[ARTCLE]

© Authors. Published under an International Creative Commons Attribution License 4.0

ISSN On-Line 2310-3883
ISSN Print 0038-2337

Psychometric Properties of the Creative Participation Assessment administered by occupational therapy undergraduate students in their work integrated learning

AUTHORS
*Kristen J. Cronje [a]
https://orcid.org/0009-0005-4377-4721
*Amanda Aitken [a]
https://orcid.org/0000-0002-0922-257X
*Elianna R de Clercq [a]
https://orcid.org/0009-0009-6274-0892
Daleen Castelein [a]
https://orcid.org/0000-0002-0611-8662

AFFILIATIONS
[a] Occupational Therapy Department, School of Health Care Sciences, University of Pretoria, Prinshof Campus, Bophelo Road, Prinshof, Pretoria, Gauteng South Africa
* Undergraduate students at the University of Pretoria at the time of study

CORRESPONDING AUTHOR
Kristen Cronje
kristencronje12@gmail.com

KEYWORDS
Vona du Toit Model of Creative Ability, clinical meaningful assessment, structured teaching, supervised practice, construct validity, good health and well-being, quality education

HOW TO CITE THIS ARTICLE
Cronje K J, Aitken A, de Clercq E R, Casteleijn D. Psychometric Properties of the Creative Participation Assessment administered by occupational therapy undergraduate students in their work integrated learning. *South African Journal of Occupational Therapy*. Volume 56 Number 2. August 2026, pp 1-11. DOI: https://doi.org/10.17159/2310-3833/2026/vol56no2a6

ARTICLE HISTORY
Received: 5 October 2024
Reviewed 1st round: 8 October 2024
1st round revision: 11 December 2025
Reviewed 2nd round: 17 December 2025
2nd round revision: 28 April 2026
Accepted: 30 April 2026
First Online: 1 August 2026

EDITOR
Hester van Biljon
https://orcid.org/0000-0003-4433-6457

DATA AVAILABILITY
Upon reasonable request from corresponding author

FUNDING
No funding was received for this study

ABSTRACT

Introduction: Undergraduate occupational therapy students in South Africa use the Creative Participation Assessment during their work-integrated learning to assess clients' level of creative ability. However, the psychometric properties of this assessment tool when administered by undergraduate students have not yet been investigated.
Aim and objectives: To investigate the inter-rater reliability, intra-rater reliability, internal consistency and construct validity of the Creative Participation Assessment when administered by undergraduate students.
Methods: The study used samples collected from third and fourth-year students at a South African tertiary institution. In Study A, prospective data was collected from students based on a case study video in 2023 to assess inter-rater reliability. In Study B, retrospective data was collected from students during their work integrated learning blocks in 2021, 2022 and 2023 to assess intra-rater reliability, internal consistency, and construct validity. Analysis of inter-rater and intra-rater reliability was conducted using inter-class correlation coefficients. Internal consistency was analysed using Cronbach's alpha. An analysis of construct validity was performed using the Rasch model.
Results: Poor to moderate inter-rater reliability (0.422) was found with improvements when considering only the levels of creative ability (0.575). Excellent intra-rater reliability was indicated, with improved reliability for the fourth year (0.960) compared to the third-year group (0.889). Good construct validity was demonstrated, with a good overall fit to the Rasch model (χ^2 probability = 0.187), ordered thresholds across all items, and good item-fit, although some evidence of uni-dimensionality was noted.
Conclusion: These findings support the use of the Creative Participation Assessment as a valid and clinically meaningful assessment within undergraduate occupational therapy training, particularly in mental health settings. They emphasise the need for structured teaching and supervised practice to enhance scoring consistency, while the strong construct validity underscores its value in informing intervention planning and monitoring client progress.

Implications for practice
- Undergraduate education of creative ability has value informing intervention planning and monitoring client progress.
- The Creative Participation Assessment when used during undergraduate training is a valid and meaningful assessment tool.
- Further research for occupational therapy student administration of assessments in practice is needed as is structured teaching and supervised practice.

INTRODUCTION

The occupational therapy undergraduate curriculum requires students to administer assessment and treatment to clients with various diagnoses during their work integrated learning[1]. These are structured learning opportunities and clinical training in collaboration with facilities such as residential care facilities, and private and public hospitals[2]. One of the assessment tools commonly used is the Creative Participation Assessment (CPA) tool.

The CPA is an assessment tool that was developed by Dain van der Reyden and colleagues in the early 1970s for use with mainly psychiatric inpatients. This was one of the first assessment tools that was based on the Vona du Toit Model of Creative Ability (VdTMoCA)[3]. It is formatted as a table with twelve items (rows) and six levels (columns). The assessor will rate each of these items on the corresponding six levels, ranging from clients showcasing low to high functioning abilities for these items. Within each level, there are an additional three distinct and sequential phases. Raters are encouraged to use three separate activities during the assessment, namely a habituated, familiar and unfamiliar task, of which scores are later averaged for the final result. To administer this assessment tool, the occupational therapist must be knowledgeable of the VdTMoCA constructs and the presentation of the levels of creative ability[4, 5]. Although students are taught the VdTMoCA in their undergraduate curriculum, the application of this model in their assessment and treatment practice has not yet been explored.

The VdTMoCA proposes that a relationship exists between volition, motivation and action through the emergence of creative ability. Creative ability emerges when volition directs the motivation to act[3]. Creative ability is presented in nine different levels, of which the first six are represented on the CPA. The level of creative ability is deduced from the quality of action performed by an individual[3,4], as expressed by the twelve items on the CPA. The phase within the level is then determined by the quality of performance and support required during activity participation[4,6]. In South Africa, the VdTMoCA is used for assessment, and intervention purposes by enabling occupational therapists to monitor the client's level of functioning. It is thus essential for students to administer this assessment tool consistently and accurately, as this informs both assessment findings and treatment programmes.

Previous research has established the psychometric properties of the CPA when administered by qualified occupational therapists. Casteleijn and Smit[13] investigated construct validity, criterion validity, internal consistency and inter-rater reliability of the CPA. These properties were found acceptable with a population of clients with schizophrenia. Samsonraj et al[14] found good concurrent validity with the Global Assessment of Functioning but a poor correlation with the Canadian Occupational Performance Measure; Casteleijn[15] used the Rasch analysis to show that the thresholds for the levels of creative ability were ordered in the CPA as well as two other outcome measures based on the VdTMoCA, the Activity Participation Outcome Measure (APOM) and the Functional Levels Outcome Measure (FLOM). This study confirmed that the levels exist and have a hierarchical progression from low to high ability. Van der Linde and Casteleijn[16] further showed evidence of high convergence with a nursing tool, the Fairview self-help scale, in a population of clients with intellectual disabilities. All of the above studies were conducted with a mental healthcare client population, with qualified occupational therapists as raters. However, there is no literature investigating the administration of the CPA by students.

Literature surrounding student administration of any assessment tools, and the psychometric properties thereof, is minimal. Assessment tools that have been investigated include the use of dynamometers to investigate grip and pinch strength[7] and the figure-of-eight circumferential assessment of hand volume[8], when assessing the physical rehabilitation population. Both studies revealed that students were reliable raters of these assessments, with the latter also indicating validity in their rating abilities. Other assessment tools frequently used by students, such as the Model of Human Occupation Screening Tool (MOHOST)[9,10], and the Functional Independence Measure + Functional assessment Measure (FIM+FAM) assessment tool[11,12] have been determined to be valid and reliable when administered by qualified occupational therapists, however no further studies have explored their application when used by students. This leaves limited valid and reliable tools for use by the student population.

The aim of this research study was therefore to investigate the psychometric properties of the CPA when administered by undergraduate students.

METHOD

Study design

This quantitative, non-experimental psychometric evaluation study was conducted in two components (Study A and Study B) to determine the measurement properties of the CPA when administered by students during their work integrated learning. The design aligns with the COSMIN guidelines[21,22] for studies on measurement properties, focusing on aspects of reliability, internal consistency and construct validity.

Study A used prospective data to investigate inter-rater reliability.

Study B used retrospective data to investigate intra-rater reliability, internal consistency and construct validity.

Study populations and sample

Study A's population were CPA records collected from fourth year students at a South African tertiary institution during a video study in 2023.

To determine the inter-rater reliability in Study A, non-probability purposive sampling was used. The available population size was 44 students within the single 4th year class. Gwet[17] recommended that a sample size of 40 is appropriate when using an ICC reliability analysis. This considers the number of raters multiplied by the number of cases as well as the number of items to be evaluated (total CPA items)[17]. As there is only one video analysis, 40 raters were required for inter-rater reliability. If 40 raters each independently rate the same video using the CPA with 12 items, it results in 480 total ratings. This is a sufficient sample and provides a good level of precision[18] with confidence intervals within an acceptable range (±0.05 to ±0.10).

Study B's population were CPA records collected from third and fourth-year occupational therapy students in 2021, 2022 and 2023, that were administered on their individual patients during their work integrated learning. Different sampling methods were applied.

For the analysis of intra-rater reliability in Study B, simple random probability sampling was used from the population of third and fourth-year students at one tertiary institution. The sample size requirements for measuring intra-rater reliability were calculated using the Kappa online sample size calculator[19]. Using a minimal acceptable Kappa value of 0.6 indicates moderate reliability[20], a significance level of 0.05; and a power (1-β) value of 80%; a sample size of 14 was calculated. In order to obtain this sample size, random samples of 14 were drawn using an online random number generator. This was repeated three times, in order to repeat the analysis several times for a more conclusive result. Additionally, a random sample of 14 was drawn from both the third and fourth-year groups, in order to compare these analyses.

In determining internal consistency and construct validity in Study B, non-probability convenience sampling from a population of CPAs administered by third and fourth-year students at the tertiary institution from the period between 2021 to 2023, was used. Based on the COSMIN conceptual framework[21], the sample size requirements for measuring construct validity are seven units per item (in this case 84 CPA forms), but not less than 100[22]. The guidelines indicate that a sample size greater than 100 will provide reliable results[22].

Inclusion and exclusion criteria

Study A included all CPAs completed by a single class of fourth year students during one video study. CPAs were excluded if any of the items were not completed.

Study B included CPAs completed by third- and fourth-year students at residential, acute, forensic and outpatient mental healthcare facilities. CPAs with one to three assessment activities and those with or without phases indicated on them were included in the study. The analysis software automatically excluded records with missing data. The inclusion and exclusion criteria remained consistent for the records sampled for the analysis of intra-rater reliability, internal consistency and construct validity.

Data Collection

For Study A, a video study was conducted with a class of fourth-year students in 2023. The students who agreed to participate all received a short introduction to the study and a reminder of how to administer the CPA. In this study, the CPA scoring was adapted from the original six levels to include the three phases within each level, resulting in an 18-point scale. This decision was motivated by the need to empirically test the hierarchical structure of both levels and phases of creative ability, which are theoretically sequential. The adapted scoring also aligns with the approach used in the Activity Participation Outcome Measure (APOM), where incorporating phases into the scoring has been shown to produce a sensitive and psychometrically robust scale[15]. This approach allowed for more refined measurement of change and better alignment with clinical practice, where therapists often indicate the phase for each item (instead of one overall phase for all items combined) to reflect subtle variations in functioning. Although this adaptation deviates from the original CPA instructions [4], it enabled a more detailed psychometric analysis of the construct and provided insights for potential refinement of the tool for educational and clinical use.

Participants watched the four-minute video where one mental healthcare client participated in an unfamiliar instrumental activity of daily living (making a pizza, handling tools, and selecting toppings for the pizza). Students were provided with background information which described the client's behaviour at home and at the day centre that the person visited daily, as well as a description of the person's social participation and use of free time. The video was shown twice. Participants completed a CPA form on hard copies which were collected after the session.

For Study B, the CPAs were digitally uploaded by third and fourth-year students during 2021, 2022 and 2023 onto the university's education platform as part of the curriculum criteria. We also accepted CPA forms as hard copies, because they adhered to the inclusion criteria, and were de-identified and processed in the same way as the digitised CPA forms. Some of these CPA uploads included an initial and follow-up assessment of the same client, with the follow-up assessment being conducted within two to six weeks of the initial assessment. These were downloaded from the platforms for processing for the intra-rater reliability analysis.

Research procedure

The CPAs for both Study A and B were processed similarly. After their collection, they were anonymised and assigned unique identification codes. The CPAs with follow-up samples were assigned identification codes that linked them together. Excel spreadsheets were created for each psychometric property for easier sampling.

When completing the CPA, students indicated the phase for each item by adding a tick mark towards the left of the column to indicate that the item score was in the therapist-directed phase, a tick mark in the middle of the column if it was in the patient-directed phase, and towards the right of the column if it was the transitional phase. During data input into the above excel spreadsheets, each of the indicated items on the CPAs were then assigned numerical values indicating phases within items, ranging from one to 18 based on their marked score. If the raters did not indicate the phase, the research group captured the score as the patient-directed phase. If the CPAs were administered with multiple assessment activities in Study B, the scores were additionally averaged in a calculation sheet for a single score.

Additionally, the CPAs for the inter-rater reliability objective within Study A were captured twice, with comparable scores. They were initially recorded with scores from one to 18 reflecting phases as above. They were subsequently re-captured in a collapsed scale from one to six, considering only the levels, and recorded in a separate Excel spreadsheet.

No post-hoc analysis was conducted in this research study.

Data analysis method

Data analysis occurred using various statistical analysis tools. The Rasch model was used to analyse construct validity using the RUMM2030® software; the IBM SPSS Statistics 29 was used to analyse internal consistency, inter-rater reliability and intra-rater reliability.

Inter-rater reliability

The analyses for inter-rater reliability in Study A were run twice, first with the scores representing the phases (scale of one to 18); and separately with the codes representing the levels only (scale of one to six).

Intraclass Correlation Coefficient (ICC) analysis, two-way mixed model with a 95% confidence interval, was used to investigate inter-rater reliability, with an absolute agreement between raters [23, 24]. ICC scores between 0.5 and 0.75 indicate moderate reliability, scores between 0.75 and 0.9 indicate good reliability, and scores above 0.9 indicate excellent inter-rater reliability[23, 24]. The ICC values derived from this analysis determined whether the students similarly understand, interpret, and subsequently administer the CPA; thus, indicating the inter-rater reliability of their administration of the CPAs in their work integrated learning placements.

Intra-rater reliability

An inter-class correlation coefficient (ICC) analysis, two-way mixed model with a 95% confidence interval and a consistency type, was used to analyse the intra-rater reliability in Study B. This was interpreted as the ICC interpretation above. Each set of scores for an individual rater (initial and follow-up assessment CPAs) were analysed in relation to one another, with an individual ICC score and p-value calculated for each. The values obtained from this analysis ascertained whether students remained consistent in their administration of the CPA after time had elapsed; thus, indicating the CPA's intra-rater reliability when administered by the same students. Additionally, ICC scores were obtained based on random samples of third and fourth-year raters, which were analysed and compared. The comparison of the ICC values obtained between the third and fourth-year students allowed for the observation of differences in administration between year groups.

Internal consistency

Cronbach's alpha was used as the outcome measure to analyse the internal consistency [25] in Study B. Cronbach's alpha measures all possible combinations of items and is a value between zero and one [26]. Cronbach's alpha values between 0.7 and 1, indicates internal consistency, with a measure above 0.8 being considered good and above 0.9 as excellent internal consistency [26]. An excellent internal consistency could indicate that the items of the CPA are consistently interpreted when administered by the students.

Construct validity

The analysis of the CPAs for construct validity in Study B was performed using the Rasch model. The Rasch model transposes ordinal data into interval data so that it can display consistent and equal intervals in the scoring of the items. The data run through the model is compared to a mathematical outline allowing for analysis if

it fits with the requirements of the Rasch model. Various tests were run, and outputs were analysed, which included the overall fit to the model, threshold map, person-item threshold distribution, item fit residuals, residual correlation matrix, and uni-dimensionality through the distribution of the t-test. The Rasch model subsequently indicates whether all the items coalesce to represent one construct or one dimension, called uni-dimensionality [27, 28].
Should the Rasch analysis indicate construct validity, it can be said that each item within the CPA contributes uniquely to the overall construct of creative ability, specifically when the form is administered by students.

Overall fit to the Model
For data to be analysed through the Rasch model, it must fit the requirements of the model. The model stipulates a mean value of 0 and a standard deviation value of 1 to indicate a perfect fit. This analysis determines the mean and standard deviation values of the given data, which can be compared to these expected values. A chi square value is also given, where a value greater than 0.05 indicates that there is no significant difference between the expected requirements (the Rasch requirements) and the observed data (the results of the CPA data) and thus the data fits the model.

Threshold map
Ordering of the response categories (scores) within a data set is a prerequisite for analysing data through the Rasch model. The threshold map is therefore outputted to visually represent the sequential order of scores that represent items of the CPA to determine if there is an incremental scale within the items of the CPA. All of the items were visually displayed on the threshold map to determine if the scoring of the item followed a linear order. In other words, did a person with less ability to perform the item, have a lower score while a person with higher ability scored higher on the scale. A disordered threshold would be found when two thresholds overlapped for that item, indicating that the scoring is not in a linear order [29]. If this occurs, scores of the items of the CPA form may need renumbering, be collapsed into adjacent categories or the item completely removed from the tool.

Person-item threshold distribution
The person-item threshold distribution analysis displays the dispersion of data on a logit scale. The person data distribution is expected to display a normal distribution curve, which represents the expected bell shape of the population. The item data distribution is expected to display a flat curve, which indicates the equal representation of each of the items on the scale. The mean score should be close to 0 while the standard deviation should be around 1.

Item fit
The item fit analysis indicates if an item in the assessment tool uniquely contributes to the assessment result. The tool items are represented as either overfitting or redundant [27, 28]. A fit residual above 2.5 indicates a misfit of the items [30], which may indicate that a specific item is redundant and therefore does not uniquely contribute.

Person-item residual correlation matrix
The correlation matrices indicate the extent to which the different items correlate to one another when measuring the construct. A correlation above 0.3 shows a medium correlation and a correlation above 0.4 indicates high correlation between items. This may indicate that both items measure the same elements of the construct, and therefore only one may be necessary [31]. If an item has a high correlation, it may be considered to remove this item from the instrument.

Uni-dimensionality
Uni-dimensionality is achieved when all items are measuring one underlying construct. In RUMM2030 uni-dimensionality is performed by principal component analysis by analysing the negatively correlated items and the positively correlated items using a two-tailed t-test. These are represented in a distribution graph. Uni-dimensionality is indicated by a normal distribution bell curve with the peak around zero, with no more than 5% of the data falling below -1.96 or above 1.96 [32].

Ethical Approval
Ethical approval was obtained from the Faculty of Health Sciences Research Ethics Committee of the University of Pretoria with ethical clearance number 647/2022.

RESULTS
Two different demographics were used in this study. Firstly, in Study A and B, the students who administered the CPAs, who studied or were currently enrolled at a South African tertiary institution that offers occupational therapy as an undergraduate course. Secondly, in Study B, are the clients assessed by the students in residential, acute, forensic and outpatient mental healthcare facilities. These included three residential care facilities, six acute facilities, one chronic facility and one facility admitting both acute and chronic cases. Students were expected to remove any personal demographic characteristics when submitting their CPAs to the academic platform to maintain confidentiality, thus the research team was unable to determine at which institution these clients were admitted.

For Study A, thirty-three records were completed during the video study and submitted for use. After exclusions and data cleaning, 23 records could be used for statistical analysis.
For study B, a total of 106 records were entered for Rasch analysis, while 40 records were extracted for intra-rater reliability

Inter-raterreliability
To investigate inter-rater reliability in Study A, two ICC analyses were conducted, firstly including levels and phases, and secondly including levels only. Both analyses had p-values below 0.05, indicating statistically significant results. When analysing the samples with phases included (scale of one to 18), the ICC was found to be 0.422, indicating poor agreement between the raters [23, 24]. When analysing the samples with levels only (scale of one to six), the ICC was 0.575, which indicates moderate agreement among the raters [23, 24]. (Refer to Appendix 1: ICC results (page 10)

Intra-rater reliability
Of the forty raters sampled for the intra-rater reliability analysis, 15 were third-year students, and 25 were fourth-year students. All the ICC scores of the forty samples had a p-value below 0.05; thus, indicating their statistical significance. The forty ICC scores of the third and fourth years were averaged from three random samples of 14. Thereafter, two additional random samples of 14 were separately taken from the 25 fourth-year and the 15 third-year ICC scores.

The average ICC scores for the three random samples of both the third and fourth-year students combined were 0.914, 0.920, and 0.968. These samples had ICC values above 0.9, which indicates excellent intra-rater reliability[23,24]. The average ICC score for the random sample of only third-year students was 0.889, and the average for the random sample of only fourth-year students was 0.960. Both measures indicate good intra-rater reliability; however, the fourth-year students indicate excellent intra-rater reliability[23,24].

Internal consistency
The Cronbach Alpha value derived from the retrospective samples review, through the analysis of 106 CPAs, was 0.981. This value is indicative of excellent internal consistency between the items of the CPA[26].(Refer to Appendix 1: ICC results(page 10)

Construct validity

Overall fit to the model

In the Rasch analysis, the mean of the data was -0.688. This was indicative of a good level of fit to the model, with the expected mean on the logit scale being close to zero. The only score which was indicative of invariance to the model was the standard deviation, which was 2.1088. This was higher than the expected score of one[33 3]. The Chi-Square probability obtained was 0.187, which indicates little variance from the requirements of the model and thus demonstrating good construct validity[29].

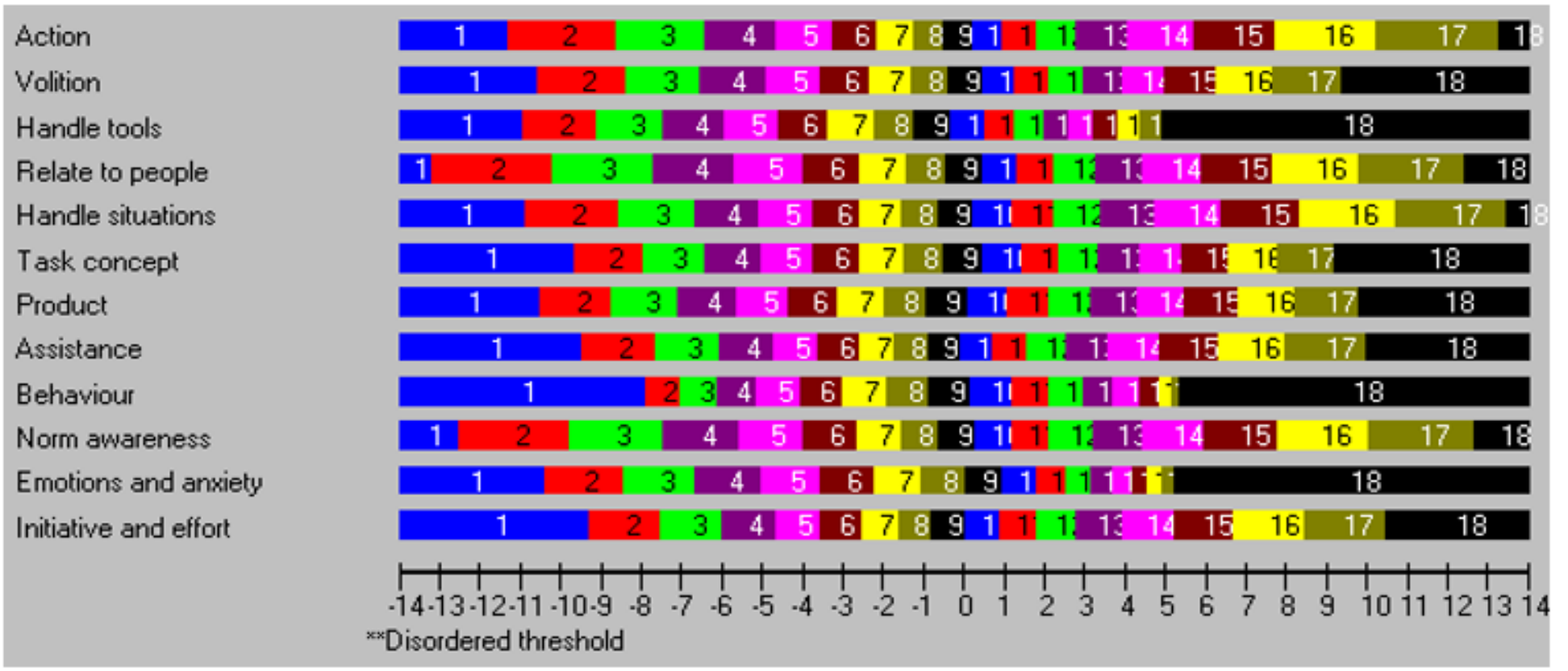


Figure 1: Threshold map of the dataset as displayed on the RUMM2030® software

The threshold map (Figure 1, above) indicated that each item of the CPA appeared on a linear scale from one to 18. None of the vertical bars were excluded for any score, thus no disordered thresholds were found for any of the items. This indicated that each of the phases within the levels represented a unique level of ability (in this case level of creative participation), as each of the scores from one to 18 represented a level of creative participation on the map. Persons with higher scores were accurately measured as having more ability than those with lower scores. Additionally, it is indicative of an incremental scale within the tool, with a linear progression between the levels.

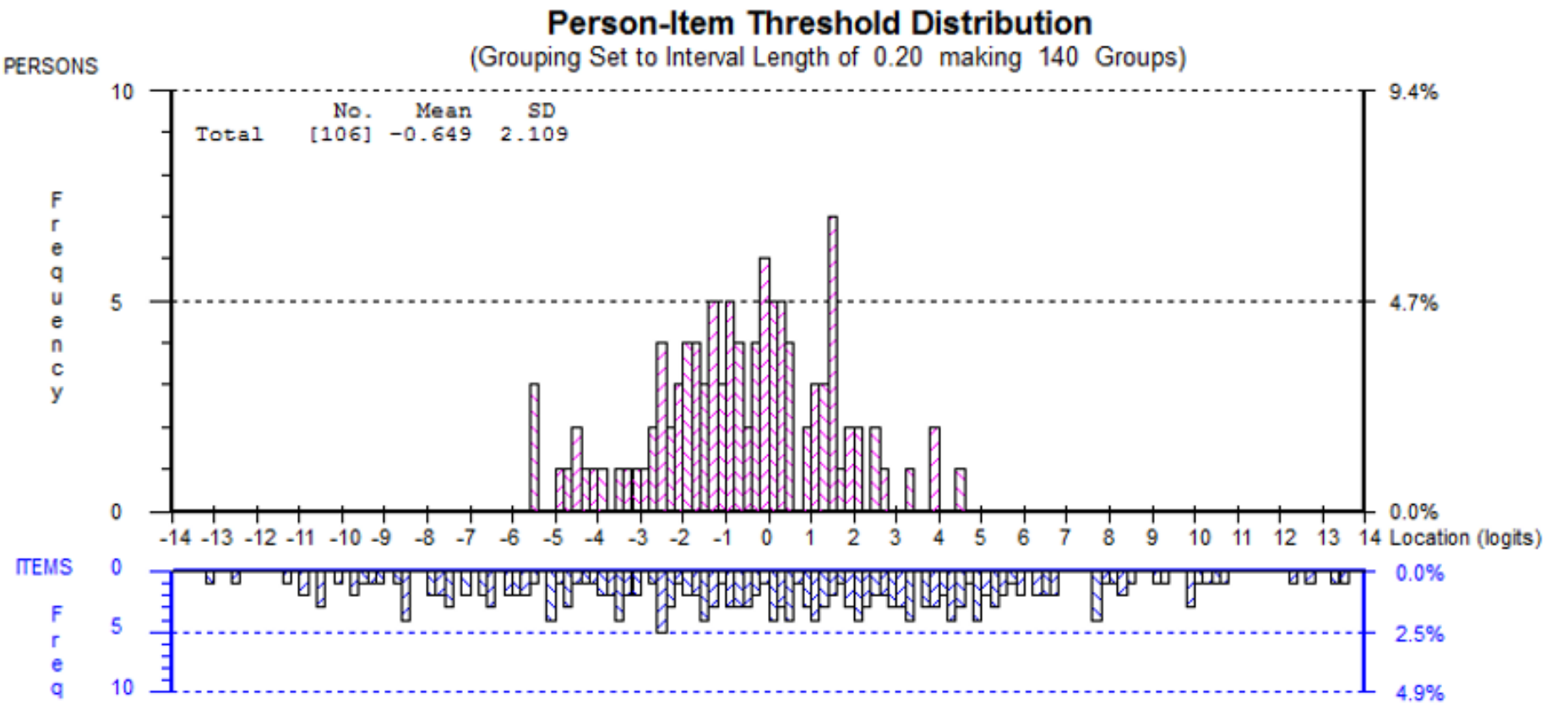


Figure 2: Bar graph of the dataset indicating the person-item threshold distribution as displayed on the RUMM2030® software Person-item threshold distribution

The person distribution, (Figure 2, above) represented by the top (pink) histograms, reflects the scores provided by the raters, while the bottom (blue) histograms represent the item threshold distribution for each item of the CPA. This data was plotted on a logit scale from-14 to 14.

Within the person distribution, a mean of -0.648 and a large standard deviation of 2.109 was observed within the sample size of 106. The distribution displayed a bell curve which was slanted slightly more to the negative side (the lower ability of creative participation)

yet close to a normal distribution of the patient demographic. The standard deviation of 2.109 could be due to a large number of persons on the 1,5 logit scale.

Within the item-threshold distribution (the categories on the scale), none of the item threshold distributions exhibited a frequency of larger than five, and a flat curve is shown.

Item fit

In the item-fit analysis, only item 4, "Relate to people", had a residual fit score above 2.5, being 3.062. This indicates a possible misfit of this item.

Uni-dimensionality

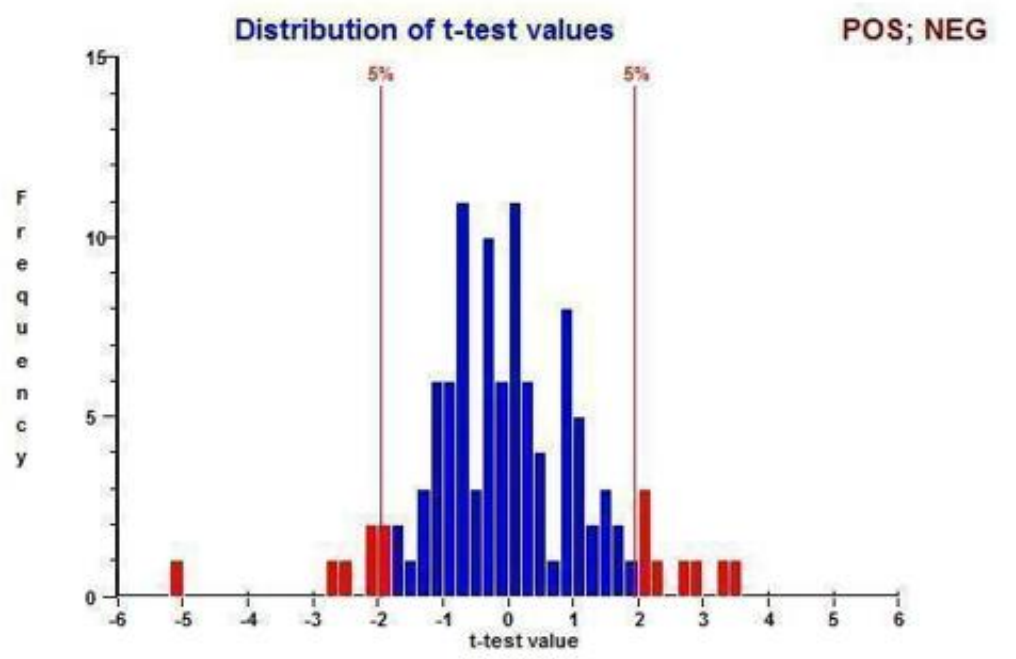


Figure 3: Bar graph as displayed on the RUMM2030® software indicating the uni-dimensionality of the dataset

In the analysis, 12.3% of the data fell outside of the confidence interval, with 7.5% of the dataset falling above 1.96, and 4.7% falling below -1.96. To achieve uni-dimensionality, less than 5% of the data should fall on either side of these values. Therefore, this analysis indicated that the data are not unidimensional. (Refer to Appendix 2 for summary statistics (page 11)

DISCUSSION

Both Study A and Study B indicated significant results pertaining to the psychometric properties of the CPA when administered by students, which have implications into the future use of the assessment tool at an undergraduate level.

The poor and moderate inter-rater reliability results in Study A indicate that the students had little agreement in the manner in which they administered the CPA. This could indicate a lack of understanding of the theory of creative ability or the CPA itself. However, it is also important to consider that the CPA requires interpretation of abstract constructs such as motivation and action, which are inherently less observable and more dependent on clinical judgement. As such, some variability between raters is expected, particularly among students who are still developing their clinical reasoning.

Notably, there was a higher agreement between raters when only the level of creative ability was analysed, compared to when both levels and phases were included. This may suggest that the students were better able to make broader distinctions between levels of creative ability, but experienced greater difficulty when making more detailed distinctions between phases. The more detailed 18-point phase scoring system requires finer clinical judgement, which may increase the variability between raters.

The structure of the undergraduate curriculum could support these results. Training may not be provided in proportion to the complexity of creative ability theory, and subsequent CPA administration. For example, tools such as dynamometer for grip and pinch strength or the figure-of-eight circumferential assessment for hand volume, which students have been shown to be reliable raters of[7, 8], are less complex and more objective than the CPA and thus require a less intensive training approach. In contrast, the CPA demands interpretation and synthesis of observed behaviour, which places greater demands on clinical reasoning.

Through the comparison of this result and that obtained from Casteleijn and Smit[13], showing significant inter-rater reliability of the CPA when administered by qualified occupational therapists, it can be inferred that experience in the field, and subsequent exposure to the administration of the CPA, is an important factor in improving the reliability of the rater[13]. This discrepancy between student and qualified therapist performance highlights the role of experience, training, and calibration in achieving agreement between raters.

This interpretation is further supported by the results obtained from the analysis of the intra-rater reliability. Although a combination of the third- and fourth-year's CPAs yielded excellent intra-rater reliability, the fourth-year intra-rater reliability was superior to the third-year reliability. This suggests that individual students were able to apply the CPA consistently over time, and that this consistency improves with increased exposure and experience. The fourth years having had more opportunities for client interaction and practice in administration of the CPA, likely developed a more stable internal understanding of the assessment, which contributed to their higher intra-rater reliability.

It should be noted that the intra-rater reliability result in Study B does not indicate that the results of the CPA are the same in the follow-up but rather that the manner of completion of the form was consistent. Additionally, this is a method of reliability and does not indicate the validity of the scores and the accuracy of the subsequent conclusion on the clients' creative ability.

The findings from the Rasch analysis strengthen confidence in the construct validity of the CPA, indicating that it produces scores that meaningfully reflect differences in clients' levels of creative ability, even when administered by students. The confirmation of a linear and incremental structure across both levels and phases extends earlier findings by Casteleijn[5], where linearity was demonstrated primarily at the level of creative ability when administered by qualified

Person-item residual correlation matrix

In the person-item residual correlation matrix, three pairs of items indicated medium correlation between one another, with a correlation value above 0.3. Item three (Handle Tools and Materials) and four (Relate to People) had a correlation of 0.311; items four and seven (Product) had a correlation of 0.362; and items six (Task Concept) and nine (Behaviour) had a correlation of 0.311.

(Refer to Appendix 3 for Person-item residual correlation matrix (page 11)

occupational therapists. The current study therefore contributes additional evidence that the full structure of the CPA, including phases within levels, operates as a coherent measurement continuum.

Importantly, this has implications for the sensitivity of the CPA in detecting change over time. The 18-point scale allows for more refined differentiation of client functioning and progression. In contrast, while assessments such as the Model of Human Occupation Screening Tool use a 4-point rating across multiple items (yielding a broader total score for psychometric analysis), the scaling structure provides less detailed differentiation at the level of individual ratings. The CPA's hierarchical scaling of levels and phases therefore offers a distinct advantage in capturing incremental changes in creative ability.

This enhanced sensitivity supports its use as a re-assessment tool, enabling clinicians to identify subtle but clinically meaningful changes over the course of intervention. This is consistent with findings by Carpenter[5], who reported that occupational therapists experienced the CPA as particularly useful in detecting nuanced changes in client participation and progression.

Further analysing of construct validity indicated that the item "Relate to people", demonstrated a high fit residual compared to other items. In the correlation matrix, this item showed modest associations with "Handle tools and materials" (0.311) and "Product" (0.362). These values suggest some shared variance between items, which is not unexpected in measures assessing related aspects of functioning, but do not indicate problematic overlap or redundancy.

Importantly, from a theoretical perspective grounded in the construct of creative ability, "Relate to people" represents a distinct dimension of functioning, specifically reflecting interpersonal engagement and communication[6,35]. This information cannot be inferred from task-related items such as "Handle tools and Materials" or "Product" formation, and it is essential for clinical decision-making, including planning for group participation, social interaction demands, and the therapeutic use of self [6,35].

Similarly, items such as "Behaviour" and "Task concept" captures conceptually distinct aspects of creative ability and contribute uniquely to the overall profile of a client. Furthermore, the result of excellent internal consistency of the CPA indicates that students can consistently interpret and administer these items of the CPA, which supports their understanding of the value of the distinct items.

Uni-dimensionality is a foundational psychometric requirement for valid measurement, referring to the extent to which all items in an instrument measure a single underlying construct[29]. In Rasch analysis, this implies that each item contributes meaningfully to the same latent trait, without excessive redundancy or measuring unrelated attributes[33,36]. An instrument demonstrating uni-dimensionality provides evidence that its scores form a coherent and interpretable continuum, which is essential for meaningful comparison and outcome monitoring.

In occupational therapy, where constructs such as motivation, task performance, and social participation are inherently complex, uni-dimensionality ensures that a total score reflects a theoretically grounded construct rather than a combination of loosely related abilities[32,36,37]. From a clinical perspective, this strengthens confidence that observed changes in scores represent real changes in functioning, thereby supporting sound intervention planning and evaluation. While some variability is expected given the multidimensional nature of human occupation, establishing sufficient uni-dimensionality remains critical for the valid interpretation and use of assessment results.

Only one analysis indicated a possible lack of uni-dimensionality and thus weaker construct validity. This may be indicative that some items in the CPA, when administered by students, are not contributing uniquely towards the assessment of creative participation as a single construct. However, the sample demographics should be considered, as the CPAs in this study were primarily administered to institutionalised clients with moderately low levels of functioning. In the overall fit to the model, a slightly negative mean was evident, which may suggest that the scores on the lower range of the scale were represented more frequently in the data. This is supported by the very high Cronbach's Alpha score indicating excellent internal consistency, which may be due to consistently low CPA scores within the population. Furthermore, a high standard deviation was evident, which may indicate outliers caused by clients in this population having very high or low scores. With the highest negative-negative and positive-positive correlations being laid on a scale to analyse uni-dimensionality, these outliers of the data set will be included and will deviate the data further from the mean. It may be possible that a larger or more representative sample of all possible levels would produce fewer outliers and thus result in improved uni-dimensionality. Regardless, the overall result from the Rasch analysis indicated that the CPA had good construct validity when administered by undergraduate students.

Limitations

Various limitations were experienced, which may have affected the results. Due to difficulty in interpretation for coding the CPAs, eight were excluded in general. In the research procedure, due to interpretation that was necessary for averaging phases, and approximating markings, human error may have affected the data. While care was taken to minimise these errors, they may have influenced the results.

A key limitation relates to the achieved the sample size for inter-rater reliability analysis. Although a minimum of 40 CPAs was initially targeted, due to non-participation, only 33 were obtained, and after exclusions and data cleaning, the final sample available for analysis was substantially smaller (n=23). The reduced sample size increases the risk of Type II error, meaning that true differences between raters may not have been detected. This is particularly important given the study aimed to demonstrate agreement between raters, and therefore results should be interpreted with caution.

Furthermore, the exclusion of incomplete or incorrectly completed CPAs may have introduced bias, as students with a poorer understanding of the CPA were likely underrepresented in the final dataset. This may have resulted in an overestimation of agreement and consistency.

Finally, this study was conducted within mental health practice settings, with CPAs primarily completed on institutionalised clients with lower levels of functioning. This reflects the typical clinical training context in which undergraduate occupational therapy students are introduced to and apply the CPA. As such, the findings are directly relevant to student learning and assessment practices within this context.

While the results therefore provide important evidence supporting the use of the CPA in mental health training settings, caution should be exercised in generalising these findings to other practice areas. Differences in the nature and expression of creative ability across clinical populations may influence how students interpret and assess clients' creative participation when using the CPA. Future research may explore its application in additional settings; however, the current study makes a meaningful contribution by establishing its validity within a key and representative area of occupational therapy training.

Recommendations

The findings highlight the importance of strengthening undergraduate training in both the VdTMoCA and the administration of the CPA. Given the complexity of the assessment, students should receive structured teaching that integrates theoretical understanding with supervised practical application. Particular emphasis should be placed on differentiating between items, as well as accurately interpreting levels and phases of creative ability. The use of applied teaching strategies, such as case-based learning, the use of videos and

guided observation, may further support consistent and confident administration.

In addition, consideration should be given to standardising CPA training within undergraduate curricula to promote consistency in its application. For example, universities could implement a structured training module that includes core theoretical teaching of the VdTMoCA, guided scoring of standardised case scenarios (e.g. video-based assessments), and the use of common marking rubrics or calibration exercises to align students' scoring practices. Collaboration between universities is encouraged, including the sharing of teaching resources, case materials, and assessment guidelines, to support a more unified and standardised approach to CPA training across institutions in South Africa.

Future research should aim to replicate these findings with larger and more representative samples to address the risk of Type II error identified in this study. Further investigation into students' competency in administering the CPA, as well as its application across different clinical training contexts, is also recommended.

CONCLUSION

The study yielded evidence that the CPA demonstrates acceptable reliability and good construct validity when administered by undergraduate occupational therapy students, particularly when applied to clients in mental health settings. While variability in scoring at the phase level and across third- and fourth-year groups was observed, the overall findings suggest that students are able to apply the CPA in a consistent and theoretically coherent manner. The Rasch analysis provides strong evidence of internal construct validity, confirming that the CPA functions as a coherent and theoretically aligned measure of creative participation when administered by undergraduate students. The CPA has a unique contribution to the occupational therapy discipline with significant impact on client care. Therefore, it is essential to ensure that both practitioners and students can administer it reliably and validly to ensure quality clinical practice.

Conflicts of interest

No conflicts of interest to declare.

Acknowledgements

Acknowledgement is made of Nabeel Ally Mohamed, Chantell Chiloane, Tasha Erasmus, and Karabo Makgopela for contributions to the study in the initial phases of literature review, data collection, and writing of the initial draft of the manuscript.

Author Contributions

This study was supervised by Daleen Casteleijn, who also suggested the outline of the study. Amanda Aitken, Kristen Cronje, Elianna de Clercq developed the protocol, collected and analysed the data, and wrote the manuscript. The final manuscript was prepared by the research team and submitted by Kristen Cronje.

Artificial Intelligence (AI) declaration

The initial manuscript was conceptualised and drafted in 2023 without the use of artificial intelligence (AI) tools. During the revision process, AI-assisted tools were used to support language refinement, clarity of expression, and editing in response to reviewer feedback. AI was not used for the generation of original ideas, data analysis, interpretation of findings, or the development of the study design. All content was critically reviewed, verified, and approved by the authors, who take full responsibility for the integrity and accuracy of the manuscript.

REFERENCES

1. Allison H, Turpin MJ. Development of the student placement evaluation form: A tool for assessing student fieldwork performance. Australian Occupational Therapy Journal. 2004;51(3):125-32 https://dx.doi.org/https://doi.org/10.1111/j.1440-1630.2004.00414.x
2. Evenson ME. Fieldwork: The Transition from Student to Professional. Willard and Spackman's occupational therapy. 13th edition. ed. Philadelphia: Wolters Kluwer; 2019. p. 1988-2023.
3. du Toit V. An investigation into the correlation between volition and its expression in action. South African Journal of Occupational Therapy. 1974;6(9) https://dx.doi.org/https://doi.org/10.1177/030802267403701123.
4. Van der Reyden D, Sherwood W. The Vona du Toit model of creative ability core constructs and concepts. The Vona du Toit model of creative ability : origins, constructs, principles and application in occupational therapy. 1st ed. Pretoria, South Africa: The Vona and Marie du Toit Foundation; 2019. p. 58-105.
5. Carpenter C. What are Occupational Therapist's experiences of utilising the Vona du Toit Model of Creative Ability: Creative Participation Assessment in forensic mental health settings in the U.K? : Oxford Brookes University; 2016
6. Crouch R, Alers V. Creative ability: a model for individual and group occupational therapy for clients with psychosocial dysfunction. Occupational therapy in psychiatry and mental health: John Wiley & Sons; 2014.
7. Lindstrom-Hazel D, Kratt A, Bix L. Interrater reliability of students using hand and pinch dynamometers. The American Journal of Occupational Therapy. 2009;63(2):193-7 https://dx.doi.org/https://doi.org/10.5014/ajot.63.2.193.
8. Nadar MS, Taaqi M. Reliability of occupational therapy students using the figure-of-eight technique of measuring hand volume. Hong Kong Journal of Occupational Therapy. 2013;23(1):20-5 https://dx.doi.org/https://doi.org/10.1016/j.hkjot.2013.04.001.
9. Kielhofner G, Burke JP. A model of human occupation, part 1. Conceptual framework and content. The American Journal of Occupational Therapy. 1980;34(9):572-81 https://dx.doi.org/https://doi.org/10.5014/ajot.34.9.572.
10. Kielhofner G, Fan C-W, Morley M, Garnham M, Heasman D, Forsyth K, et al. A psychometric study of the Model of Human Occupation Screening Tool (MOHOST). Hong Kong Journal of Occupational Therapy. 2010;20(2):63-70 https://dx.doi.org/https://doi.org/10.1016/s1569-1861(11)70005-5.
11. Hall K. Characteristics and comparisons of functional assessment indices; Disability Rating Item, Functional Independence Measure, and Functional Assessment Measure. Journal of Head Trauma Rehabilitation. 1993;8:60-70 https://dx.doi.org/https://doi.org/10.1097/00001199-199308020-00008.
12. Turner-Stokes L, Siegert RJ. A comprehensive psychometric evaluation of the UK FIM+ FAM. Disability and rehabilitation. 2013;35(22):1885-95 https://dx.doi.org/https://doi.org/10.3109/09638288.2013.766271.
13. Casteleijn D, Smit C. The psychometric properties of the creative participation assessment. South African Journal of Occupational Therapy. 2002;32(1):6-11.
14. Samsonraj RS, Loughran MF, Secker J. Evaluating outcomes of therapies offered by occupational therapists in adult mental health. Journal of Mental Health. 2012;21(6):531-8 https://dx.doi.org/https://doi.org/10.3109/09638237.2011.638000.
15. Casteleijn D. Using measurement principles to confirm the levels of creative ability as described in the Vona du Toit Model of Creative Ability. South African Journal of Occupational Therapy. 2014;44(1):14-9. https://journals.assaf.org.za/index.php/sajot/article/view/19684
16. Van der Linde J, Casteleijn D. A comparison of two assessments of levels of functioning in clients with intellectual disability between occupational therapists and nursing staff within a long-term mental healthcare facility in South Africa. Curationis. 2016;39(1):1-10 https://dx.doi.org/https://doi.org/10.4102/curationis.v39i1.1665.
17. Gwet KL. Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters: Advanced Analytics, LLC; 2014.

18. Jia, B., Lynn, H.S. A sample size planning approach that considers both statistical significance and clinical significance. Trials 16, 213 (2015). https://doi.org/10.1186/s13063-015-0727-9
19. Arifin WN. A Web-based Sample Size Calculator for Reliability Studies. Education in medicine journal. 2018;10(3) https://dx.doi.org/https://doi.org/10.21315/eimj2018.10.3.8.
20. Harvey ND. A Simple Guide to Inter-rater, Intra-rater and Test-retest Reliability for Animal Behaviour Studies. 2021 https://dx.doi.org/https://doi.org/10.31219/osf.io/8stpy.
21. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. Journal of clinical epidemiology. 2010;63(7):737-45 https://pubmed.ncbi.nlm.nih.gov/20169472/.
22. Mokkink LB, Prinsen C, Patrick DL, Alonso J, Bouter LM, de Vet HC, et al. COSMIN Study Design checklist for Patient-reported outcome measurement instruments. Amsterdam, The Netherlands. 2019:1-32. https://pubmed.mcbi.nlm.nhi.gov/27330520/
23. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. PubMed Central. 2016;15(2):155-63 https://dx.doi.org/10.1016/j.jcm.2016.02.012.
24. Landers RN. Computing Intraclass Correlations (ICC) as Estimates of Interrater Reliability in SPSS. The Winnower. 2015 https://dx.doi.org/10.15200/winn.143518.81744.
25. Spiliotopoulou G. Reliability reconsidered: Cronbach's alpha and paediatric assessment in occupational therapy. Australian Occupational Therapy Journal. 2009;56(3):150-5 https://dx.doi.org/https://doi.org/10.1111/j.1440-1630.2009.00785.x.
26. Glen S. Cronbach's Alpha: definition, interpretation SPSS: Statistics How To; 2022 [Available from: https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/cronbachs-alpha-spss/].
27. Baghaei P. The Rasch Model as a Construct Validation Tool. Rasch Measurement Transactions 2008;22(1):1145-62 https://dx.doi.org/https://doi.org/10.21317/ksd.75.6.3.
28. Salzberger T. Reporting a Rasch analysis. Rasch Models in Health. 2012:347-62 https://dx.doi.org/https://doi.org/10.1002/9781118574454.ch19.
29. Linacre JM, Wright B. Rasch measurement transactions. MESA Press Chicago; 1995. p. 883-92.
30. Chatterjee S. Residuals. In: Lovric M, editor. International Encyclopedia of Statistical Science. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. p. 1227-9.
31. Wilberforce M, Sköldunger A, Edvardsson D. A Rasch analysis of the Person-Centred Climate Questionnaire–staff version. BMC Health Services Research. 2019;19(1):1-9 https://dx.doi.org/https://doi.org/10.1186/s12913-019-4803-9.
32. Tennant A, Pallant J. Unidimensionality matters!(A Tale of Two Smiths?). Rasch Measurement Transactions, 20 (1), 1048–1051. 2006.
33. Bond T, Yan Z, Heene M. Applying the Rasch model: Fundamental measurement in the human sciences: Routledge; 2020. https://doi.org/10.1080/02813432.2019.1568703
34. Christensen KS, Oernboel E, Nielsen MG, Bech P. Diagnosing depression in primary care: a Rasch analysis of the major depression inventory. Scandinavian Journal of Primary Health Care. 2019;37(1):105-12 https://dx.doi.org/ https://doi.org/10.1080/02813432.2019.156870.
35. Sherwood W, De Witt P. Assessment of Creative Ability. The Vona du Toit model of creative ability : origins, constructs, principles and application in occupational therapy. 1st ed. Pretoria, South Africa: The Vona and Marie du Toit Foundation; 2019. p. 148-99.
36. Hagell P. Testing rating scale unidimensionality using the principal component analysis (PCA)/t-test protocol with the Rasch model: the primacy of theory over statistics. Open Journal of Statistics. 2014;4(6):456-65 https://doi.org/10.4236/ojs.2014.46044.
37. Tennant A, Küçükdeveci AA. Application of the Rasch measurement model in rehabilitation research and practice: early developments, current practice, and future challenges. Frontiers in Rehabilitation Sciences. 2023; Volume 4 - 2023 doi: https://doi.org/10.3389/fresc.2023.1208670

Appendix 1: Intra-class Correlation Coefficient Results

Phases

Intraclass Correlation Coefficient

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0		
		Lower Bound	Upper Bound	Value	df1	df2
Single Measures	.030[a]	.005	.110	2.769	11	253
Average Measures	.422[c]	.116	.747	2.769	11	253

Intraclass Correlation Coefficient

	F Test with .. Sig
Single Measures	.002
Average Measures	.002

Levels

Intraclass Correlation Coefficient

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0		
		Lower Bound	Upper Bound	Value	df1	df2
Single Measures	.063[a]	.017	.201	3.352	11	231
Average Measures	.596[c]	.278	.847	3.352	11	231

Intraclass Correlation Coefficient

	F Test with .. Sig
Single Measures	<,001
Average Measures	<,001

Appendix 2: Construct Validity - Summary Statistics

SUMMARY STATISTICS for Analysis Name RUN1

ITEM - PERSON INTERACTION

ITEMS

	Location		Fit Residual
Mean	0.0000	Mean	0.1617
Std Dev	0.5092	Std Dev	1.2582
Skewness	-1.5690	Skewness	0.5830
Kurtosis	3.5853	Kurtosis	0.1209
		Correlation [location/stdResidual]	-0.1471

PERSONS

	Location		Fit Residual
Mean	-0.6488	Mean	-0.5385
Std Dev	2.1088	Std Dev	1.6633
Skewness	-0.1987	Skewness	0.1276
Kurtosis	0.0199	Kurtosis	-0.7561
Include Extremes	N = 106	Correlation [location/stdResidual]	0.0574

ITEM - TRAIT INTERACTION

Total - Item Chi Square	29.9027
Degrees of Freedom	24
Chi Square Probability	0.187998

RELIABILITY INDICES

	Run1
PerSepIdx:	
* with extms	0.98134
* NO extms	0.98134
CoefficAlpha	
* with extms	N/A
* NO extms	N/A

[Coefficient alpha not applicable with missing data]

LIKELIHOOD RATIO TEST

Analysis	Likelihood		
anaName1		ChiSq	
anaName2		DegF	
		Prob	

POWER OF ANALYSIS OF FIT

Excellent
Good
Reasonable
Low
Too Low

EXCELLENT

This display is intended as a guide ONLY and should be used in conjunction with other analysis indicators

< Display Control

File Text Format: Fixed / Tab Delimit

Save

Appendix 3: Construct Validity - Person-item residual correlation matrix

Person-Item Residual Correlation Matrix

Principal Component Summary | PC Normalised Vectors | PC Loadings | Varimax Rotation Loadings | Residual Correlation Matrix | Person-Item Residuals

Item	I0001	I0002	I0003	I0004	I0005	I0006	I0007	I0008	I0009	I0010	I0011	I0012
I0001	1.000											
I0002	0.222	1.000										
I0003	-0.088	0.025	1.000									
I0004	-0.240	-0.195	-0.311	1.000								
I0005	-0.137	-0.055	-0.147	0.262	1.000							
I0006	-0.110	-0.207	-0.278	0.020	0.091	1.000						
I0007	0.120	-0.182	0.052	-0.362	-0.149	-0.014	1.000					
I0008	-0.012	-0.114	0.144	-0.262	-0.247	0.014	0.059	1.000				
I0009	-0.077	-0.116	-0.161	-0.049	-0.092	-0.311	-0.153	-0.216	1.000			
I0010	-0.200	-0.115	-0.003	0.016	-0.111	0.047	-0.016	-0.199	-0.004	1.000		
I0011	-0.283	-0.157	-0.128	-0.016	-0.139	-0.070	-0.154	-0.080	0.052	-0.092	1.000	
I0012	-0.158	0.063	-0.218	-0.036	-0.234	-0.101	-0.080	-0.142	0.126	-0.178	-0.025	1.000