adaptations to obtain a reliable and valid adaptation of the questionnaire.7 The translation was carried out without major difficulties. Several minor issues arose regarding the translation of the aforementioned phrases, all of which were resolved during the expert committee meeting.

---

**Table III.** Test-retest scores of the Indonesian Oxford Hip Score.

| Variable | First test, mean (SD) | Second test, mean (SD) | ICC (95% CI) | p-value * |
|----------|-----------------------|------------------------|--------------|-----------|
| Total    | 15.19 (6.99)          | 15.37 (7.02)           | 0.983 (0.976 to 0.988) | 0.108     |
| OHS 1    | 0.62 (0.59)           | 0.68 (0.62)            | 0.924 (0.890 to 0.947) | 0.008     |
| OHS 2    | 1.38 (0.97)           | 1.38 (0.86)            | 0.929 (0.900 to 0.949) | 0.796     |
| OHS 3    | 1.37 (0.80)           | 1.35 (0.83)            | 0.961 (0.944 to 0.972) | 0.480     |
| OHS 4    | 1.28 (0.92)           | 1.31 (0.88)            | 0.892 (0.850 to 0.923) | 0.394     |
| OHS 5    | 1.19 (0.97)           | 1.22 (0.89)            | 0.903 (0.865 to 0.931) | 0.513     |
| OHS 6    | 1.56 (0.87)           | 1.56 (0.85)            | 0.946 (0.924 to 0.962) | 0.997     |
| OHS 7    | 1.26 (0.87)           | 1.26 (0.83)            | 0.950 (0.930 to 0.965) | 0.739     |
| OHS 8    | 1.42 (0.82)           | 1.40 (0.80)            | 0.872 (0.822 to 0.908) | 0.513     |
| OHS 9    | 0.98 (0.92)           | 1.00 (0.89)            | 0.937 (0.912 to 0.955) | 0.405     |
| OHS 10   | 1.57 (0.81)           | 1.60 (0.79)            | 0.887 (0.844 to 0.920) | 0.346     |
| OHS 11   | 1.23 (0.77)           | 1.29 (0.79)            | 0.902 (0.863 to 0.930) | 0.071     |
| OHS 12   | 1.33 (0.96)           | 1.31 (0.89)            | 0.907 (0.869 to 0.933) | 0.655     |

*p*Analysis of variance with Friedman’s test.

ICC, intraclass correlation coefficient; OHS, Oxford Hip Score; SD, standard deviation.

**Table IV.** Construct validity: correlation coefficient between the Indonesian Oxford Hip Score, visual analogue scale, and Harris Hip Score.

| Comparison | Mean (SD) | p     | p-value * | Hypothesis confirmed |
|------------|----------|-------|-----------|----------------------|
| OHS-ID score vs VAS | 5.9 (1.5) | -0.831 | < 0.001   | Yes                  |
| OHS-ID score vs HHS  | 32.1 (15.2) | 0.945 | < 0.001   | Yes                  |
| SF-36 physical functioning | 41.7 (13.4) | 0.782 | < 0.001   | Yes                  |
| SF-36 role limitation physical | 33.2 (21.0) | 0.461 | < 0.001   | Yes                  |
| SF-36 role limitation emotional | 57.3 (16.7) | 0.452 | < 0.001   | Yes                  |
| SF-36 mental health | 50.7 (10.9) | 0.585 | < 0.001   | Yes                  |
| SF-36 bodily pain | 40.0 (11.9) | 0.858 | < 0.001   | Yes                  |
| SF-36 general health perceptions | 49.3 (13.7) | 0.675 | < 0.001   | Yes                  |
| SF-36 vitality (energy/fatigue) | 48.8 (11.9) | 0.551 | < 0.001   | Yes                  |
| SF-36 social functioning | 49.5 (10.1) | 0.482 | < 0.001   | Yes                  |

*Spearman’s test.

*p*, correlation coefficient; HHS, Harris Hip Score; OHS, Oxford Hip Score; SF-36, Short-Form 36 questionnaire; VAS, visual analogue scale.
Due to the fact that it was administered in a clinical setting, the response rate of OHS-ID was higher than previous validation studies. Having no missing data and short time required to complete the questionnaire showed that OHS-ID was easily administered and understood by the patients. The average time to complete the questionnaire was 4.5 minutes, which was within the expected interval (2 to 15 minutes).

The Cronbach’s alpha correlation coefficient for OHS-ID (0.89) showed a good internal consistency that is comparable with the other translated versions. The internal consistencies from previous studies were ranged from 0.84 to 0.99 (Supplementary Table I). The Pearson coefficient for item total (ranging from 0.41 to 0.74) also indicated a moderate-strong correlation between each item and total score.

Test-retest reliability shows the consistency of a questionnaire within a time interval. The seven-day interval was usually chosen since it was short enough to avoid changes due to disease progression/resolution, but not too short to allow recalling of previous answers. An interval of seven to 14 days was the most frequently used interval in previous cross-cultural adaptation studies. We also added a transition question/interview to ensure the patient’s condition did not change during this period. The ICC for OHS-ID (0.98) was considered of good reproducibility, which was also in accordance with the previous validation studies.

The SEM and MDC are important features of a questionnaire. The MDC (which is derived from the SEM) has been used to detect whether the change is clinically relevant and is not caused by the measurement error. The MDC of OHS-ID was 5.8 points, which is similar to the original OHS (MDC 4.85 points) and the Spanish OHS (MDC 5.5 points). However, in order to determine whether this change is clinically important or not, different metrics are used. Beard et al. described anchor-based methods to explore change or difference in score on the OHS after a hip surgery. Instead of using minimal clinically important difference (MCID), they used minimal important change (MIC) to detect the smallest change of OHS for a single individual/group over time and minimal important difference (MID) to detect the difference between two independent groups of patients. For OHS, the MIC values were 11 for a single group and 7.5 at individual level. Furthermore, the ratio between the MIC and MDC in OHS-ID was higher than 1, indicating that the MIC can be discriminated clearly from measurement error.

In order to assess construct validity, we compared OHS-ID to HHS, VAS, and all the subscales of SF-36 using the hypotheses testing method. The hypotheses were developed based on the direction and magnitude of the correlations obtained from previous cross-culturally adapted versions. Although the hypotheses on the association of OHS-ID with all subscales of SF-36, HHS, and VAS were met, the strength of correlations was varied. OHS-ID showed similar VAS correlation when compared with the other adapted versions \((r = -0.79 \text{ to } -0.53; \text{ as shown in Table IV})\). Moderate correlation was found with HHS, which is similar with German \((r = 0.63)\) and French \((r = 0.6)\) version.

Since OHS generally evaluates the patient’s pain and physical function aspect, strong correlation with physical function and bodily pain subscales in SF-36 was expected. The ID-OHS exhibited a strong correlation with physical function and bodily pain subscales, which was also consistent with the other versions. Interestingly, no correlation was found with the general health (GH) subscales, which was also observed in Japanese and Korean versions. However, in the Japanese validation study, significant correlation with GH subscales was found later after THA surgery. The authors suggested that THA subjects responded in a manner similar to the general population after experiencing significant reductions in pain.

The responsiveness is an important characteristic of a questionnaire since it reflects the sensitivity to clinical changes. The result of responsiveness assessment revealed that ID-OHS was able to detect changes after the surgical treatment with comparable responsiveness with the previous OHS validation studies.

Our study was subjected to several limitations. First, the sample used might not reflect entire Indonesian population. Given there are over 300 different native languages in Indonesia, it is possible that some populations still used their native language in daily living. However, since Indonesia’s literacy rate is 95%, we are confident that at least those portions of Indonesian are fluent in the Indonesian language. The second limitation of this study was the patient selection. We only included patients with hip OA or end-stage osteonecrosis scheduled for THA, which may account for the low OHS-ID score in our study. Finally, since there was no other hip-specific questionnaire available in Indonesian, we were only able to measure the construct validity using the generic HHS (physician-based questionnaire), VAS, and SF-36 (which has already been validated previously).

In conclusion, the Indonesian version of Oxford Hip Score maintains the reliability, validity, and the psychometric characteristics of the original English version. This questionnaire will be a suitable instrument in assessing hip OA-related disability and THA outcome for Indonesian-speaking patients.

**Take home message**
- The Indonesian version of Oxford Hip Score maintains the reliability, validity, and psychometric characteristics of the original English version.
- This questionnaire will be a suitable instrument in assessing disability related to the hip osteoarthritis and the outcome of total hip arthroplasty for Indonesian-speaking patients.

**Supplementary material**
Characteristics of the published cross-cultural adaptation studies of the Oxford Hip Score.

**References**
1. Culliford DJ, Maskell J, Kiran A, et al. The lifetime risk of total hip and knee arthroplasty: Results from the UK general practice research database. *Osteoarthritis Cartilage* 2012;20(8):S19–S24.
