Chung Shan Medical Journal 2024; 35:93-100

airiti

Original Article

The Direct Observation of Procedural Skills (DOPS) Validity and Reliability Set for Occupational Therapy of Splint Fabrication

Ming-I Chen [1, 2*]

[1] *Occupational Therapy Room, Chung Shan Medical University Hospital, Taichung City, Taiwan*
[2] *Department of Occupational Therapy, Chung Shan Medical University, Taichung City, Taiwan*

Purpose: The objectives of this study were to investigate the content validity, test-retest and intra-rater reliability characteristics of the Directly Observed Procedural Skills (DOPS), measuring Occupational Therapy Students in performing the fabricating skill of splint and to help inform decisions about the outcome of students' and postgraduate staffs' training.
Methods: We developed Occupational Therapist DOPS for Taiwan, established content items and verify the validity, reliability by conducted 14 Registered Occupational Therapists (ROT) from Chung Sun Medical University Hospital in Taichung City Taiwan.
Results: DOPS for Taiwan has excellent content validity, total CVI =0.94. Individual item-CVI, had 4 items to give total agreements. For reliability analysis, we conducted the Item-Deleted Cronbach's Alpha from 0.61 to 0.98 and the correlation coefficient from 0.55 to 0.90. For test-retest reliability showed good stability over time, Pearson's correlation coefficient α was 0.812 for the two separate time raters, and intra-rater reliability, there was satisfactory with good internal consistency for subscales, 'Cronbach's Alpha = 0.906, 95% C.I. = 0.800-0.968 ($p<0.01$).
Conclusion: This is the first attempt to evaluate the use of comprehensive workplace assessment across occupational therapists in Taiwan. The methods are feasible to conduct and can make reliable distinctions between OTs performances. With our adaptation, they may be appropriate for assessing the performance of other grades and specialties of ROT. This may be helpful in informing foundation assessment.

Keywords: Direct Observation of Procedural Skills (DOPS), Workplace-Based Assessments (WBA), Validity, Test-retest, Ankle-Foot-Orthosis (AFO)

1. Introduction

The outbreak of severe acute respiratory syndrome (SARS) epidemic that occurred during 2003 exposed serious deficiencies especially in Medical Education Program in Taiwan. For the reason, the Department of Health in Taiwan had in efforts to engage the reform of promoting "Post Graduate Year (PGY)" programs which strengthen medical practitioners' education and improve the quality of patients' care. The establishment of valid and reliable clinical performance assessment tools was the first step of reform in Medical Education (1). A paper-and-pencil knowledge test would not recognize the realistic technical knowledge's aspects of clinical skills. Traditionally, occupational therapy educator used the direct observation and scored with checklists to evaluate the performance of OTs and PGY trainees in clinical practices (2-5). Those

* Corresponding Author: Ming-I Chen
Address: No. 110, Sec. 1, Jianguo N. Rd., Taichung City, 40201, Taiwan
Tel: +886-4-2473-9595 ext. 21520
E-mail: cmi1239@gmail.com

airiti

checklists had been subjective that could be under or over estimating the students' or trainees' performance for all of the evaluators and were not subtly settled by scientific investigation. There is not an objective criterion of the assessment process in occupational therapy clinical practice facility in Taiwan. We would have not universally standardized tests (i.e. valid and reliable) for occupational therapy students' and postgraduates' performance in clinic skill especially in a fabricating splint which was considered critical skill in occupational therapy core practical education. However, there is a usual agreement of the necessity to develop and apply reliable measurements of students' performance. But we had commented on the lack of studies in occupational therapy for assessing the validity and reliability of workplace-based assessment, even though those methods are fairly widely used to assess the competency of others' medical professional residents (6).

Workplace-Based Assessments (WBA) provided structured and criterion based tools to assess the clinical tasks and practices. The tools should be used in varying clinic contexts to evaluate not only practical and students' cognitive abilities but also attitudes and professionalism (7-10). For the reason, we deeply concerned that we must settle a tool which were valid and reliable to be objectively assessing the performance of occupational therapist students and postgraduate staffs in clinical faculties. The Direct Observation of Procedural Skills (DOPS) was developed by The Royal College of Physicians (RCP) as a tool for physicians' performance in clinical skills. It was well-established and widely use in all medical practitioners' WBA (11). The DOPS has been confirmed to be valid, reliable and feasible in assessing medical practitioners (12-13). However, there is no published literature about DOPS validity and reliability studies for OT facility context in usefulness and appropriateness.

The objectives of this study were to investigate the content validity, test-retest and intra-rater reliability characteristics of the Directly Observed Procedural Skills (DOPS), a 10-item measure of Occupational Therapy Students in performing the fabricating skill of splint and to help inform decisions about the outcome of students' and postgraduate staffs' training.

2. Materials and Methods

2.1. Development of the DOPS for Taiwan

We translated RCP version of DOPS from English into Chinese for conventional understanding as the official language in Taiwan is Chinese. The development of the Occupational Therapist DOPS for Taiwan and the establishment of its content items was achieved from two senior occupational therapists. The DOPS for Taiwan included 10 items. There were four main domains, Clinical assessment, Patients' consent, Communication; Safety issues; Splinting skills during fabrication; and Diagnostic and therapeutic ability. Each of four main domains contained several individual items (see Table 1).

Table 1. The Items of DOPS for Taiwan Domains for Occupational Therapy.

1. Demonstrates comprehension of the indications, anatomy.
2. Performs an adequate preparation of instruments/material to be used before the procedure of splint fabrication.
3. Awareness about patient's condition during the procedure.
4. Diagnosis and management of complications.
5. The technique of the procedure of splint fabrication.
6. Seeks helps where appropriate.
7. Post procedure management.
8. Communication skills.
9. Professionalism during the procedure execution, consideration of the patients' welfare and safety.
10. Overall ability to perform the procedure.

airiti

2.2. Content Validity Assessment

We conducted a feasibility study and generalizability analysis based on the application of these assessment methods and the resulting data from modified DOPS. A total of 14 Registered Occupational Therapists (ROT) from Chung Sun Medical University Hospital in Taichung City Taiwan took part. We performed the Content Validity Index with 5-scale assessment for each item of validity measurement before confirming each item of assessment would be recognized by raters. Lynn considered 3 or 5 points rating scales might be accepted and appreciated (14). We appeared to be in use with 1= highly not relevant, 2 = somewhat not relevant, 3 = no command, 4 = quite relevant, 5 = highly relevant. To prevent a neutral and ambivalent midpoint rating, investigators randomized arrange of the 5-point rating scales and introduced in front of the items. We use the Content Validity Index (CVI) for each assessment of 10 items of DOPS; they would be counted by item-CVI, scale-CVI, and universal agreement (UA). The item-CVI represented each item of DOPS Content Validity scored by all raters and mean from the dichotomy of 5-point rating (i.e. 1 and 2 represented score 0; 3-5 represented score 1). Scale content validity index (S-CVI) represented the mean from total scores of all items scores. Universal Agreement (UA) represented the rate from all items were all agreement (i.e. all items were 3, 4, 5 scores). The assessors scored the items independently using the modified DOPS forms; the completed DOPS and evaluation forms were proofread. We set free-text comments for assessors to entered other opinions about their concerns onto a spreadsheet. The comments were reviewed by 2 investigators.

2.3. Test-Retest Reliability and Intra-Rater Reliability Assessment

High reliability means that the assessment could reach the consistency to assess trainees in the same circumstances. It would be considered the good assessment tool as if the reliability coefficient was high enough (15). For test-retest reliability and intra-rater reliability measurement, all raters would watch the processing of two Occupational Therapy students in CSMU Hospital Occupational Therapy Department to fabricate the Ankle-Foot-Orthosis splint. For test-retest reliability, we conducted the raters' scores for stability and equivalence were set when the same procedure was acted by the same students after a period of time interval. We conducted the two separate time between two weeks to prevent the acquire effect and randomized items-arranging to prevent arbitrary rating. The rater might be the same, and the procedure can be an equivalent one. Our study passed CSMU IRB approval, number #.CS17117. We want to know those 14 raters in one assessment process. We conducted the Pearson correlation coefficient for test-retest reliability, in generally; the coefficient over than 0.8 was considered high correlation (16). The intra-rater reliability coefficient was introduced by Cronbach alpha which the value is represented among the trainers' internal consistency. We appeared an appropriate cutoff point for assessments is 0.8 (17). Variance component analysis of means scores and derivation of 95%e confidence interval for individual therapists scores based on the standard deviation of measurement for intra-rater reliability (18).

3. Results

3.1. Validity Analysis

Table 2 showed the Content Validity Index (CVI for short) for rating on DOPS for Taiwan by 14 Experts. The DOPS for Taiwan has excellent content validity, based on the advice of 14 ROT, total CVI =0.94. For the individual item-CVI, we found the item #1, item #2, item #7, and item #8 gave total agreement. For the total agreement (i.e. the scale for the individual item was 3, 4, and 5) we used the universal agreement (UA) of raters in all agreement for all items. In Table 2, we demonstrated the ratio of Scale-level-CVI/UA was 0.4. The other index for CVI was the Proportion relevant for individual expert among 0.7 to 1.0; the expert # 12 gave the lowest scale.

3.2. Reliability Analysis

Reliability Analysis for Item-Total Statistics

Table 3 showed the Item-Total Statistics for Case 1, we concluded the good reliability analysis Item-Total Cronbach's Alpha = 0.928. If we conducted the Item-Deleted Cronbach's Alpha from 0.61 to 0.98 and the correlation coefficient from 0.55 to 0.90. It represents this DOPS for Taiwan evaluation form was excellent

airiti

Table 2. CVI Validity Analysis of DOPS for Taiwan by 14 Experts.

	CVI Rating															
ITEM	Ex 1	Ex 2	Ex 3	Ex 4	Ex 5	Ex 6	Ex 7	Ex 8	Ex 9	Ex 10	Ex 11	Ex 12	Ex 13	Ex 14	Number in Agreement	Item CVI
1	5	4	4	5	4	5	4	5	4	5	5	4	4	5	14	1.00
2	4	4	4	5	4	5	5	5	5	5	5	4	4	5	14	1.00
3	4	5	4	4	3	4	2	4	4	5	4	2	4	4	12	0.86
4	1	4	2	5	3	5	4	5	5	5	5	4	5	4	12	0.86
5	5	5	3	5	2	5	2	5	5	5	5	4	5	5	12	0.86
6	4	4	4	5	4	4	3	4	4	5	4	2	5	4	13	0.93
7	5	4	5	5	4	5	4	5	5	5	5	4	5	5	14	1.00
8	4	3	5	5	3	5	4	5	5	5	4	4	5	4	14	1.00
9	4	3	3	4	3	4	4	5	4	5	3	2	5	3	13	0.93
10	4	4	4	5	4	5	4	4	5	5	5	1	5	4	13	0.93
Proportion relevant	0.9	1	0.9	1	0.9	1	0.8	1	1	1	1	0.7	1	1		

1. Mean I-CVI= 0.94; S-CVI/UA = 0.4
2. Ex = Expert

Table 3. Reliability Analysis for Item-Total Statistics in Case 1

Item-Total Statistics in Case 1

ITEM	Scale Mean if Item Deleted	Scale Variance if tem Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted	Mean	SD
1	39.86	51.43	41.19	0.61	0.95	5.36	0.93
2	39.43	51.14	42.59	0.63	0.87	5.64	0.75
3	39.79	51.14	37.36	0.81	0.88	5.64	1.08
4	38.21	50.79	42.95	0.55	0.61	6.00	0.78
5	40.36	51.71	37.60	0.74	0.87	5.07	1.14
6	39.71	50.93	37.00	0.90	0.96	5.86	1.03
7	39.57	51.21	42.95	0.50	0.96	5.57	0.85
8	39.07	50.86	40.59	0.87	0.96	5.93	0.73
9	38.71	50.64	40.25	0.86	0.98	6.14	0.77
10	40.00	51.21	39.72	0.82	0.94	5.57	0.85

Cronbach's Alpha = 0.928

reliability tool for OT students' performance. Nevertheless, we found same trend over Item-Total Statistics for Case 2 (Table 4), but exempted the item #6 (α=0.15).

Test-Retest Reliability and Intra-Rater Reliability

airiti

Table 4. Reliability Analysis for Item-Total Statistics in Case 2

Item-Total Statistics in Case 2

ITEM	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted	Mean	SD
1	39.86	36.29	0.70	0.90	0.83	4.00	0.78
2	39.43	31.96	0.78	0.89	0.82	4.43	1.16
3	39.79	31.26	0.80	0.69	0.81	4.07	1.21
4	38.21	36.34	0.56	0.63	0.84	5.64	0.93
5	40.36	36.86	0.51	0.75	0.84	3.50	0.94
6	39.71	38.68	0.15	0.83	0.89	4.14	1.46
7	39.57	35.80	0.63	0.89	0.83	4.29	0.91
8	39.07	34.69	0.69	0.97	0.83	4.79	0.98
9	38.71	37.60	0.49	0.95	0.84	5.14	0.86
10	40.00	38.77	0.67	0.91	0.84	3.86	0.54

Cronbach's Alpha = 0.853

The test-retest reliability procedure showed good stability over time, Pearson's correlation coefficient r was 0.812 for the two separate time raters. The intra-rater reliability, there was satisfactory with good internal consistency for subscales, 'Cronbach's Alpha = 0.906, 95% C.I. = 0.800-0.968 ($p<0.01$).

4. Discussion

The DOPS for Taiwan was Excellent Internal Consensus Assessment Tool

In our investigation, we used an expert consensus to set a content valid DOPS tool for Taiwan Occupational Therapy trainees' context, which was Workplace-based-tested to analyze its psychometric properties and to evaluate its acceptability to the senior ROTs. We considered the DOPS for Taiwan was high valid evaluation tool for educators in OT trainees' clinical practices.

From Eastern to Western countries, we found the previous studies of DOPS were good to excellent validity (12, 13), and wide range of application in fields of work-place-assessment including surgeon, nursing care, endoscopy, intubation, internal medicine, radiology…etc. (7, 8, 10, 12, 13, 17,19, 20). It meant that DOPS were cross-disciplines and cross-cultures work-placed-assessment tool.

In our study, the universal agreement was set by 0.4. It represented that 4 out of ten items were full consensus for all of 14 raters. It might be a hard task for the high rate of universal agreement within multiple raters (21-23). Although, "Universal agreement" was an unusual term of validity psychological parameter, it might be an important predictor for multiple raters.

In the content validation processing, we considered that the original DOPS (i.e., the UK version for the physician) were similar with our modified DOPS for Taiwan, but with some differences. For example, the items of the original version were 11, we set for Taiwan version were 10 items. We removed the #6 item of Aseptic technique, due to the whole procedure of Taiwan DOPS in OT splinting technique did not necessarily higher-level Aseptic technique in content. We consider the shrinking term would not change the holistic conception of original DOPS (Table 3 and Table 4), and be easy to calculate, to comprehend, to understand and to explain the results.

DOPS for Taiwan was high Test-Retest Reliability and Intra-Rater Reliability

We estimate test-retest reliability when we

airiti

administer the same test to the same sample on two different occasions. This approach assumes that there is no substantial change in the construct being measured between the two occasions. The wide range in correlation among assessing performance was the argumentative view of DOPS test-retest reliability. Due to the reason, investigators concluded that the DOPS tool did not appear to be a reliable tool in assessing performance and not been consistent (24). Contrastingly, some authors considered the DOPS was a valid, reliable tool for assessing trainees' performance (9, 12, 13, 19). The amount of time allowed between measures is critical. We know that if we measure the same thing twice that the correlation between the observations will depend in part by how much time elapses between the two measurement occasions. The time elapses in our study we set was two weeks, we considered that was suitable for DOPS evaluation of splinting in Occupational Therapy practice. The shorter the time gap, the higher the correlation; the longer the time gap, the lower the correlation. This is because the two observations are related over time. To prevent the relative error and to decrease the learning effect of time issue we randomized items for raters.

In our results showed the modified version of DOPS demonstrated high intra-rater reliability and test-retest reliability over time (19). High reliability could be hard to accompany with an excessive number of raters (23). The new tool was highly acceptable to our senior staffs. It represents the tool is reliable for assessment of Occupational Therapy trainees. However, further enhancement of feedback provision from other facilities around the Taiwan appeared to be necessary to achieve its educational challenge and great acceptability.

The RCP originally developed DOPS to assess practical skills in the UK (11). Amidst the cultural, educational, and language differences between the UK and Taiwan, the majority of the original assessment areas were directly adaptable. This may well be due to the universality of procedural assessment in the practice across countries. To our knowledge, this is the first study, in which a process of validating and reliable an adapted version of DOPS had been reported. The outcome of this study expresses almost perfect content validity. Therefore, 10 assessment areas will be useful to any splinting skills context.

Strengths and limitations

The present investigation was the first Taiwan study to assess the intra-reliability, test-retest reliability and content validity of splinting technique educational curriculum. Furthermore, our study has several limitations including:

1. We did not recognize the effect of impression rating over the trainees due to the raters would cover the trainee's training programs. For further investigation, we suggest that other educators who did not cover trainees' programs or another facility could take the assessment to prevent the impressing effect on assessing the process.
2. Although the DOPS would be the important assessment tool in OT training programs of work-placed-assessment, we also suggest that for the global view of trainees' skill, 360-degree evaluation scores and mini-CEX would be included in further studies (11,20).
3. In reliability test of Item-Total Statistics for Case 2, we found item #6 (i.e. Seeks helps where appropriate) had disputed opinion (α=0.15). However, in Case 1 we did find the middle-correlation (α=0.37) among the raters. We postulated that would be difficult to rate the performance of student who was poor manipulation when seeking helps.
4. The study included only 14 occupational therapist assessors from one hospital in Taiwan. This small, homogeneous sample limits generalizability of the results. The validity and reliability testing of the DOPS would be strengthened by including raters from multiple sites and diverse settings. The small number of raters also provides less precise estimates of the tool's psychometric properties like internal consistency. Expanding the study to include: (1). Larger sample size of raters across different institutions. (2). Correlation analysis between DOPS and other clinical skills assessments.
5. The analysis of test-retest reliability only used 2 occupational therapy students as research subjects, and the sample size was indeed too limited. For analysis of the stability and consistency of assessment tools such as test-retest reliability, the larger the number of samples, the better the effect.

airiti

6. A validity analysis of DOPS for Taiwan was conducted by 14 experts, as shown in Table 2. The results indicated that the 12th expert gave a relatively low score of relevance (0.7) compared to the other 13 experts, who had an average score of 0.94. This discrepancy may affect the overall validity of the instrument and should be considered in future studies using this method. A possible solution is to exclude the 12th expert's assessment from the analysis and re-evaluate the impact on the overall validity of the instrument.
7. A possible direction for further research is to apply back translation and examine the alignment of the original and the back translated English texts.

Conclusions

The DOPS for Taiwan can be used for occupational therapists after a short period of training. This is the first attempt to evaluate the use of comprehensive workplace assessment across occupational therapists in Taiwan. The methods are feasible to conduct and can make reliable distinctions between OTs performances. With our adaptation, they may be appropriate for assessing the performance of other grades and specialties of ROT. This may be helpful in informing foundation assessment.

Reference:

1. Yang, Y. Y., Lee, F. Y., Hsu, H. C., Huang, C. C., Chen, J. W., Cheng, H. M., Huang, C. C. Assessment of first-year post-graduate residents: Usefulness of multiple tools. Journal of the Chinese Medical Association. 2011; 74(12), 531-538.
2. Edwards, M., & Martin, A. The objective structured clinical examination as a method of occupational therapy student evaluation. Canadian Journal of Occupational Therapy. 1989; 56(3), 128-131.
3. Hagemann, E., Williams, C. K., McKee, P., Stefanovich, A., & Carnahan, H.. Using Model Hands for Learning Orthotic Fabrication. American Journal of Occupational Therapy. 2014; 68(1), 86-94
4. Stefanovich, A., Williams, C., McKee, P., Hagemann, E., & Carnahan, H. Development and validation of tools for evaluation of orthosis fabrication. American Journal of Occupational Therapy. 2012; 66(6), 739-746
5. Westmorland, M., Salvatori, P., Tremblay, M., Jung, B., & Martin, A. The once and future programme: Educational innovation in occupational therapy. Canadian Journal of Occupational Therapy. 1996; 63(1), 44-54.
6. Doucet, B. M. Five Years Later: Achieving Professional Effectiveness to Move Neurorehabilitation Forward. American Journal of Occupational Therapy. 2013; 67(5), e106-e119.
7. Eardley, I., Bussey, M., Woodthorpe, A., Munsch, C., & Beard, J. Workplace-based assessment in surgical training: experiences from the Intercollegiate Surgical Curriculum Programme. ANZ Journal of Surgery. 2013; 83(6), 448-453. doi:10.1111/ans.12187
8. Moonen-van Loon, J. M., Overeem, K., Donkers, H. H., van der Vleuten, C. P., & Driessen, E. W. Composite reliability of a workplace-based assessment toolbox for postgraduate medical education. Advances in Health Sciences Education. 2013; 18(5), 1087-1102.
9. Rauf, S., Aurangzeb, W., Abbas, S., & Sadiq, N.. Work place based assessment in foundation year: Foundation University Medical College experience. Journal of Ayub Medical College, Abbottabad: JAMC. 2011; 23(4), 76-79
10. Singh, T., & Modi, J. N. Workplace based assessment: a step to promote competency based postgraduate training. Indian Pediatrics. 2013; 50(6), 553-559.
11. Wragg, A., Wade, W., Fuller, G., Cowan, G., & Mills, P. Assessing the performance of specialist registrars. Clinical Medicine. 2003; 3(2), 131-134.
12. Barton, J. R., Corbett, S., van der Vleuten, C. P., English Bowel Cancer Screening, P., & Endoscopy, U. K. J. A. G. f. G. The validity and reliability of a Direct Observation of Procedural Skills assessment tool: assessing colonoscopic skills of senior endoscopists. Gastrointestinal Endoscopy. 2012; 75(3), 591-597.
13. Delfino, A. E., Chandratilake, M., Altermatt, F. R., & Echevarria, G. Validation and piloting of direct observation of practical skills tool to assess intubation in the Chilean context. Medical Teacher. 2013; 35(3), 231-236.
14. Lynn, M. R. (1986). Determination and quantification of content validity. Nursing research. 1986; 35(6), 382-386.

airiti

15. Shumway, J. M., & Harden, R. M. AMEE Guide No. 25: The assessment of learning outcomes for the competent and reflective physician. Medical teacher. 2003; 25(6), 569-584.
16. Ku, P. W., Sun, W. J., & Chen, L. J. Development of the Sedentary Behavior Questionnaire for the Elderly. Sports Exerc. Res. 2016;18, 41-55.
17. Bari, V.. Direct observation of procedural skills in radiology. AJR. American Journal of Roentgenology. 2010; 195(1), W14-18.
18. O'Loughlin, E. J., Swann, A. D., English, J. D., & Ramadas, R. Accuracy, intra-and inter-rater reliability of three scoring systems for the glottic view at videolaryngoscopy. Anaesthesia. 2017;72(7), 835-839.
19. Naeem, N. Validity, reliability, feasibility, acceptability and educational impact of direct observation of procedural skills (DOPS). Jcpsp, Journal of the College of Physicians & Surgeons – Pakistan. 2013; 23(1), 77-82.
20. Chien, W.-T., & Norman, I. The validity and reliability of a Chinese version of the family burden interview schedule. Nursing research. 2004;53(5), 314-322.
21. Polit, D. F., & Beck, C. T.. The content validity index: are you sure you know what's being reported? Critique and recommendations. Research in nursing & health. 2006; 29(5), 489-497.
22. Sauls, D. J. The labor support questionnaire: Development and psychometric analysis. Journal of nursing measurement. 2004;12(2), 123-132.
23. Wilkinson, J. R., Crossley, J. G., Wragg, A., Mills, P., Cowan, G., & Wade, W. Implementing workplace-based assessment across the medical specialties in the United Kingdom. Medical Education. 2008;42(4), 364-373.
24. Mohamed, R., Shaheen, A. A., & Raman, M. Evaluation of colonoscopy skills - how well are we doing? Canadian Journal of Gastroenterology. 2011; 25(4), 198-200.